A univariable and multivariable analysis of 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)

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: December 19th, 2019

DOI: <https://doi.org/10.21203/rs.2.18724/v1>

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 univariable analyses, 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 (Difference in RMST = $-109.3 \ln(\text{HR}) - 0.83$, $R^2 = 0.97$, and Difference = $-127.7 \ln(\text{HR}) - 9.49$, $R^2 = 0.93$, respectively.)

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 curves, and then log-rank test or Wilcoxon test is carried out to compare survival curves. Finally, univariable and multivariable Cox proportional hazard 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 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. A number of methods have been developed to assess if data meet the criteria of PH assumption, including graphical analysis (Kaplan Meier curves and estimated survival function with double logarithmic transformation), the use of time-dependent covariates and goodness of fit[4]. When PH assumption is not satisfied, estimated HR is invalid and incorrect conclusion may be achieved [5]. 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[6, 7]. It is estimated through calculating the area under the survival curve between 0 and pre-specific

time (τ). RMST is a simple statistic in the presence of non-proportional hazards and can be interpreted in a clinically meaningful perspective. Difference in RMST presents benefit or harm for the participants. Difference 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 (τ). Generally, unadjusted estimation on difference in RMST is completed adopting Kaplan-Meier plot, while adjusted one via standardized survival curves or covariance analysis. In such situation, only one binary 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[8, 9].

Log-rank test, Wilcoxon test and Cox proportional hazard regression are traditional and common survival analyses, while the estimation of difference in RMST, RMST-based regression 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 RMST when reanalyzing 54 randomized trials [10]. 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 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 non-missing outcome from 1970s to 2000s was captured. Cases with a survival time of zero were excluded from the analysis because of lacking significant information.

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 ≤ 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 mm 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 well-designed flowchart was designed to integrate conventional survival analyses and RMST-based analyses (Fig. 1) and it was divided into difference section and regression section.

There were two key assumptions located in diamonds that need to check before selecting suitable methods, namely proportional hazard (PH) and the homogeneity of censoring mechanism. Proportional hazard assumption was inspected employing two graphics. 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. 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.

Difference analysis was suitable only for categorical variables. Comparing distributions of two or more survival curves was performed firstly. Log-rank test and Wilcoxon test were two common methods. PH assumption was required in log-rank test, but

not for Wilcoxon test[11]. Meanwhile, difference in RMST between 2 or more groups was estimated. Truncation time, τ , was specified utilizing the smallest value of the largest follow-up time across groups minus 5 months.

PH assumption was the first key step to select suitable regression. If it was satisfied, cox proportional hazard model would be applied and HR would be calculated as effect measures. If not, RMST-based regressions would be applied. In general, there are two kinds of RMST-based regression, inverse probability of censoring weighting regression (IPCW) and Pseudo-value regression (PV).

IPCW method calculates weights using Kaplan-Meier technique and includes them into iterative estimation; and it implicitly assumes that all subjects share the same censoring mechanism, namely the homogeneity of censoring mechanism assumption. If this assumption is questionable, it would be more appropriate to adopt group-specified weights, in which censoring mechanism was homogenous in each group and weights were estimated for each group. When performing multivariable regression, a series of covariates would be included, but only one categorical variable can be specified for calculating group-specified weights, so IPCW regression with group-specified weights was limited in multivariable analysis at present.

PV method employs jackknife leave-one-out estimation to generate pseudo-values, and then pseudo-values are used to model the effects of covariates on the outcome of interest through generalized estimating equation. PV method is not restricted by the homogeneity of censoring mechanism assumption.

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 a well-designed flowchart based on statistical knowledge (Fig. 1).

For comparing results among these methods, all methods were conducted ignoring the two assumptions in difference analyses and regressions. Multiple comparison was employed in difference analyses if there were more than 2 groups. Both univariable and multivariable regressions were conducted for Cox proportional hazard regressions and RMST-based regressions. τ specified in univariable analyses was same as that obtained in difference analysis. In multivariable analyses, τ was fixed at 480 months (24 years).

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 $\sim \log_e(\text{HR})$).

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 $P < .05$. Difference analysis was 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. 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 100 mm (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.

Table 1

Basic characteristics of osteosarcoma patients from SEER

Covariates	n (%)
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 ~ 1980 s	803(18.23%)
1990s ~ 2000 s	3602(81.77%)
Race	
White and black	4015(91.15%)
others	390(8.85%)
Tumor Size	
< 100 mm	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%)
Outcome	
All-cause death	

Q1, the first quartile, Q3, the third quartile

AJCC, American Joint Committee on Cancer for staging.

Covariates	n (%)
Alive	
Q1, the first quartile, Q3, the third quartile	
AJCC, American Joint Committee on Cancer for staging.	

Assumption tests of proportional hazard and the homogeneity of censoring mechanism of all covariates on 10 covariates were summarized into Table 2. Dichotomous age, diagnosis year, race, the extend of disease, surgery status and radiation status met proportional hazard assumption, while others failed to pass the test. Dichotomous age, sex, race and radiation met the homogeneity of censoring mechanism, and the resting 7 covariates failed, which implied IPCW regression with group-specified weights was preferable to IPCW regression with individual weights.

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				

Overall, all covariates were associated with overall survival in univariable analyses (Table 3). In difference analysis, log-rank test, Wilcoxon test and 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 in overall survival between male and female patients. Difference in RMST, however, claimed male patients would live shorter by 24.31 months than female ones significantly ($p = 0.004$). In regressions, PV regressions exhibited similar effect estimates and conclusion to difference estimation in RMST from Kaplan-Meier method; similar effect estimates of covariates were shown in IPCW regression with group-specified weights, but sex, race and chemotherapy turned to be insignificant, $p = 0.42$ and $p = 0.35$, respectively.

Table 3
univariable analysis based on conventional survival analysis and restricted mean survival time

Covariates	Difference analysis			Regression			
	Survival rate		Difference in	Cox	RMST-based regression		
	Log-rank	Wilcoxon	RMST ^a	regression ^b	PV regression ^a	IPCW regression (individual weights) ^a	IPCW regression (group-specified weight) ^a
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 ~ 1980 s	< 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 ~ 2000 s	-	-	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	< 0.001*	< 0.001*	< 0.001				
> 100 mm	-	-	-50.03, < 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
< 100 mm(reference)	-	-	-	ref	ref	ref	ref

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

^a Presented as point estimation of coefficient (difference in RMST) and corresponding p value

^b Presented as point estimation of hazard ratio and corresponding p value

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

Covariates	Difference analysis			Regression			
	Survival rate		Difference in	Cox	RMST-based regression		
Extend of disease	< 0.001*	< 0.001*	RMST ^a	regression ^b			
Metastasis	-	-	-164.09, < 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	< 0.001*	< 0.001*	< 0.001*				
IV	-	-	-76.92, < 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							
^a Presented as point estimation of coefficient (difference in RMST) and corresponding p value							
^b Presented as point estimation of hazard ratio and corresponding p value							
* Overall test for categorical covariates with more than 2 groups.							

coefficients estimated by IPCW regression with group-specified weights were quite different from those estimated by IPCW regression with individual weights ignoring homogeneity of censoring mechanism, for instance, radiation variable held the homogeneity of censoring mechanism, and estimated coefficient was – 133.39 in IPCW regression with group-specified weights and – 29.15 in IPCW regression with individual weights. Surgery variable broke the assumption and estimated coefficients were 121.05 and 18.17, respectively. Two IPCW methods were sensitive to weights and produce unstable estimation.

It was noticeable that difference analysis and IPCW regression with group-specified weights were no longer applicable when age was included as a 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.

In Fig. 2, differences in RMST estimated by Kaplan-Meier method were almost equal to those by PV regressions and IPCW regression with group-specified weights ($Y = 1.04X + 0.24$, $R^2 = 0.99$; $Y = 1.00X - 0.14$, $R^2 = 1.00$, respectively), but were different from those by IPCW regression with individual weights ($Y = 0.37X - 18.47$, $R^2 = 0.09$). In addition, the relationship between HR estimated by COX regressions and coefficients estimated by PV regressions was illustrated in Fig. 3. when covariates met PH assumption, a well-fitted logarithmic relationship was observed (coefficients = $-109.3 \ln(\text{HR}) - 0.8343$, $R^2 = 0.9702$); But if PH assumption was not met, goodness of fit declined markedly (coefficients = $-16.02 \ln(\text{HR}) - 15.06$, $R^2 = 0.39$). Because coefficients from PV regressions agreed with those from IPCW regression with group-specified weights, there is no further plot in univariable analyses.

All covariates were included in multivariable regressions. IPCW regression with group-specified weights was limited. P values were highly consistent between COX regressions and PV regressions, but changed chaotically and greatly in IPCW regression with individual weights (Table 4). For simplification, negative difference in RMST was defined as survival time lost (STL) and positive difference was defined as survival time gain (STG). The overall survival was independently associated with age (HR:1.03, STL: 3.86), sex (HR:1.18, STL:24.23), diagnosis in 1970 ~ 1980s (HR: 1.39, STL:27.49), > 100 mm tumor size (HR:1.48, 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 (HR:1.26, STL:49.49).

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 ~ 1980 s	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 ~ 2000 s	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
< 100 mm(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

Parameter	Cox regression		Pseudo-Value regression		IPCW regression (individual weights)	
	HR	p	Beta (95% CI), months	p	Beta (95% CI), months	p
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

The relationship between HR and coefficients estimated by PV regressions were also fitted. Figure 4 showed that the slope of logarithmic transformed HR on coefficients from PV regressions changed slightly from -109.3 to -128.2 with stably high goodness of fit ($R^2 = 0.92$) when covariates meeting PH assumption, and the slope changed dramatically from -16.02 to -57.92 but obtained increased goodness of fit ($R^2 = 0.91$) when covariates broke PH assumption. However, no significant and stable relationship was observed between HR and coefficients estimated by IPCW regression with individual weights in Fig. 5.

Discussion

Firstly, a flowchart was designed for right-censoring survival data to guide the selection and combination of statistical methods, namely 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 analysis using RMST-based regressions, pseudo-value regression provided a more robust estimation than inverse probability censoring weighting regression with individual weights regardless of the homogeneity of censoring mechanism assumption was met or not; meanwhile, although parameter estimates from IPCW regression with group-specified weights approximated to those from PV regressions, statistical power decreased. In both univariable and multivariable variables, a robust negative logarithmic linear relationship was observed between HRs estimated by Cox regression and coefficients (changes in RMST) by PV regression when proportional hazard assumption was hold.

Survival function was created for lifetime data using a Kaplan-Meier method (product limit estimator), the 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; meanwhile, clinical interpretation of HR may not be straightforward for clinicians[12]. Thereby, RMST shows several inferential and clinical advantages over other statistics in survival analysis. RMST has gradually become an essential effect measure in survival analysis. 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[13]. 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[14]. In addition to original studies, RMST and difference in RMST have been considered as effect measures in meta-analysis[15, 16]. RMST was regarded as an obligatory end point, while HR and difference in RMST as complementary methods to summarize treatment effect in oncological trials[17]. In conclusion, RMST and difference in RMST have been accepted by many researchers and widely used.

The estimation of difference in RMST often focuses on only one binary independent variable, such as treatment arms (experimental group and control group). 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, so confounders-adjusted difference in RMST is essential. Furthermore, precision and statistical power can be improved after adjusting prognostic covariates when compared to RMST[18]. If only one categorical variable is interested and all other variables are potential confounders for adjustment, a standardized survival curve taking account of confounders can be drawn for categorical variable and then adjusted difference in RMST can be acquired. To

study the effects of multiple factors on RMST, RMST-based regressions would be appropriate. Regression methods can measure the effects of continuous variables, not limited to categorical variables. Moreover, varied regression-based methods can be adopted, such as non-linear fitting, interaction effect and subgroup analysis. Compared to Cox proportional hazard regression, RMST-based regression could model RMST directly and facilitate model-based inference.

Two types of direct modeling of RMST are developed using generalized linear modeling technique. Pseudo-value regression was proposed by Anderson in 2003 to deal with time-to-event data and to make an inference[9]. In this method, the pseudo-value of RMST was firstly calculated using jackknife leave-one-out method for each observation, and then a generalized estimating equation method was used to estimate the regression coefficients of covariates using the previous pseudo-values of RMST as dependent variables. Codes for pseudo-value method are available for three main platforms (SAS, R and Stata) [19, 20]. Another RMST-based regression is based on inverse probability censoring weighting technique proposed by Tian in 2014[8]. IPCW regression models the probability of being censored as weights with given covariates and assigns the weight calculated by inverse of probability to each observation when solving the regression coefficients. Fortunately, both PV regression and IPCW regression on RMST have been integrated into RMSTREG procedure in the latest version of SAS9.4[21]. In univariable analyses of the current research, the estimated difference of RMST was identical between Kaplan-Meier method, PV regressions and IPCW regression with grouped weights. It is impossible to specify more than one group to estimate group-specific weights, and only IPCW regression with individual weights is available in multivariable analysis when more than one covariates violate the homogeneity of censoring mechanism assumption. The estimated coefficients, however, were quite different from those obtained from PV regressions in multivariable analyses. IPCW regression with grouped weights is limited to categorical variables, and testing the homogeneity of censoring mechanism is not suitable for continuous variables. In addition, inverse probability may create extreme weight when the probability is close to 0, and small sample size will make inverse probability instable[22]. 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 [23]. 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 τ with RMST has been illustrated in details[24].

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 no matter PH assumption is met or not. 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, such as parametric survival analyses and survival analyses with interval-censored data, which 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. PV regression can be a preferred method for RMST-based regressions.

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:
 - XT. Z: study design and literature review.
 - QH: study design, data analysis and writing.
 - JL: data collection and interpretation.
 - BH.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 Huang¹, Jun Lyu², Bing-hui Li¹, Lin-lu Ma¹, Tong Deng³, Xian-Tao Zeng^{1,4*}

* 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

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

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

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

⁴Department 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. 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.
5. Hernan MA: **The hazards of hazard ratios.** *EPIDEMIOLOGY* 2010, **21**(1):13-15.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. C. RLM, Naranjo MD: **A pretest for choosing between logrank and wilcoxon tests in the two-sample problem.** *METRON* 2010, **68**:111-125.
12. 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.
13. 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.

14. 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.
15. 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.
16. 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.
17. 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.
18. Karrison T, Kocherginsky M: **Restricted mean survival time: Does covariate adjustment improve precision in randomized clinical trials?** *CLIN TRIALS* 2018, **15**(2):178-188.
19. 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.
20. Parner ET, Andersen PK: **Regression Analysis of Censored Data Using Pseudo-observations.** *The Stata Journal* 2010, **10**(3):408-422.
21. SAS Institute Inc. 2018. SAS/STAT® 15.1 User's Guide. Cary, NC: SAS Institute Inc. Available online: <https://support.sas.com/documentation/onlinedoc/stat/examples/151/index.html>.
22. 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.
23. 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.
24. 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.

Figures

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 analyses

Figure 4

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