Annals of Epidemiology 64 (2021) 149–154

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original Article

Stratified Restricted Mean Survival Time Model for Marginal Causal Effect in Observational Survival Data

Ai Ni, PhD, Zihan Lin, MS, Bo Lu, PhD*

The Ohio State University College of Public Health, Columbus, OH

ARTICLE INFO

Article history:
Received 16 November 2020
Revised 25 September 2021
Accepted 27 September 2021
Available online 4 October 2021

Keywords:
Marginal effect
Noncollapsibility bias
Confounding bias
Restricted mean survival time
Propensity Score Stratification

ABSTRACT

Time to event outcomes is commonly encountered in epidemiologic research. Multiple papers have discussed the inadequacy of using the hazard ratio as a causal effect measure due to its noncollapsibility and the time-varying nature. In this paper, we further clarified that the hazard ratio might be used as a conditional causal effect measure, but it is generally not a valid marginal effect measure, even under randomized design. We proposed to use the restricted mean survival time (RMST) difference as a causal effect measure, since it essentially measures the mean difference over a specified time horizon and has a simple interpretation as the area under survival curves. For observational studies, propensity score adjustment can be implemented with RMST estimation to remove observed confounding bias. We proposed a propensity score stratified RMST estimation strategy, which performs well in our simulation evaluation and is relatively easy to implement for epidemiologists in practice. Our stratified RMST estimation includes two different versions of implementation, depending on whether researchers want to involve regression modeling adjustment, which provides a powerful tool to examine the marginal causal effect with observational survival data.

Introduction

In epidemiologic studies, time to a certain clinical or health event is commonly used as an outcome measure (often referred to as survival outcome) [1–3]. Investigating the causal relationship between an intervention and the survival outcome is an important topic, with either randomized or observational designs. Hazard ratio (HR) [4] is often used to measure the association between two groups, which can be interpreted as the instantaneous risk ratio. In the epidemiologic literature, the Cox proportional hazards (pH) model is routinely used as a regression adjustment tool for controlling confounding in survival data [5,6]. Despite its popularity, HR is not an appropriate measure of causal effect in many situations. It suffers from the noncollapsibility bias when the treatment has some effect, is likely to change over time (non-proportional hazards) and has built-in selection bias [7–9].

To overcome drawbacks of HR as a causal effect measure, the restricted mean survival time (RMST) difference was proposed [4]. The RMST difference is defined as the integral of the survival probability difference, which is considered as risk difference. Since risk difference is collapsible [10] and integration is a linear operation that preserves the collapsibility of the integrand, the RMST difference is collapsible and therefore provides a valid marginal causal effect interpretation. It is easy to use in randomized trials and provides clinically valuable information on the relative difference between two groups. For observational studies, there was only limited literature on this topic that incorporated propensity score weighting into RMST estimation [11,12]. To facilitate the use of RMST difference in epidemiology research, we proposed an easier propensity score stratification adjustment to remove observed confounding bias. We also conducted a simulation study to examine the statistical properties of our proposed method under different scenarios. The propensity score stratified RMST method was applied to the Atherosclerosis Risk in Communities (ARIC) study [13], examining the causal effect of smoking on risk of stroke, to illustrate its practical utility.

Abbreviations: HR, hazard ratio; Ph, proportional hazards; RMST, restricted mean survival time; PATE, population average treatment effect.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

* Corresponding author. The Ohio State University College of Public Health, Columbus, OH, Tel: 6142477913, Fax: 6142923572.

E-mail address: lu.232@osu.edu (B. Lu).

https://doi.org/10.1016/j.annepidem.2021.09.016
1047-2797/

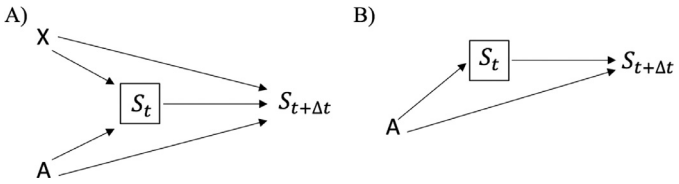

Fig. 1. (A) Randomized trials with prognostic factor. (B) Randomized trials without prognostic factor.

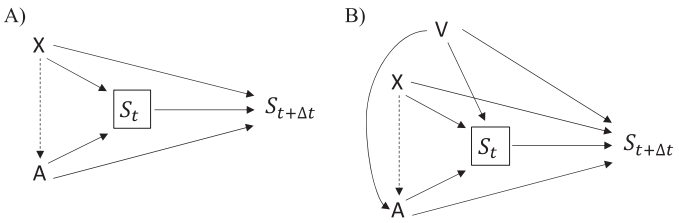

Fig. 2. (A) Observational studies with observed confounders only. (B) Observational studies with both observed and unobserved confounders. Dashed arrow indicates that observed confounding is controlled by appropriate adjustment.

Method

HR is not a valid marginal causal effect measure

Under the potential outcome framework [14], with a dichotomous treatment, each individual has a pair of potential outcomes, that is, Y^1 as the outcome under treatment and Y^0 as the outcome under control. In epidemiologic research, population average treatment effect (PATE) is of more interest, for example, $E(Y^1 - Y^0)$. PATE can be estimated in randomized studies without strong assumptions and can be estimated in observation data with appropriate adjustment under the ignorable treatment assignment assumption [14]. It is also referred to as marginal effect, with the interpretation of comparing mean outcomes for the entire study population by assigning the treatment to everyone versus withholding the treatment from everyone. Another common effect measure is conditional effect, usually resulting from using regression models to control confounding.

Hernan et al. [8] pointed out that the noncollapsibility of HR is a result of selection bias, which distorts the relationship between two variables when conditioning on their common descendent. Figure 1A depicts a randomized trial with treatment indicator A and a covariate X. The outcomes are two binary survival statuses, S_t and $S_{t+\Delta t}$, at time points t and $t+\Delta t$, respectively. X is a prognostic factor, which is related to survival outcomes regardless of treatment assignment (arrows from X to S_t and $S_{t+\Delta t}$). There is no arrow between X and A, since the study is randomized. The arrows from A to S_t and $S_{t+\Delta t}$ imply that the treatment has some effect. By definition, the HR at $t+\Delta t$ is conditioned on S_t, $HR(t) = \lim_{\Delta t \to 0} \frac{\Pr(S_{t+\Delta t}=0|S_t=1)}{\Delta t}$. Such conditioning (illustrated as a box surrounding S_t) introduces selection bias as S_t is the common descendent of both A and X. The situation when such selection bias is absent is depicted in Figure 1B, where there is no prognostic factor. This is, however, very unlikely to be true practically, as the survival outcome is expected to be related to some factors, either observed or unobserved.

There is one way to remove this selection bias by conditioning on X, which blocks the pathway $A \to S_t \leftarrow X \to S_{t+\Delta t}$. But this yields a conditional effect as one needs to include X in the Cox pH model. Due to the noncollapsibility of HR, the marginal effect does not equal the conditional effect when the treatment has an effect. Therefore, HR is not a measure with a valid marginal causal effect interpretation, except for some extreme cases.

Figure 2 presents an observational study scenario, with X being an observed confounder and V being an unobserved confounder. Figure 2A depicts the case with no unobserved confounder. Appropriate adjustment has been applied to remove confounding, shown as a dashed arrow between X and A. Since X is related to the survival outcome, the selection bias cannot be removed without conditioning on X. Hence, the resulting HR represents a conditional effect, just like the randomized scenario. Figure 2B depicts the unfortunate case with an unobserved confounder, where there is no way to remove the selection bias.

As HR is not a valid marginal effect measure for survival outcome, we introduce RMST difference as an alternative measure, which is essentially a mean difference over a specified time horizon. RMST estimation for randomized trials has been discussed substantially [4,15,16]. We will further extend it to observational studies by combining RMST with propensity score stratification to remove confounding.

Restricted mean survival time

RMST is defined as the mean survival time truncated at a fixed time point and is equivalent to the area under the survival curve up to the truncation time point. Treatment effect measured by the RMST difference can be asymptotically unbiasedly estimated without the pH assumption. RMST difference also offers a more intuitive interpretation than HR. Moreover, this effect measure is collapsible, which makes it an attractive choice for marginal causal effect estimation [17,18].

For a given subject, denote the event time as T and the restricted event time as $U = \min(T, \tau)$, where τ is the fixed truncation time point, usually prespecified at the design stage based on clinical relevance and study feasibility. Both T and U are subject to censoring by a random variable C that is assumed to be independent of T and U conditional on covariates. The observable quantities are $(T^*,\ \Delta_T,\ U^*,\ \Delta_U)$, where $T^* = \min(T, C)$, $\Delta_T = I(T < C)$, $U^* = \min(U, C)$, and $\Delta_U = I(U < C)$. The RMST is defined as $\mu(\tau) = E(U) = \int_0^\tau S(t)dt$, where $S(t)$ is the survival function of T.

Existing RMST estimation methods

A natural estimator of $\mu(\tau)$ is $\hat{\mu}(\tau) = \int_0^\tau \hat{S}(t)dt$, where $\hat{S}(t)$ can be estimated by the nonparametric Kaplan-Meier method [19] or parametric methods assuming a distribution function for T [20]. In this paper we estimated $\hat{S}(t)$ by the Kaplan-Meier method due to its robustness and popularity in epidemiological research [21].

Several methods of regressing RMST on covariates, to control for confounding or increase estimation efficiency, have been developed [22–24]. Karrison [22] considered RMST regression under a piecewise exponential model. Andersen et al. [23] devised a pseudo-outcome that converts censored survival time to uncensored pseudo-survival time and then used a generalized linear model on the pseudo-survival time to estimate RMST. Tian et al. [24] proposed an inverse probability of censoring weighted (IPCW) estimating function to estimate RMST. Wang and Schaubel [25] developed an estimating function to estimate RMST under general censoring mechanisms. We will use the IPCW estimating function with identity link [24] as a comparator to our proposed method in the simulation studies due to its ease of implementation and popularity. Let Z_i be a p-dimensional covariate vector (including treatment indicator A_i) of subject i with $i = 1, \ldots, n$, where n is the sample size. The regression coefficients β can be consistently estimated by solving the estimating equation

$$\Psi_n(\beta) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_{Ui}}{\hat{G}(U_i^*)} Z_i \left(U_i^* - \beta^T Z_i\right) = 0,$$

where β^T denotes transpose of vector β and $\hat{G}(\cdot)$ is the Kaplan-Meier estimator of the censoring time C based on $\{(U_i^*,\ 1-\Delta_U),\ i=1,\ldots,n\}$. The estimated regression coefficient of treatment, $\hat{\beta}_A$, represents the conditional treatment effect on RMST conditioning on covariates. Since RMST difference is collapsible, the conditional treatment effect equals the marginal treatment effect. The variance of $\hat{\beta}_A$ can be derived as a sandwich-type estimator [24].

Proposed propensity score stratified rmst estimation in observational studies

Propensity score adjustment is a popular strategy to removing observed confounding in non-randomized studies [26,27]. Stratification achieves a good balance between the technical complexity and easiness of understanding and implementation [14]. By splitting the sample into several propensity score strata, the covariate distribution becomes nearly identical between control and treatment group within each stratum. Therefore, the strong ignorability condition approximately holds and stratum-specific causal effects can be unbiasedly estimated. As RMST difference is collapsible, the stratum-specific effects can be pooled by weighted average to obtain the marginal causal effect estimate. Specifically, it involves the following steps:

Propensity score estimation

The propensity score is defined as the conditional probability of treatment $A=1$ given a vector of observed covariates X [26]. We estimated the propensity score by fitting a logistic regression on treatment indicator A with covariates X following the common practice, although other estimation options are available [28,29].

Propensity score stratification

Depending on the sample size, we may stratify the sample into five or ten strata [30]. Since our real data example has a large sample size, we chose to use ten equal-sized strata based on deciles of the propensity score distribution. To ensure asymptotically unbiased RMST estimates and nonzero variance estimates, τ must lie between the minimum event time and maximum follow-up time (event or censored) of each treatment group in each stratum [25,31]. When this requirement is not met for a given stratum, it should be merged to the adjacent stratum. A more reasonable value of τ should be considered if this requirement is still not met when the number of strata becomes less than five after merging.

Treatment effect estimation

The marginal treatment effect is estimated as a weighted average of stratum-specific treatment effects weighting by stratum size. Let L be the number of strata and n_l be the size of stratum l with $l=1,\ldots,L$. The total sample size $n=\sum_{l=1}^{L} n_l$. The stratum-specific treatment effect is estimated by $\hat{v}_l=\int_0^\tau(\hat{S}_{l1}(t)-\hat{S}_{l0}(t))dt$, where $\hat{S}_{l1}(t)$ and $\hat{S}_{l0}(t)$ are the Kaplan-Meier estimates of survival function for treatment and control groups in stratum l, respectively. The marginal treatment effect is calculated by $\hat{v}=\sum_{l=1}^{L}\Pr(l)\hat{v}_l=\sum_{l=1}^{L}(n_l/n)\hat{v}_l$. Since strata are independent of each other, the variance of $\hat{v}$ is estimated by $\hat{V}(\hat{v})=\sum_{l=1}^{L}(n_l/n)^2\hat{V}(\hat{v}_l)$, where $\hat{V}(\hat{v}_l)$ is the stratum-specific variance [21]. A possible improvement of this stratified adjustment is to include regression modeling, where we could use strata as a categorical covariate in an IPCW estimating equation [24] with identity link function. Unlike the above stratified estimation where treatment effect can vary across strata, this method assumes that treatment effect is the same in all strata. The benefit here is the potential gain in variance estimation due to the use of regression models.

Simulation study

Our simulation study consists of two parts—the first part concerns with a randomized design with no confounding and the second part concerns with an observational design with known confounders and other covariates.

Randomized assignment

A common misconception is that, due to the randomization, a simple Cox pH model with only treatment indicator A provides an unbiased estimate of the marginal treatment effect. In Figure A1 in Appendix A, we demonstrated that, under a true Cox pH model with treatment A and an independent covariate X, the marginal HR of treatment marginalizing over X is non-proportional, unless there is no treatment effect. Thus, the marginal HR of treatment cannot be summarized as a single value. The derivation of marginal HR is included in Appendix A.

Our simulation generated the treatment indicator A from Bernoulli distribution, an independent prognostic factor X from *N(0,1)*, potential survival times (T^1,T^0) from Weibull distribution [32], and truncated potential survival times $U^j=\min(T^j,\tau),\ j=0,1$. The censoring variable C was generated from an exponential distribution with rate parameter 0.45 and the truncation time τ was set to 2.67, the 70th percentile of the censoring distribution. Detailed description of the simulation setup is included in Appendix B. Five hundred datasets of sample size 500 were simulated.

We considered four simulation scenarios: proportional hazards with no treatment effect (referred to as pH_*null*), proportional hazards with some effect (pH_*alt*), non-proportional hazards with no effect (*NPH_null*), and non-proportional hazards with some effect (*NPH_alt*). We compared three estimation methods: Cox regression with only A, Cox regression with A and X, and RMST difference between treatment groups without adjusting for X. The performance metrics include the average of the 500-point estimates of treatment effect, model-based standard error (SEM) which was the average of the 500 estimated standard errors, empirical standard error (SEE) which was the standard deviation of the 500-point estimates, and 95% coverage probability (CP) which was the proportion of the 500 95% confidence intervals that covered the true value.

As shown in Table 1, under pH_*null*, all three methods generate almost unbiased estimates of their target parameters and adequate coverage probabilities. Under pH_*alt*, the marginal HR is a non-constant function of time as illustrated in Appendix Figure A1, and therefore cannot be estimated correctly by either of the two Cox models. The Cox model with A and X unbiasedly estimates β_A, the conditional HR of A, whereas the Cox model with only A cannot even estimate β_A unbiasedly as shown by the large bias and poor coverage probability. On the other hand, RMST difference gives unbiased estimates of the marginal treatment effect on RMST. Under *nonPH_null* and *nonPH_alt*, the Cox model with A and X gives biased estimates of β_A since it misspecifies the true non-proportional model as proportional. RMST difference still provides almost unbiased estimates of the marginal treatment effect and adequate coverage probability.

Non-randomized assignment

The second simulation considers an observational setup with four estimation methods. Since the first stimulation demonstrated that the Cox model on HR is not valid for marginal treatment effect estimation, we did not further consider it here.

Table 1
Performance of Estimation Methods for Marginal Treatment Effect Under Randomized Treatment Assignment. (500 Simulated Datasets with $n = 500$ Each). True TE: true treatment effect. Estimated TE: the average of the 500 estimated treatment effects. SEM: the average of the 500 model-based standard error estimates. SEE: the empirical standard error as the standard deviation of the 500 TE estimates. CP: 95% confidence interval coverage probability

Scenarios	Method	True TE*	Estimated TE	SEM	SEE	CP
pH_null						
	Cox with A	0	0.00	0.12	0.12	0.95
	Cox with A and X	0	−0.01	0.12	0.12	0.95
	RMST difference	0	0.00	0.088	0.087	0.96
pH_alt						
	Cox with A	–	−0.36	0.13	0.13	0.79
	Cox with A and X	−0.5	−0.51	0.13	0.13	0.95
	RMST difference	0.25	0.25	0.090	0.090	0.95
nonPH_null						
	Cox with A	–	−0.039	0.13	0.12	0.94
	Cox with A and X	0	−0.14	0.13	0.14	0.79
	RMST difference	0.087	0.086	0.077	0.077	0.96
nonPH_alt						
	Cox with A	–	−0.28	0.13	0.13	0.60
	Cox with A and X	−0.5	−0.47	0.13	0.14	0.94
	RMST difference	0.27	0.27	0.079	0.079	0.95

* The true treatment effect (TE) for Cox regression with only A is the marginal HR of A that varies over time and therefore cannot be represented by a single value except when the true conditional log-HR of A is zero; the true TE for Cox regression with A and X is the conditional log-HR of A; the true TE for RMST difference is the marginal treatment effect on RMST.

Propensity score stratified estimation

This is the proposed method described in the previous section. Samples were stratified into ten equal-sized strata based on the deciles of the propensity score distribution. RMST difference was estimated nonparametrically in each stratum separately and then pooled by weighted average to obtain the marginal treatment effect estimation.

Propensity score strata adjusted regression estimation

This is the second way of using propensity score strata, as described in the previous section. Ten strata are included as a categorical variable with ten levels in the IPCW estimating equation with a linear model on RMST.

Confounder-adjusted regression

We again used the IPCW estimating equation to estimate treatment effect but included individual confounders instead of the propensity score strata indicators as covariates in the regression model on RMST.

Crude comparison

RMST was directly compared between control and treated groups without adjusting for confounders.

We generated nine independent covariates from either normal or Bernoulli distribution. Potential survival times (T^1, T^0) were generated from Weibull distributions in a similar way as in the first part of simulation. We considered two data-generating models, one using the linear form of covariates and the other using various nonlinear functional forms of covariates (details in Appendix B).

The treatment status A was generated from Bernoulli distribution with $P(A = 1)$ defined by a logit model

$$\text{logit}(P(A = 1)) = -1.7 + 0.5X_3 + 0.25X_4 - 0.25X_5 + 0.5X_6 - 0.5X_7 + 0.5X_8 + 0.5X_9.$$

The sample size was set to 2000. The proportion of treated subjects in the sample was around 20%. The censoring variable C was generated depending on X_5 and X_6. Truncation time τ were determined similarly as before. In this setting, covariate X_3, X_4, X_5, and X_6 were confounders. We considered the same four simulation scenarios as in the randomized case. Five hundred datasets were simulated for each combination of data-generating model and scenario.

The simulation results of the first data-generating model with only linear covariates are summarized in Table 2. Both propensity score stratification and propensity score strata adjusted regression give virtually unbiased estimates of marginal treatment effect on RMST and adequate coverage probabilities in all four scenarios. The latter method provides smaller estimated variance than the former method, which reflects the advantage of regression modeling over nonparametric estimation. Confounder-adjusted regression method exhibits some degree of bias and unsatisfactory coverage probabilities. The crude estimation method gives highly biased estimates and poor coverage probabilities across all scenarios. Similar findings are observed with the second data-generating model (results summarized in Appendix B).

Real data analysis

The ARIC study [13] is an ongoing large prospective cohort study in four U.S. communities. One of its major themes is to investigate risk factors of stroke [33]. We applied the four methods compared in non-randomized simulation to estimate the marginal causal effect of baseline smoking status on the incident ischemic stroke measured as 20-year RMST. The analysis results are summarized in Table 3. The pattern of the results of the four methods is consistent with the simulation results. Based on the propensity score stratified analysis, the average ischemic stroke-free time over 20 years is 4.5 months (95% confidence interval: 3.0, 6.0 months) shorter among current smokers compared to that among non-current smokers. Details of the ARIC study and our analysis were relegated to Appendix C.

Discussion

In this paper, we clarified the utility of HR as a causal effect measure, which is generally not a valid marginal effect measure even under randomized designs. Alternatively, RMST difference provides a valid marginal causal effect measure for survival

Table 2
Performance of Estimation Methods for Marginal RMST Treatment Effect under the Data-generating Model with Linear Covariates (500 Simulated Datasets with $n = 2000$ Each). True TE: true treatment effect. Estimated TE: the average of the 500 estimated treatment effects. SEM: the average of the 500 model-based standard error estimates. SEE: the empirical standard error as the standard deviation of the 500 TE estimates. CP: 95% CI coverage probability

Scenario	Method	True TE	Estimated TE	SEM	SEE	CP
pH_null						
	PS Stratification	0.00	−0.02	0.15	0.14	0.95
	PS Strata Regression		−0.08	0.12	0.10	0.93
	Confounder Adjusted		−0.11	0.12	0.11	0.87
	Crude		−0.52	0.11	0.11	0.006
pH_alt						
	PS Stratification	0.42	0.40	0.16	0.14	0.95
	PS Strata Regression		0.34	0.12	0.11	0.92
	Confounder Adjusted		0.31	0.12	0.11	0.88
	Crude		−0.11	0.12	0.12	0.008
nonPH_null						
	PS Stratification	0.13	0.12	0.12	0.11	0.96
	PS Strata Regression		0.15	0.10	0.087	0.97
	Confounder Adjusted		0.17	0.063	0.063	0.91
	Crude		−0.27	0.090	0.087	0.006
nonPH_alt						
	PS Stratification	0.43	0.43	0.13	0.12	0.95
	PS Strata Regression		0.45	0.10	0.091	0.97
	Confounder Adjusted		0.45	0.066	0.065	0.94
	Crude		0.021	0.096	0.092	0.008

Table 3
Estimated effect of smoking status (current smoker vs nonsmoker) on 20-y RMST of incident ischemic stroke in ARIC study

	Estimated difference in RMST in months	Estimated SE	95% CI (Lower Bound)	95% CI (Upper Bound)
Crude	−4.19	0.68	−5.52	−2.85
PS stratified	−4.51	0.75	−6.00	−3.03
PS strata-adjusted	−4.48	0.70	−5.85	−3.11
Covariate-adjusted	−5.19	0.67	−6.51	−3.87

outcomes. It is more advantageous because (i) it does not change over time; (ii) it is collapsible; (iii) it does not depend on proportional hazards assumption. We have seen an increased acceptance of RMST measures in randomized clinical studies. However, the use of RMST in observational epidemiologic studies is still limited.

One reason for the limited use of RMST in epidemiologic research may be that its interpretation is quite different from that of a hazard. It must be interpreted in the context of a specified time horizon. However, life expectancy over a time horizon τ, such as 10-year life expectancy, may be more intuitive to patients, clinicians, and epidemiologists [16,34–36]. Another reason for its limited use in epidemiology may be due to the lack of methodology development, which requires additional confounding adjustment. Only a couple of papers discussed incorporating propensity score weighting adjustment into RMST estimation [11,12]. In this article, we propose to estimate covariate adjusted RMST difference by the popular propensity score stratification method, which is easily understood by epidemiologists, and thereby promote the use of RMST in epidemiologic research. Nevertheless, more propensity score-based adjustment methods should be developed for epidemiologists to use, including matching-based RMST estimation and more refined stratified RMST methods.

One limitation of our work is that propensity score adjustment is known to only remove confounding due to observed covariates. In many epidemiologic studies, the hidden bias due to unobserved covariates remains a major threat to the validity of the scientific findings. A popular approach to assessing the potential impact of unmeasured confounding is via sensitivity analysis [37]. There is only limited discussion on sensitivity analysis comparing survival curves and it would be desirable to develop sensitivity analysis strategies for RMST estimators [38].

Acknowledgments

This work was partially supported by grant DMS-2015552 from National Science Foundation. This work was also partially supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Grant Number UL1TR002733. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation. The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700005I, HHSN268201700004I). The authors thank the staff and participants of the ARIC study for their important contributions.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.annepidem.2021.09.016.

References

[1] Hernán MÁ, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000:561–70.
[2] Sarfati D, Blakely T, Pearce N. Measuring cancer survival in populations: relative survival vs cancer-specific survival. Int J Epidemiol 2010;39(2):598–610.
[3] Ahmad T, Munir A, Bhatti SH, Aftab M, Raza MA. Survival analysis of heart failure patients: a case study. PLoS ONE 2017;12(7):e0181001.

[4] Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. J Clin Oncol 2014;32(22):2380.
[5] Karim ME, Gustafson P, Petkau J, Zhao Y, Shirani A, Kingwell E, et al. Marginal structural cox models for estimating the association between β-Interferon exposure and disease progression in a multiple Sclerosis Cohort. Am. J. Epidemiol 2014;180(2):160–71.
[6] Carslake D, Davey Smith G, Gunnell D, Davies N, Nilsen TIL, Romundstad P. Confounding by ill health in the observed association between BMI and mortality: evidence from the HUNT Study using offspring BMI as an instrument. Int J Epidemiol 2017;47(3):760–70.
[7] Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. Stat Sci 1999;14:29–46.
[8] Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004;11:615–25.
[9] Hernán MA. The hazards of hazard ratios. Epidemiology (Cambridge, Mass) 2010;21(1):13.
[10] Huitfeldt A, Stensrud MJ, Suzuki E. On the collapsibility of measures of effect in the counterfactual causal framework. Emerg Themes Epidemiol 2019;16(1):1–5.
[11] Zhang M, Schaubel DE. Double-robust semiparametric estimator for differences in restricted mean lifetimes in observational studies. Biometrics 2012;68(4):999–1009.
[12] Conner SC, Sullivan LM, Benjamin EJ, LaValley MP, Galea S, Trinquart L. Adjusted restricted mean survival times in observational studies. Stat Med 2019;38(20):3832–60.
[13] The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol 1989;129(4):687–702.
[14] Imbens GW, Rubin DB. Causal inference in statistics, social, and biomedical sciences. Cambridge University Press; 2015.
[15] 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(1):152.
[16] Kim DH, Uno H, Wei LJ. Restricted mean survival time as a measure to interpret clinical trial results. JAMA Cardiol 2017;2(11):1179–80.
[17] Chen PY, Tsiatis AA. Causal inference on the difference of the restricted mean lifetime between two groups. Biometrics 2001;57(4):1030–8.
[18] Stensrud MJ, Aalen JM, Aalen OO, Valberg M. Limitations of hazard ratios in clinical trials. Eur. Heart J 2019;40:1378–83.
[19] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53(282):457–81.
[20] Nemes S, Bülow E, Gustavsson A. A brief overview of restricted mean survival time estimators and associated variances. Stats 2020;3(2):107–19.
[21] Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. Springer Science & Business Media; 2006.
[22] Karrison T. Restricted mean life with adjustment for covariates. J Am Stat Assoc 1987;82(400):1169–76.
[23] Andersen PK, Hansen MG, Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations. Lifetime Data Anal 2004;10(4):335–50.
[24] Tian L, Zhao L, Wei L. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. Biostatistics 2014;15(2):222–33.
[25] Wang X, Schaubel DE. Modeling restricted mean survival time under general censoring mechanisms. Lifetime Data Anal 2018;24(1):176–99.
[26] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41–55.
[27] Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics 2005;61(4):962–73.
[28] McCaffrey DF, Ridgeway G, Morral AR. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. Psychol Methods 2004;9(4):403.
[29] Westreich D, Lessler J, Funk MJ. Propensity score estimation: machine learning and classification methods as alternatives to logistic regression. J Clin Epidemiol 2010;63(8):826.
[30] Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1984;79(387):516–24.
[31] Tian L, Jin H, Uno H, Lu Y, Huang B, Anderson KM, et al. On the empirical choice of the time window for restricted mean survival time. Biometrics 2020;76(4):1157–66.
[32] Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. Stat Med 2005;24(11):1713–23.
[33] Ni A, Cai J. A regularized variable selection procedure in additive hazards model with stratified case-cohort design. Lifetime Data Anal 2018;24(3):443–63.
[34] Saad ED, Zalcberg JR, Péron J, Coart E, Burzykowski T, Buyse M. Understanding and communicating measures of treatment effect on survival: can we do better? JNCI 2018;110(3):232–40.
[35] 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):1–10.
[36] Kloecker DE, Davies MJ, Khunti K, Zaccardi F. Uses and limitations of the restricted mean survival time: illustrative examples from cardiovascular outcomes and mortality trials in type 2 diabetes. Ann. Intern. Med 2020;172(8):541–52.
[37] Rosenbaum PR. Design of observational studies. New York: Springer; 2010.
[38] Lu B, Cai D, Tong X. Testing causal effects in observational survival data using propensity score matching design. Stat Med 2018;37(11):1846–58.