

# A univariable and multivariable analysis of right censored time-to-event data based on restricted mean survival time: A combination with traditional survival methods.

**Qiao Huang**

Wuhan University Zhongnan Hospital

**Jun Lyv**

Xi'an Jiaotong University

**Bing-hui Li**

Wuhan University Zhongnan Hospital

**Lin-lu Ma**

Wuhan University Zhongnan Hospital

**Tong Deng**

Henan University

**Xian-Tao Zeng** (✉ [zengxiantao1128@163.com](mailto:zengxiantao1128@163.com))

Wuhan University Zhongnan Hospital <https://orcid.org/0000-0002-3566-1023>

---

## Research article

**Keywords:** survival analysis; restricted mean survival time; regression; pseudo-value; inverse probability of censoring weighting

**Posted Date:** April 30th, 2020

**DOI:** <https://doi.org/10.21203/rs.2.18724/v2>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background** Hazard ratio is considered as an appropriate effect measure of time-to-event data. However, hazard ratio is only valid when proportional hazards (PH) assumption is met. The use of the restricted mean survival time (RMST) is proposed and recommended without limitation of PH assumption.

**Method** 4405 osteosarcomas were captured from Surveillance, Epidemiology and End Results Program Database. Traditional survival analyses and RMST-based analyses were integrated into a flowchart and applied for univariable and multivariable analyses, using hazard ratio (HR) and difference in RMST (survival time lost or gain, STL or STG) as effect measures. The relationship between difference in RMST and HR were explored when PH assumption was and was not met, respectively.

**Results** In group comparison and univariable regressions, using difference in RMST calculated by Kaplan-Meier methods as reference, pseudo-value regressions ( $R^2=0.99$ ) and inverse probability of censoring probability (IPCW) regressions with group-specific weights ( $R^2=1.00$ ) provided more consistent estimation on difference in RMST than IPCW with individual weights ( $R^2=0.09$ ). In multivariable analysis, age (HR:1.03, STL: 3.86 months), diagnosis in 1970~1980s (HR:1.39 STL:27.49 months), metastasis (HR:4.47, STL: 202 months), surgery (HR:0.58, SLG:35.55 months) and radiation (HR:1.46, SLT:44.65 months) met PH assumption and were main independent factors for overall survival. In both univariable and multivariable variables, a robust negative logarithmic linear relationship between HRs estimated by Cox regression and differences in RMST by pseudo-value regressions was only observed when PH assumption was hold.

**Conclusion** The flowchart will be intuitive and helpful to instruct appropriate use of RMST based and traditional methods. RMST-based methods provided an absolute effect measure to inspect effects of covariates on survival time and promote evidence communication with HR. Difference in RMST should be reported with hazard ratio routinely.

## Background

Survival data are frequently adopted in prospective and oncological studies which involves both survival time (until the occurrence of an event of interest) and status (event or censor) [1]. This feature introduces a series of special methods called survival analysis. The features of censoring types make survival analyses different. The most common form of censoring is right censoring, which means the event of interest is beyond the end of follow-up. For such data, survival analysis generally starts from data exploration through plotting Kaplan-Meier (KM) curves, and then log-rank test or Wilcoxon test is carried out to compare survival curves. Finally, univariable and multivariable regressions are used to investigate the association between participants' survival time and one or more factors[2]. In most situations, hazard ratio (HR) from Cox proportional hazard regression (referred as Cox regression) is routinely considered as the preferred and main effect measure in oncology. Note that a key assumption for appropriately using Cox regression is proportional hazards (PH) which requires a constant HR over time [3]. PH assumption should be tested for each included covariate, which may increase workload. When PH assumption is not satisfied, estimated HR is invalid and incorrect conclusion may be achieved [4]. Moreover, the interpretation of HR may be a challenging since it is not an intuitively and clinically summary statistic.

Restricted mean survival time (RMST) has been promoted as an alternative and better summary statistic for survival data when PH assumption is violated[5, 6]. It is estimated through calculating the area under the survival

curve between 0 and pre-specific time ( $\tau$ ). RMST is a simple statistic in the presence of non-proportional hazards and can be interpreted in a clinically meaningful perspective. Difference in RMST indicates that patient will live longer (positive) or shorter (negative) and its magnitude presents the size of average gain or lost in life expectancy within pre-specific time ( $\tau$ ). Generally, unadjusted estimation on difference in RMST is completed adopting Kaplan-Meier method, while adjusted one via standardized survival curves or covariance analysis. In such situation, only one categorical variable of interest can be analyzed, such as interventional group and control group. Pseudo-value (PV) technique and inverse probability censoring weighting (IPCW) technique make it possible to explore the effects of several covariates on RMST in multivariable regression[7, 8].

Log-rank test, Wilcoxon test and Cox regression are traditional and common survival analyses, while the estimation of difference in RMST, RMST-based regressions by PV and IPCW are considered as novel and promising model-free methods. Few studies make a summary and comparison among those methods to help researchers to choose correctly and conveniently. Furthermore, it was found that different conclusions were achieved from between estimated HR and difference in RSMT when reanalyzing 54 randomized trials [9]. The relationship among HR, difference in RMST and coefficients estimated from RMST-based regression should be further discussed. In the current research, our objective is to try to build a clear flow chart and to compare different methods based on an actual oncological data.

## Methods

### Design and setting of the study

The Surveillance, Epidemiology and End Results (SEER) Program Database was accessed and information on all osteosarcoma cases with complete data on status and survival time from 1970s to 2000s was captured. Cases who died or censored at time 0 were excluded from the analysis because of different processing strategies in general statistical softwares and potential risk of incorrect explanation. .

### Covariates and outcome measures

Patients' demographic covariates included age, sex, the year of diagnosis and race. According to age, the patients could be categorized into  $>60$ years and  $\leq 60$  years while in terms of race, they were into white/black and others. Data on tumor included tumor size, the extension of disease, and American Joint Committee on Cancer for staging (AJCC), all of which contained a type of 'UNKNOWN'. Tumor size was categorized into  $>100$ mm,  $<100$ nm and 'UNKNOWN'. Moreover, data on whether surgery, chemotherapy and/or radiation were performed were recorded, all of which were classified into "yes" and "no". Patients' outcome variables of interest referred to overall survival, including dead for all causes or alive and survival time in months.

### Flowchart for statistical methods

A flowchart was designed to integrate conventional survival analyses and RMST-based analyses on right-censored data (figure 1) and it was divided into two steps: graphic description and group comparison, univariable and multivariable regressions.

Graphic description should be the first step to explore survival curves for both single group or multiple groups. Group comparison is suitable for a categorical covariate. Comparing distributions of two or more survival rates is performed by Log-rank test or Wilcoxon test. PH assumption is required for log-rank test, but not for Wilcoxon

test[10]. Meanwhile, difference in areas under survival curve between 2 or more groups can be estimated with pre-specific truncation time,  $\tau$ (RMST).

In univariable and multivariable regressions, we consider PH assumption as the first key step to select suitable regression. Compared with Cox regression, parametric survival regression required to test and pre-specify probability distribution of survival times, such as Weibull and Exponential distribution, meanwhile, time-dependent covariates (covariates changing over follow-up) and stratified analysis controlling for nuisance covariates, cannot be incorporated easily. If PH assumption was satisfied, Cox regression would be a preferred and robust method because of its attractive features. If not, two modified Cox regression (Cox regression with time-dependent covariates and stratified Cox regression) and parametric survival regression can be performed instead of standard Cox regression. In addition, RMST-based regression can be an alternative and promising method to deal with non-proportional hazard data, pseudo-value regression (PV) and inverse probability of censoring weighting regression (IPCW). Both of them assumes that censoring distribution is non-informative, where distribution of censoring time does not provide information for distribution of survival time.

In IPCW regression, censoring weights required to be estimated initially, which suggests censoring distribution must be correctly estimated [7]. IPCW method used Kaplan-Meier method to estimate the censoring distribution (modeling censoring mechanism) and conditional probability of having remained uncensored until time  $t$  is calculated as weight for each participant. Then, inverse of the estimated weights are included into iterative estimation of coefficients of IPCW regression. Use of Kaplan-Meier method assumes that all subjects share the same censoring distribution, namely homogeneity of censoring mechanism assumption. When a categorical covariate exists and censoring distribution in each group are not the same, it is appropriate to use group-specific weights where censoring mechanism keep homogenous in each group and weights were estimated using Kaplan-Meier technique for each group separately. Otherwise, biased weights will be used in coefficient estimation, resulting in incorrect conclusion. There are two methods in IPCW regression, IPCW regression with individual weights and IPCW regression with group-specified weights. If only one categorical variable is of interest and homogeneity of censoring mechanism is questionable, it would be more appropriate to adopt IPCW regression with group-specified weights.

Pseudo-values technique has been developed for direct regression modeling survival function, RSMT and competing risks based on right censored data[11]. In PV regression on RMST, jackknife leave-one-out estimation is employed to generate pseudo-values. Firstly, the area under the Kaplan-Meier curve up to time  $\tau$  is estimated as  $\hat{\theta}$  on the complete data with  $n$  sample size, then remove  $i$ th subject and repeat the above estimation to get leave-one-out estimator as  $(\hat{\theta}^{(-i)})$ . Finally, the  $i$ th pseudo-values of  $(\hat{\theta}_i)$  can be calculated based on the difference between the complete sample and the leave-one-out estimator:  $(\hat{\theta}_i) = n\hat{\theta} - (n-1)(\hat{\theta}^{(-i)})$ . After calculating the pseudo-value for each subject, generalized estimating equation is used to model the effects of covariates on the RMST [8]. Since no weights are estimated from censoring distribution, PV method is not restricted by the homogeneity of censoring mechanism assumption.

PH assumption and homogeneity of censoring mechanism need to check before selecting suitable methods. There are three different methods to evaluate PH assumption: 1) graphical assessment (KM curves and  $\ln(-\ln(S(t)))$  vs  $\ln(t)$  Curves); 2) add a time interaction,  $x \times \log(t)$ , into COX regression and test its significance; 3) check global good of fitness by plotting and testing association between ranked survival time and Schoenfeld residuals[12, 13]. Violations of the PH assumption suggests existence of interactions between one or more covariates and time. Homogeneity of censoring mechanism was evaluated applying Kaplan-Meier curves and log-rank test. Note that the

event of interest was coded as censoring while others as event to draw Kaplan-Meier plot and conduct log-rank test. The curves showing substantial separation or small P value from log-rank test supported the violation of the homogeneity of censoring mechanism, which suggested censoring patterns of groups were different.

## Statistical analysis

The characteristics of the participants were summarized for descriptive statistics. Categorical variables were described using frequencies and percentages, while continuous ones as means with SD or as medians with interquartile ranges when data were skewed. Statistical inference, including conventional survival analyses and RMST-based analyses, complied with the flowchart based on statistical knowledge (figure 1).

Graphics assessment was selected because large sample size may make hypothesis tests in aforementioned method 2 and method 3 be sensitive to little deviation from PH assumption. This assumption would be met when curves in Kaplan-Meier plot did not cross and curves in the log (-log (survival)) versus log of survival time graph paralleled with each other approximately.

For comparing results among these methods, all methods were conducted ignoring the two assumptions in group comparison and regressions.  $\tau$  was specified utilizing the smallest value of the largest follow-up time across groups minus 5 months. Multiple comparison was employed in group comparison if there were more than 2 groups. Both univariable and multivariable regressions were conducted for Cox regressions and RMST-based regressions.  $\tau$  specified in univariable regression was same as that obtained in group comparison. In multivariable analyses,  $\tau$  was fixed at 480 months (24years).

To explore the relationship between HR and difference in RMST as well as the influence of PH assumption on such relationship, we performed data visualization and model fitting. Since HR was expressed as ratio while coefficients based on RMST as difference, a log transformed non-linear regression was considered when fitting the relationship between HR and difference in RMST (difference in RMST  $\sim$   $\log_e$ (HR)).

All analyses were carried out using the SAS software for windows, version 9.4 TS1M6 (SAS Institute Inc, Cary, NC) with a 2-sided significance threshold of  $P < .05$ . Graphic description and group comparison were carried out by LIFETEST procedure, and regressions were realized by PHREG procedure, and RMSTREG procedure. Data visualization were conducted in Microsoft Excel 2016.

## Results

Information for 4505 patients who met the inclusion criteria was extracted from SEER database. Of those patients, 100 patients with 0 survival time were excluded and 4405 were included for final analysis. Of 4405, 2389(54.23%) deaths were documented at the end of study during a median follow-up of 35 months. Descriptive characteristics were summarized into table 1. The median age of these patients was 30 years old, and 19.09% of them were older than 60 years. About half of the patients were male (57.05%). The minority of the patients had tumors more than 100mm (4.74%), metastasis (5.93%), III stage (0.61%) and IV stage (8.94%). Furthermore, the majority of the patients experienced surgery (77.80%) and chemotherapy (67.17%), while few cases (14.10%) underwent radiation.

### Group comparison and univariable regressions

Assumption tests for all covariates were summarized into table 2 and graphic assessment were collected into supplement. In group comparison (step 1), log-rank test, Wilcoxon test and unadjusted difference in RMST presented similar results at  $\alpha=0.05$  in all covariates but sex. Since sex did not satisfy PH assumption, Wilcoxon test was more appropriate and showed no statistical difference between male and female patients. Difference in RMST, however, showed that male patients would live shorter by 24.31 months than female ones significantly ( $P=0.004$ ). See in table 3.

In univariable regressions (step 2), majority of covariates were categorical variables. Using difference estimation in RMST based on Kaplan-Meier method as reference, PV regression and IPCW regression with group-specified weights exhibited similar effect ( $R^2=0.99$  and  $R^2=1.00$ , respectively), but IPCW regression with individual weights showed inconsistent effects ( $R^2 = 0.09$ ), see in table 3 and figure 2. Even though dichotomous age, sex, race and radiation met the homogeneity of censoring mechanism, IPCW regression with group-specified weights gave more consistent and stable estimates than those estimated by IPCW regression with individual weights (table 3 and figure 2). However, both sex, race and chemotherapy turned to be insignificant in the IPCW regressions with group-specific weights.

It was noticeable that group comparison in survival rate and IPCW regression with group-specified weights were no longer applicable for age as continuous variable. At this point, PV regression indicated that the average survival time was reduced by 4.09 months for each additional year in age ( $p<0.001$ ). Meanwhile, similar regression coefficient, 4.35, was acquired by IPCW regression with individual weights.

### **Multivariable regressions with Cox regression and RMST-based regressions**

In multivariable regressions, all covariates were included (table 4). IPCW regression with group-specified weights was not applicable since more than one categorical covariates broke the homogeneity of censoring mechanism. To compare with Cox regression, interactions between time and covariates that broke PH assumption were not included, their estimated HRs were a sort of average effect over time and may made misleading conclusions.

Compared with results from PV regressions, IPCW regression with individual weights presented great change in both P values and coefficients which suggested that individual weights was not appropriate. Negative difference in RMST was defined as survival time lost (STL) and positive difference was defined as survival time gain (STG). Moreover, HR was only presented for those covariates met PH assumption. The overall survival was independently associated with age (HR:1.03, STL: 3.86), sex (STL:24.23), diagnosis in 1970~1980s (HR: 1.39, STL:27.49), >100 mm tumor size (STL:49.56), metastasis and local invasion, higher AJCC staging, surgery (HR:0.58, STG:35.55), radiation (HR:1.46, STL:44.65) and chemotherapy (STL:49.49).

### **Relationship between HR and RMST**

Relationship between HR estimated by COX regressions and coefficients estimated by PV regressions were presented in both univariable analysis and multivariable analysis (table3, table 4 and figure 3). In both univariable and multivariable analyses, when covariates met PH assumption, a well-fitted and stable logarithmic relationships were observed between HR estimated by COX regressions and coefficients estimated by PV method ( $R^2 = 0.97$  and  $R^2=0.92$ ). But for those covariates that broke PH assumption, the estimated HRs were biased, both coefficients and goodness of fit of logarithmic relationship changed dramatically, -16.02 vs. -58.11 for coefficients and 0.39 vs. 0.91 for  $R^2$ , respectively.

## Discussion

The flowchart designed for right-censored survival data could guide the selection and combination of appropriate survival analysis methods (traditional methods and RMST-based methods). The estimation of difference in RMST between groups demonstrated a more intuitive and clinical interpretation than HR, survival time lost and gain. In univariable and multivariable regressions for RMST, pseudo-value regression provided a more robust estimation than inverse probability censoring weighting regression with individual weights. Although parameter estimates from IPCW regression with group-specified weights approximated to those from PV regressions in univariable regressions, statistical power decreased. In both univariable and multivariable regressions, a robust negative logarithmic linear relationship was observed between HRs estimated by Cox regression and coefficients (differences in RMST) by PV regression when PH assumption was hold.

Area under survival curve from time 0 to a pre-specific follow-up point was called RMST. It can be considered as the average survival time until an event of interest occurs during a defined period and be estimated without the limitation from proportional hazard assumption. If non-proportional hazards appears, biased estimates will be given by Cox proportional hazard model because HR is inconstant over time; meanwhile, clinical interpretation of HR may not be straightforward for clinicians[14]. Thereby, RMST and difference in RMST show several inferential and clinical advantages over other statistics in survival analysis. RMST and difference in RMST have been accepted by many researchers and widely used. In a randomized controlled trial, to reflect the convenience of Internet-accessed sexually transmitted infection testing, difference in RMST was tested to assess difference in the time to test[15]. A prospective cohort study exploring the relationship between midlife cardiorespiratory fitness and chronic obstructive pulmonary disease performed both Cox models and RMST analysis, estimated HR and changes in RMST of incidence of COPD and presented death simultaneously, which increased the reliability and understandability of the conclusion[16]. In addition to original studies, RMST and difference in RMST have been considered as effect measures in meta-analysis[17, 18]. RMST was regarded as an obligatory end point, while HR and difference in RMST as complementary methods to summarize treatment effect in oncological trials[19].

Unadjusted difference in RMST can be easily achieved through calculating the difference in area under the two survival curves based on Kaplan-Meier method. The existence of confounders may distort true relationship between exposure and outcome, confounders-adjusted difference in RMST is essential. Furthermore, precision and statistical power can be improved after adjusting prognostic covariates when compared to RMST[20]. An approach developed by Zucker can estimate covariates-adjusted RMST effect based on Cox regression[21]. But, it was limited to calculate RMST effect of a binary variable of interest (such as experimental group and control group) adjusted for confounders. Firstly, coefficients and model-based covariance matrix are extracted using COX regression, then a covariate-adjusted area under survival curve is constructed for each of the two groups separately, adjusted difference in RMST can be acquired finally.

When the categorical variable of interest has two or more two groups, standardized survival curve can be an alternative way is to obtain adjusted RMST and its difference[22]. This method has two steps: 1) defining a reference population for the confounders; 2) estimation of adjusted curves for that population. The definition of reference population is a critical step which determines the generalization of adjusted survival curve. There are two main approaches to compute adjusted curves: marginal analysis (balance data on all confounders using reweighting before modeling) and conditional analysis (modeling a comprehensive overall model and getting average predicted curves from a series of predicted survival curves for any combination of confounders).

Both aforementioned methods must make a clear division of included covariates into effect of interest vs. confounders, and adjusted RMST can be only estimated for the categorical effect of interest. Complex process and difficulty in extension limit their friendly applications in RMST estimation. In addition, PH assumption is assumed to be hold as utilization of Cox regression in Zucker's method. Likewise, when Cox regression is adopted in modeling part of standardized survival curves, meeting of PH assumption is also required.

RMST-based regressions, using generalized linear modeling technique, can study the effects of multiple factors or single factor on RMST without limitation on variable type and distinction between key variable and confounding variables. Compared to Cox regression based method and standardized survival curves, RMST-based regression could model RMST directly and facilitate model-based inference and prediction straightforward. Moreover, varied regression-based methods can be adopted, such as non-linear fitting, interaction effect and subgroup analysis.

Two types of direct modeling of RMST are developed. Pseudo-value regression was proposed by Anderson in 2003 to deal with time-to-event data and to make an inference[8]. Codes for pseudo-value method are available for three main platforms (SAS, R and Stata) [23, 24]. Another RMST-based regression was based on inverse probability censoring weighting technique proposed by Tian in 2014[7]. Weights estimation based on correctly specified censoring distribution is the key step. In univariable analyses of the current research, point estimation of difference of RMST were similar among Kaplan-Meier method, PV regressions and IPCW regression with grouped weights. When categorical covariate presents in IPCW regressions, individual weights or group-specified weights should be chosen based on the result of testing on homogeneity of censoring mechanism[25]. However, in this study, group-specified weights gave more accurate estimates than individual weights for both categorical covariates that met assumption and those that didn't meet. Even if categorical covariates satisfy homogeneity of censoring mechanism, censoring distribution in each subgroup can share similar parameters with that in complete data, group-specified weights can still work well. So, IPCW with group-specified weights may be a preferred method for categorical covariates. When more than one covariates included in multivariable regressions violate homogeneity of censoring mechanism assumption, it is hard to specify multiple strata in group-specific weights. At this time, forced adoption of IPCW with individual weights presents biased and misleading estimates. In addition, inverse probability may create extreme weight when the probability is close to 0, and small sample size will make inverse probability instable[26]. With such restrictions on IPCW regression, PV regression should be a preferred method in exploring the effects of more than one determinants on survival time in RMST-based regressions.

Ludovic et al reconstructed individual patient data based on survival curves for 54 randomized controlled trials, and HRs, ratio of RMST and differences in RMST were estimated and compared using those data. Accordingly, an agreement on the direction of treatment effect was observed in 50 trials, but neither linear nor non-linear relationship was fitted in their study. Furthermore, adjusted RMST cannot be derived based on KM curves in those trials for further comparison [27]. In the current study, a stable logarithmic relationship between difference in RMST by PV regression and HR was observed when proportional hazard assumption was held. Transformation between difference in RMST and HR can provide an alternative way to combine the effect size of meta-analysis in evidence-based medicine.

Generally, since a rise in HR corresponds to a reduction in restricted mean survival time, the sign slope of logarithmic transformed HR on difference in RMST is negative. Particular timespecified for RMST-based analysis determines the magnitude of difference in RMST and whether the difference is statistically significant or not. RMST should be typically calculated over a defined period that has adequate follow-up because of its time-dependent attribute. The type of end point, such as overall survival or progression-free survival, will also affect the estimation

of difference in RMST. So comparison and transformation between difference in RMST and HR should be cautious. In addition, a step by step tutorial about study design, sample size estimation, and the determination of  $\tau$  with RMST has been illustrated in details[28].

A strength in this study is that a flowchart for method selection has been made and discussed through comparing traditional methods to RMST-based analyses on oncological data with sufficient sample size. In addition, a logarithmic-transformed relationship was visualized between HR and difference in RMST with satisfied and unsatisfied PH assumption. A limitation in this study was that details on statistical theories for PV regressions and IPCW regressions were not presented. Not all of survival analyses were presented in the flowchart, survival analyses on left-censored data, interval-censored data and competing risks were not discussed here.

## Conclusion

In conclusion, HR and difference in RMST should be routinely reported with equal consideration for time-to-event data in prospective studies, which can improve the communication of clinical evidence. Furthermore, difference in RMST can make the interpretation of clinical results more straightforward than HR. The flowchart will instruct researchers and clinicians to select appropriate statistical methods for univariable analyses and multivariable analyses with the consideration of both PH assumption and the homogeneity of censoring mechanism assumption.

## List Of Abbreviations

RMST, restricted mean survival time

HR, Hazard ratio

PH, proportional hazards

STL or STG, survival time lost or gain

IPCW, inverse probability of censoring probability

PV, Pseudo-value

SEER, The Surveillance, Epidemiology and End Results

AJCC, American Joint Committee on Cancer for staging

## Declarations

**Ethics approval and consent to participate:** Not applicable.

**Consent for publication:** Not applicable.

**Availability of data and materials:**

Data and SAS code can be available upon request.

**Competing interests:** No

**Funding:** No

**Authors' contributions:**

Z: study design and literature review.

QH: study design, data analysis and writing.

JL: data collection and interpretation.

L and LL. M: Data collection.

TD: literature review.

All authors have read and approved the manuscript.

**Acknowledgements:**

This study used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the National Cancer Institute; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database.

**Authors' information (optional)**

Qiao Huang<sup>1</sup>, Jun Lyu<sup>2</sup>, Bing-hui Li<sup>1</sup>, Lin-lu Ma<sup>1</sup>, Tong Deng<sup>3</sup>, Xian-Tao Zeng<sup>1,4\*</sup>

\* corresponding author: Xian-Tao Zeng

Institution: Center for Evidence-Based and Translational Medicine, Zhongnan Hospital of Wuhan University, Wuhan, China

E-mail: zengxiantao1128@163.com

Phone: +86 13349999916

<sup>1</sup>Center for Evidence-Based and Translational Medicine, Zhongnan Hospital of Wuhan University, Wuhan, China

<sup>2</sup> Clinical Research Center, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

<sup>3</sup>Department of General Surgery, Center for Evidence-Based Medicine and Clinical Research, Huaihe Hospital of Henan University, Kaifeng, Henan, P.R. China

<sup>4</sup>Department of Urology, Zhongnan Hospital of Wuhan University, Wuhan, China

**Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

**References**

1. Clark TG, Bradburn MJ, Love SB, Altman DG: **Survival analysis part I: basic concepts and first analyses.** *Br J Cancer* 2003, **89**(2):232-238.
2. Bradburn MJ, Clark TG, Love SB, Altman DG: **Survival analysis part II: multivariate data analysis—an introduction to concepts and methods.** *Br J Cancer* 2003, **89**(3):431-436.
3. Blagoev KB, Wilkerson J, Fojo T: **Hazard ratios in cancer clinical trials—a primer.** *NAT REV CLIN ONCOL* 2012, **9**(3):178-183.
4. Hernan MA: **The hazards of hazard ratios.** *EPIDEMIOLOGY* 2010, **21**(1):13-15.
5. Royston P, Parmar MKB: **The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt.** *STAT MED* 2011, **30**(19):2409-2421.
6. Royston P, Parmar MK: **Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome.** *BMC MED RES METHODOL* 2013, **13**:152.
7. Tian L, Zhao L, Wei LJ: **Predicting the restricted mean event time with the subject's baseline covariates in survival analysis.** *BIostatistics* 2014, **15**(2):222-233.
8. Andersen PK, Hansen MG, Klein JP: **Regression analysis of restricted mean survival time based on pseudo-observations.** *LIFETIME DATA ANAL* 2004, **10**(4):335-350.
9. Trinquart L, Jacot J, Conner SC, Porcher R: **Comparison of Treatment Effects Measured by the Hazard Ratio and by the Ratio of Restricted Mean Survival Times in Oncology Randomized Controlled Trials.** *J CLIN ONCOL* 2016, **34**(15):1813-1819.
10. C. RLM, Naranjo MD: **A pretest for choosing between logrank and wilcoxon tests in the two-sample problem.** *METRON* 2010, **68**:111-125.
11. Klein JP, Gerster M, Andersen PK, Tarima S, Perme MP: **SAS and R functions to compute pseudo-values for censored data regression.** *Comput Methods Programs Biomed* 2008, **89**(3):289-300.
12. Tolosie K, Sharma MK: **Application of cox proportional hazards model in case of tuberculosis patients in selected addis ababa health centres, ethiopia.** *Tuberc Res Treat* 2014, **2014**:536976.
13. Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pélissier: **Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer.** *BMC MED RES METHODOL* 2010, **10**(1):20.
14. Calkins KL, Canan CE, Moore RD, Lesko CR, Lau B: **An application of restricted mean survival time in a competing risks setting: comparing time to ART initiation by injection drug use.** *BMC MED RES METHODOL* 2018, **18**(1):27.
15. Wilson E, Free C, Morris TP, Syred J, Ahamed I, Menon-Johansson AS, Palmer MJ, Barnard S, Rezel E, Baraitser P: **Internet-accessed sexually transmitted infection (e-STI) testing and results service: A randomised, single-blind, controlled trial.** *PLOS MED* 2017, **14**(12):e1002479.
16. Hansen GM, Marott JL, Holtermann A, Gyntelberg F, Lange P, Jensen MT: **Midlife cardiorespiratory fitness and the long-term risk of chronic obstructive pulmonary disease.** *THORAX* 2019.
17. Niglio SA, Jia R, Ji J, Ruder S, Patel VG, Martini A, Sfakianos JP, Marqueen KE, Waingankar N, Mehrazin R *et al*: **Programmed Death-1 or Programmed Death Ligand-1 Blockade in Patients with Platinum-resistant Metastatic Urothelial Cancer: A Systematic Review and Meta-analysis.** *EUR UROL* 2019.
18. Wei Y, Royston P, Tierney JF, Parmar MK: **Meta-analysis of time-to-event outcomes from randomized trials using restricted mean survival time: application to individual participant data.** *STAT MED* 2015, **34**(21):2881-

2898.

19. A'Hern RP: **Restricted Mean Survival Time: An Obligatory End Point for Time-to-Event Analysis in Cancer Trials?** *J CLIN ONCOL* 2016, **34**(28):3474-3476.
20. Karrison T, Kocherginsky M: **Restricted mean survival time: Does covariate adjustment improve precision in randomized clinical trials?** *CLIN TRIALS* 2018, **15**(2):178-188.
21. Zucker DM: **Restricted Mean Life with Covariates: Modification and Extension of a Useful Survival Analysis Method.** *J AM STAT ASSOC* 1998, **93**:702-709.
22. Therneau TM, Grambsch PM: *Modeling Survival Data: Extending the Cox Model.* New York: Springer; 2000.
23. Klein JP, Gerster M, Andersen PK, Tarima S, Perme MP: **SAS and R functions to compute pseudo-values for censored data regression.** *Comput Methods Programs Biomed* 2008, **89**(3):289-300.
24. Parner ET, Andersen PK: **Regression Analysis of Censored Data Using Pseudo-observations.** *The Stata Journal* 2010, **10**(3):408-422.
25. Willems S, Schat A, van Noorden MS, Fiocco M: **Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator.** *STAT METHODS MED RES* 2016, **27**(2):323-335.
26. Robins J, Sued M, Lei-Gomez Q, Rotnitzky A: **Comment: performance of double-robust estimators when "inverse probability" weights are highly variable.** *STAT SCI* 2007, **22**(4):544-559.
27. Trinquart L, Jacot J, Conner SC, Porcher R: **Comparison of Treatment Effects Measured by the Hazard Ratio and by the Ratio of Restricted Mean Survival Times in Oncology Randomized Controlled Trials.** *J CLIN ONCOL* 2016, **34**(15):1813-1819.
28. Pak K, Uno H, Kim DH, Tian L, Kane RC, Takeuchi M, Fu H, Claggett B, Wei LJ: **Interpretability of Cancer Clinical Trial Results Using Restricted Mean Survival Time as an Alternative to the Hazard Ratio.** *JAMA ONCOL* 2017, **3**(12):1692-1696.

## Tables

Table 1 Basic characteristics of osteosarcoma patients from SEER

Characteristics	n (%)
Outcome at the end of the study	
Death	2389(54.23%)
Free of death	2016(45.77%)
Age, years, median (Q1, Q3)	30.00(19.00, 54.00)
>60 years	841(19.09%)
<=60 years	3564(80.91%)
Sex, Male	2513(57.05%)
Year of diagnosis	
1970s~1980s	803(18.23%)
1990s~2000s	3602(81.77%)
Race	
White and black	4015(91.15%)
others	390(8.85%)
Tumor Size	
<100mm	655(14.87%)
>100 mm	209(4.74%)
unknown	3541(80.39%)
Extend of disease	
Confined	283(6.42%)
Local Invasion	887(20.14%)
Metastasis	261(5.93%)
Unknown	2974(67.51%)
AJCC	
I	207(4.70%)
II	886(20.11%)
III	27(0.61%)
IV	394(8.94%)
Unknown	2891(65.63%)
Surgery, yes	3427(77.80%)
Radiation, yes	621(14.10%)
Chemotherapy, yes	2959(67.17%)

Q1, the first quartile, Q3, the third quartile

AJCC, American Joint Committee on Cancer for staging.

Table 2 Assumption test of proportional hazard and homogeneity of censoring mechanism

Covariates	Proportional hazard assumption		Homogeneity of censoring mechanism	
	KM curves	ln (-ln(S(t))) vs ln(t) Curves	Graphical analysis	Log-rank test
Age(continuous)	-	-	-	-
Age(binary)	✓	✓	✓	✓
Sex	□	□	✓	✓
Year of diagnosis	✓	✓	□	□
Race	✓	✓	✓	✓
Tumor size	□	□	□	□
Extend of disease	✓	✓	□	□
AJCC	✓	□	□	□
Surgery	✓	✓	□	□
Radiation	✓	✓	✓	✓
Chemotherapy	□	□	□	□

KM, Kaplan-Meier. AJCC, American Joint Committee on Cancer for staging

Table3 univariable analysis based on conventional survival analysis and restricted mean survival time

Covariates	Difference analysis			Regression			
	Survival rate		Difference in RMST <sup>a</sup>	Cox regression <sup>b</sup>	RMST-based regression		
	Log-rank	Wilcoxon			PV regression <sup>a</sup>	IPCW regression (individual weights) <sup>a</sup>	IPCW regression (group-specified weight) <sup>a</sup>
Age (continuous)	-	-	-	1.13, <0.001	-4.09, <0.001	-4.35, <0.001	-
Age(binary)							
>60 years	<0.001	<0.001	-146.64, <0.001	3.82, <0.001	-147.29, <0.001	-137.17, <0.001	-146.43, <0.001
<=60 years	-	-	ref	ref	ref	ref	ref
Sex							
Male	0.005	0.058	-24.31, 0.004	1.12, 0.005	-24.22, 0.004	-60.12, 0.46	-23.98, 0.42
Female	-	-	ref	ref	ref	ref	ref
Year of diagnosis							
1970s~1980s	<0.001	<0.001	-26.89, <0.001	1.32, <0.001	-28.03, <0.001	178.96, <0.001	-26.27, <0.001
1990s~2000s	-	-	ref	ref	ref	ref	ref
Race							
White and black	0.005	0.006	-41.67, 0.01	1.24, 0.005	-39.76, 0.005	-165.32, 0.274	-41.80, 0.35
others	-	-	ref	ref	ref	ref	ref
Tumor Size							
>100 mm	<0.001*	<0.001*	<0.001	1.32, <0.001	-58.56, <0.001	-33.62, 0.010	-50.02, 0.07
unknown	-	-	-36.45, <0.001	1.42, <0.001	-39.35, <0.001	72.64, <0.001	-37.02, <0.001
<100mm(reference)	-	-	-	ref	ref	ref	ref
Extend of disease							
Metastasis	<0.001*	<0.001*	<0.001*	6.34, <0.001	-183.43, <0.001	-113.45, <0.001	-164.05, <0.001
Local Invasion	-	-	-60.15, <0.001	1.98, <0.001	-71.54, <0.001	-22.99, <0.001	-60.18, <0.001
Unknown	-	-	-82.28, <0.001	2.50, <0.001	-93.62, <0.001	2.21, <0.001	-82.94, <0.001
Confined	-	-	ref	ref	ref	ref	ref
AJCC							
IV	<0.001*	<0.001*	<0.001*	15.36, <0.001	-68.27, <0.001	-70.48, <0.001	-77.06, <0.001
III	-	-	-45.31, <0.001	6.511, <0.001	-41.12, <0.001	-51.98, <0.001	-45.77, <0.001
II	-	-	-28.72, <0.001	3.99, <0.001	-23.05, <0.001	-34.50, <0.001	-28.89, <0.001
Unknown	-	-	-39.58, <0.001	5.62, <0.001	-33.66, <0.001	-16.99, <0.001	-39.49, <0.001
I	-	-	ref	ref	ref	ref	ref
Surgery, yes	<0.001	<0.001	121.41, <0.001	0.37, <0.001	121.11, <0.001	18.17, 0.88	121.05, <0.001
Radiation, yes	<0.001	<0.001	-133.53, <0.001	2.54, <0.001	-134.41, <0.001	-29.15, 0.73	-133.39, <0.001
Chemotherapy, yes	<0.001	<0.001	16.26, 0.048	0.85, <0.001	17.04, 0.05	-137.623, 0.001	16.06, 0.27

RMST, restricted mean survival time. IPCW, inverse probability of censored weighting. AJCC, American Joint Committee on Cancer for staging

<sup>a</sup> Presented as point estimation of coefficient (difference in RMST) and corresponding p value

<sup>b</sup> Presented as point estimation of hazard ratio and corresponding p value

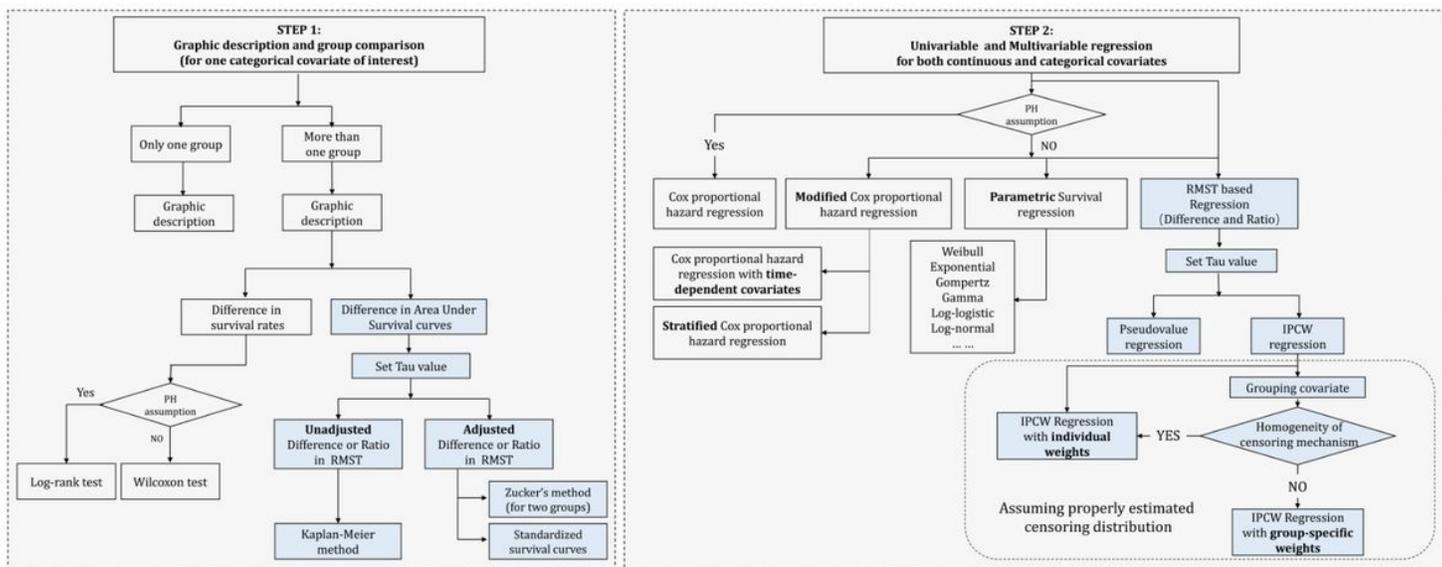
\* Overall test for categorical covariates with more than 2 groups.

Table 4 multivariable analysis based on cox proportional hazard regression and restricted mean survival time based regression.

Parameter	Cox regression		Pseudo-Value regression		IPCW regression (individual weights)	
	HR	p	Beta (95% CI), months	p	Beta (95% CI) , months	p
Age (continuous)	1.03(1.03,1.03)	<0.001	-3.86(-4.20,-3.53)	<0.001	-2.58(-3.12,-2.04)	<0.001
Sex, male	1.18(1.08,1.28)	<0.001	-24.23(-38.02,-10.45)	<0.001	-18.34(-47.03,10.35)	0.21
Year of diagnosis						
1970s~1980s	1.39(1.23,1.57)	<0.001	-27.49(-1.80,-53.19)	0.036	-197.65(-227.40,-167.91)	<0.001
1990s~2000s	ref		ref		ref	
Race						
White and black	1.02(0.88,1.19)	0.79	11.00(-12.32,34.32)	0.355	-14.67(-55.39,26.05)	0.48
other	ref		ref		ref	
Tumor Size						
>100 mm	1.48(1.21,1.80)	<0.001	-49.56(-85.01,-14.12)	0.006	-15.01(-29.61,-0.42)	0.044
unknown	1.12(0.97,1.31)	0.126	-11.51(-37.92,14.90)	0.393	-5.23(-19.48,9.03)	0.472
<100mm(reference)	ref		ref		ref	
EOD						
Metastasis	4.47(3.49,5.73)	<0.001	-202.47(-237.54,-167.39)	<0.001	-22.48(-45.64,0.69)	0.057
Local Invasion	1.66(1.33,2.07)	<0.001	-89.60(-121.37,-57.83)	<0.001	0.95(-13.51,15.41)	0.897
Unknown	1.81(1.43,2.29)	<0.001	-113.22(-147.32,-79.11)	<0.001	38.59(23.69,53.49)	<0.001
Confined	ref		ref		ref	
AJCC staging						
IV	11.44(7.29,17.93)	<0.001	-174.21(-201.43,-146.99)	<0.001	25.90(-4.87,56.68)	0.099
III	4.81(2.48,9.36)	<0.001	-105.87(-179.12,-32.62)	0.005	21.26(-7.95,50.47)	0.154
II	3.56(2.28,5.56)	<0.001	-61.97(-86.77,-37.17)	<0.001	20.22(-4.02,44.47)	0.102
Unknown	4.21(2.71,6.54)	<0.001	-89.53(-114.75,-64.31)	<0.001	75.49(56.17,94.81)	<0.001
I	ref		ref		ref	
Surgery, yes	0.58(0.53,0.64)	<0.001	35.55(18.57,52.53)	<0.001	53.15(17.20,89.10)	0.004
Radiation, yes	1.46(1.32,1.62)	<0.001	-44.65(-62.19,-27.11)	<0.001	-4.95(-58.00,48.09)	0.855
Chemotherapy, yes	1.26(1.14,1.39)	<0.001	-49.49(-67.41,-31.57)	<0.001	-55.42(-94.40,-16.45)	0.005

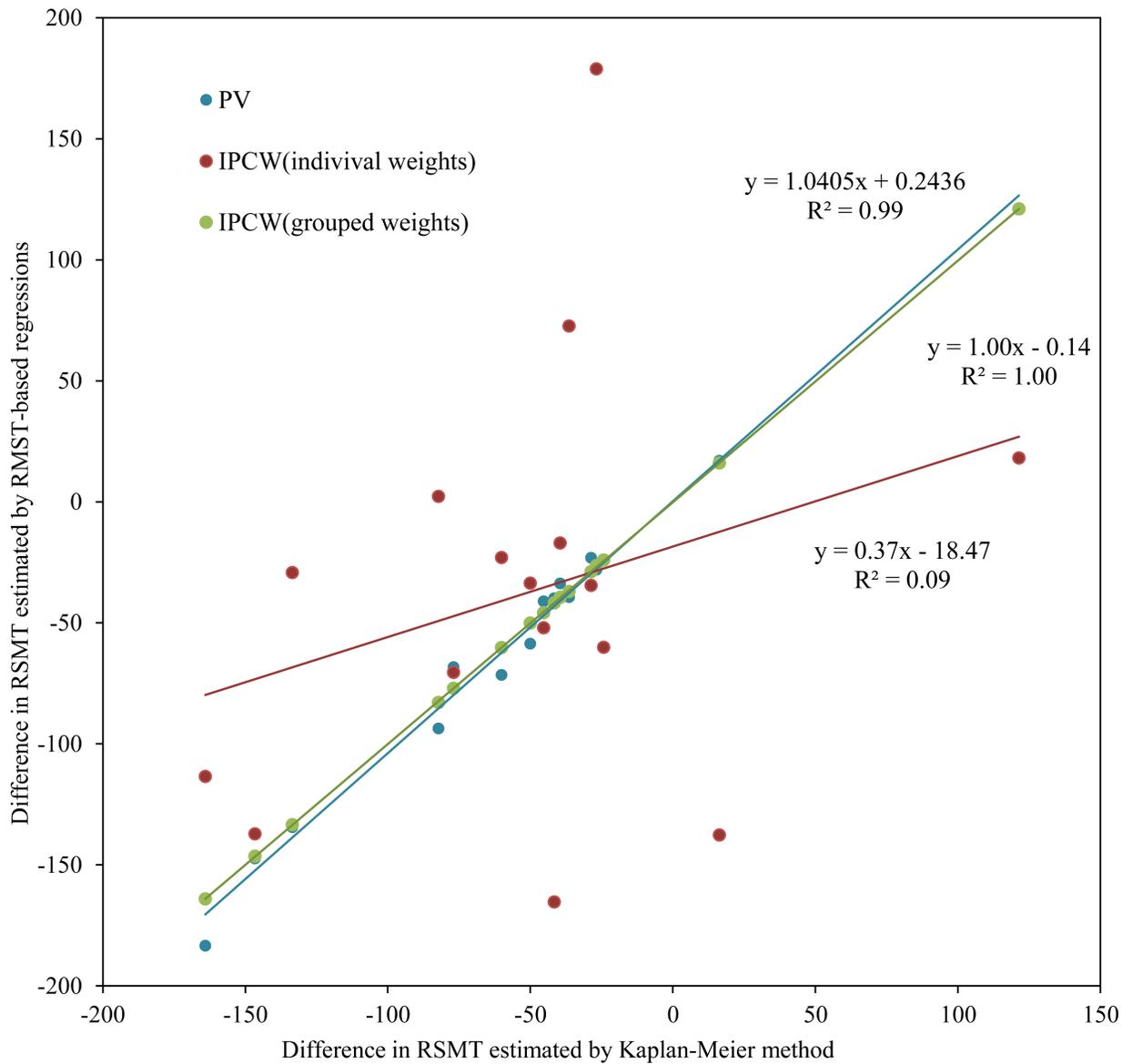
## Figures

### Survival data (Right-censored time-event data)



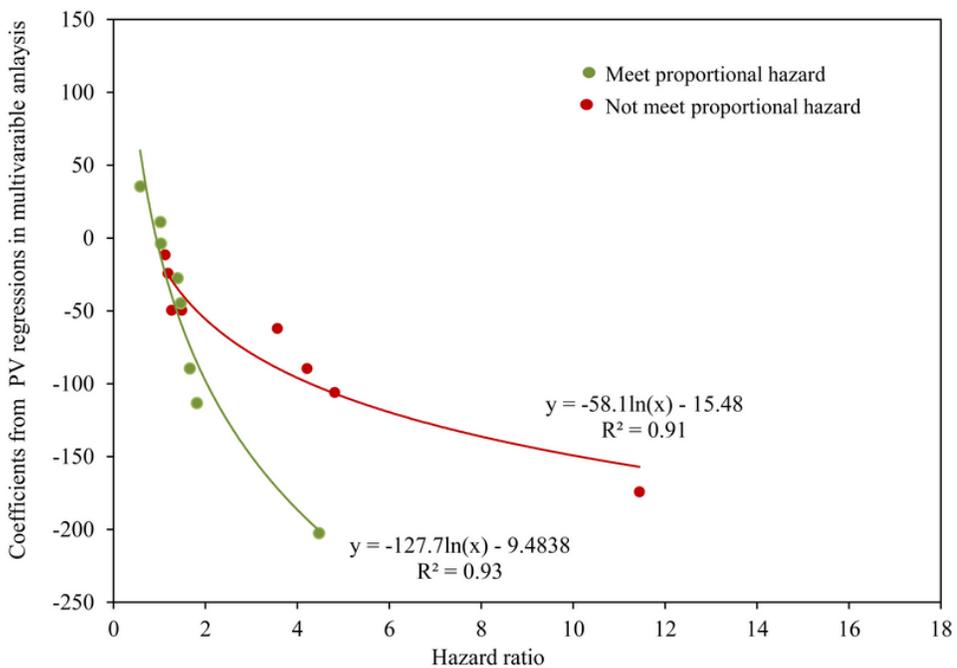
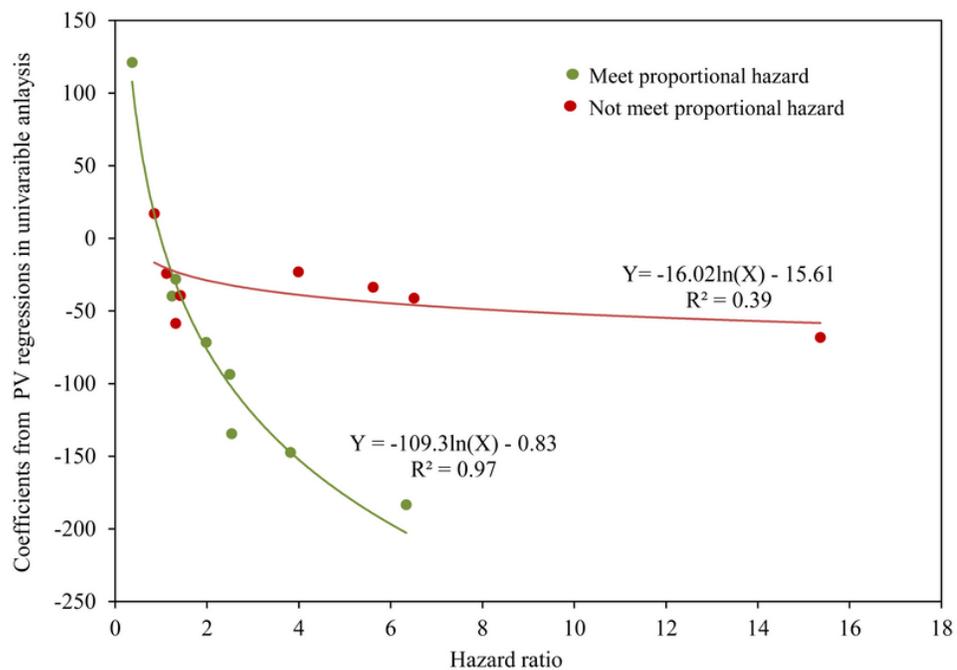
**Figure 1**

flowchart combined conventional survival analysis and restricted mean survival time based regressions.



**Figure 2**

Difference in restricted mean survival time estimated from 3 methods in univariable analyses



**Figure 3**

Relationship between hazard ratios and differences in restricted mean survival time in univariable and multivariable analyses

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GraphicassessmentofPHassumption.pptx](#)