The Stata Journal (2016)
16, Number 3, pp. 702–716

strmst2 and strmst2pw: New commands to compare survival curves using the restricted mean survival time

Angel Cronin
Dana–Farber Cancer Institute
Boston, MA
AngelM_Cronin@dfci.harvard.edu

Lu Tian
Stanford University
Stanford, CA
lutian@stanford.edu

Hajime Uno
Dana–Farber Cancer Institute
Boston, MA
huno@jimmy.harvard.edu

Abstract. We present `strmst2`, a new command to implement k-sample comparisons using the restricted mean survival time (RMST) as the summary measure of the survival-time distribution. Unlike model-based summary measures such as the hazard ratio, the validity of which relies on the adequacy of the proportional-hazards assumption, the measures based on the RMST (that is, the difference in RMST, the ratio of RMST, and the ratio of the restricted mean time lost) provide more robust and clinically interpretable results about the between-group differences. `strmst2` performs analysis of covariance-type adjusted analyses for k-sample comparisons as well as the corresponding unadjusted analyses. Pairwise comparisons for the adjusted analyses are summarized using the new postestimation command `strmst2pw`. We briefly describe the issues of the hazard ratio, introduce details of the method for the RMST, and then illustrate how to use the new commands.

Keywords: st0451, strmst2, strmst2pw, restricted mean survival time, restricted mean time lost, survival analysis, time-to-event data

1 Introduction

The Cox proportional hazards (PH) model (Cox 1972) has been widely used for survival-data analysis in medical research. Specifically, when the objective of the study is to compare a new intervention with the standard of care, the hazard-ratio estimate is almost routinely used to summarize the difference between the two groups. Cox regression analysis is indeed a great method that is supported by elegant statistical theories (Andersen and Gill 1982; Gill 1984), but there are some issues with using the hazard ratio as the between-group summary measure. Because detailed discussions about the issues of the hazard ratio have been described elsewhere (for example, Hernán [2010], Uno et al. [2014], and Uno et al. [2015]), we briefly summarize the three most important issues below.

The first issue is about the difficulty in interpretation because of the lack of a single summary of the baseline event rate. For example, suppose that we want to compare the survival-time distribution between a new intervention group and a control group and observe a hazard-ratio estimate of 0.8 (new intervention/control). In this simple example, the hazard of the control group is simply the baseline hazard function in the PH regression model. If the baseline hazard function is a constant, then that constant can be estimated and serves as a "reference level" in interpreting the resulting hazard ratio of 0.8. However, in general, the baseline hazard function is not necessarily constant over time. While the PH model can still provide a valid estimate for the hazard ratio in such a case, the interpretation of the hazard ratio is less transparent because there is no simple summary for the hazard function in the control group. Theoretically, the hazard function in the control group can be estimated nonparametrically. However, this requires a large sample size and is difficult to implement and thus is rarely done in practice. In sum, the hazard ratio of 0.8 represents a 20% risk reduction from the hazard of the control group, but the PH regression-based analysis unfortunately does not provide a valid summary for this baseline hazard to assist our interpretation of the hazard ratio.

The second issue is that the precision of the hazard-ratio estimate essentially depends on the number of observed events. Thus, when the event rate is very low, a 95% confidence interval for the hazard ratio will be very wide even when many patients have been followed up for many years. This becomes a critical issue especially in clinical trials aiming to confirm the safety of a new treatment. For example, suppose we randomized 20,000 patients into either the new treatment group or the placebo group to evaluate the noninferiority of the new treatment in terms of safety and followed patients for 10 years. Suppose only one event occurred at year five in each group. Given these data, heuristically, we could reasonably conclude that the new treatment is safe because only one subject out of the 10,000 experienced the event during a long term of follow-up in both groups. With these data, the hazard-ratio estimate is going to be 1. However, the 95% confidence interval of the hazard ratio is wide (0.06 to 16.0) because of the small number of events. Thus we could not conclude the noninferiority of the new treatment with these data, if we summarized the between-group difference by the hazard ratio (Uno et al. 2015).

Note that these two issues are problematic even when the PH assumption is correct. If the PH assumption is violated, it becomes even more problematic to summarize the group difference by the hazard ratio. Nonetheless, the theory shows that even when the PH assumption is not correct, the hazard-ratio estimate converges to a constant when the sample size goes to infinity (Hjort 1992; Lin and Wei 1989); mathematically, this might be one of the most attractive properties of this method. However, the constant value depends not only on the differences in the survival-time distribution between the two groups but also on the underlying study-specific censoring time distribution, which does not contain any information on the survival times. This is the third issue of the hazard-ratio approach. Note that other model-based measures have the same issue when their respective modeling assumption is violated.

In this article, we introduce the between-group contrast measures calculated from the restricted mean survival time (RMST) as alternatives to the hazard ratio. Those summary measures based on the RMST do not require a specific relationship between groups; that is, they are model free, and they do not have the aforementioned issues of the hazard ratio. We introduce a new command, `strmst2`, to make inferences for these measures and use the data from the primary biliary cirrhosis (PBC) study to illustrate `strmst2`.

2 Details of the methods

2.1 The RMST and its inference

We consider right-censored data. Let $S(t) = \Pr(T > t)$ be the survival function of the time-to-event variable T. Let C be the underlying censoring time, which is considered independent of the event time T. Let $\{(T_i, C_i) \mid i = 1, \ldots, n\}$ be the independent and identically distributed copies from (T, C). Because we can observe only either T or C, we denote the observed data by $\{(X_i, \Delta_i) \mid i = 1, \ldots, n\}$, where $X_i = \min(T_i, C_i)$ and Δ_i takes 1 if $T_i \leq C_i$ and 0 otherwise. Also, let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be the ordered statistics of $X_1, X_2, \ldots, X_n$, and let $\Delta_{(i)}$ be the Δ that corresponds to $X_{(i)}$.

The RMST is defined as the area under the curve of the survival function up to a truncation time point $\tau(< \infty)$,

$$\mu = \int_0^\tau S(t)dt$$

where $S(t)$ is the survival function for the time T. Interpreting this quantity is straightforward. For example, suppose τ is equal to five years. Then, μ can be interpreted as 5-year life time expectancy, that is, "what is the life expectancy for the next 5 years" (Royston and Parmar 2011, 2013). We can also interpret μ in the following way—"when we follow up subjects for 5 years in a study, those subjects will survive for μ on average" (Royston and Parmar 2011, 2013). The truncation time τ is considered because it is practically infeasible to follow up the study subjects until all of them experience the event. If there were no censored observations, one could use the mean survival time

$$\mu_\infty = \int_0^\infty S(t)dt$$

instead of the RMST. However, this will not be the case in practice.

A natural estimator for μ is given by

$$\widehat{\mu} = \int_0^\tau \widehat{S}(t)dt$$

where $\widehat{S}(t)$ is the Kaplan–Meier (KM) estimator for $S(t)$, and the asymptotic variance of $\widehat{\mu}$ (Miller 1981) is estimated by

$$\widehat{\mathrm{Var}}\left(\widehat{\mu}\right) = \sum_{i=1}^{n} \left\{ \int_{X_{(i)}}^{\tau} \widehat{S}(u)du \right\}^2 \frac{\Delta_{(i)}}{(n-i)(n-i+1)}$$

The restricted mean time lost (RMTL) is defined as the area "above" the curve of the survival function up to a time τ:

$$\tau - \mu = \int_0^{\tau} \{1 - S(t)\}dt$$

This is estimated by $\tau - \widehat{\mu}$, and the asymptotic variance estimate is identical to that for the RMST.

Note that because the RMTL measures time lost from τ, it may not be reasonable to use the RMTL when τ is never theoretically achievable. For example, suppose we investigate "time from birth to death" in a cohort and follow each subject in the cohort for 120 years. In this case, it is not reasonable to think that $\tau = 120$ years is achievable. Thus the time lost from $\tau = 120$ years may not be sensible. For other cases, interpreting the RMTL is as straightforward as the RMST.

2.2 Between-group comparison without covariate adjustment

Because the RMST (or the RMTL) is a summary of the survival-time distribution, we can simply compare these metrics between the two groups using the difference or the ratio. Let us consider the following three measures:

1. Difference in RMST
$$D_1 = \mu_1 - \mu_0$$
2. Ratio of RMST
$$D_2 = \mu_1/\mu_0$$
3. Ratio of RMTL
$$D_3 = (\tau - \mu_1) \,/\, (\tau - \mu_0)$$

μ_1 and μ_0 are the RMST for treatment groups 1 and 0, respectively. The estimators for those metrics are given by simply replacing μ_1 and μ_0 by their empirical counterparts (that is, $\widehat{\mu}_1$ and $\widehat{\mu}_0$, respectively). For example,

$$\widehat{D}_1 = \widehat{\mu}_1 - \widehat{\mu}_0$$

and

$$\widehat{\mathrm{Var}}\left(\widehat{D}_1\right) = \widehat{\mathrm{Var}}\left(\widehat{\mu}_1\right) + \widehat{\mathrm{Var}}\left(\widehat{\mu}_0\right)$$

For inference on the ratio-type metrics, D_2 and D_3, we use the delta method to calculate their variance estimates. Specifically, because

$$\begin{aligned}\log\left(\widehat{D}_2\right) &= \log\left(\widehat{\mu}_1\right) - \log\left(\widehat{\mu}_0\right)\\ \mathrm{Var}\left\{\log\left(\widehat{D}_2\right)\right\} &= \mathrm{Var}\left\{\log\left(\widehat{\mu}_1\right)\right\} + \mathrm{Var}\left\{\log\left(\widehat{\mu}_0\right)\right\}\\ &\approx \widehat{\mu}_1^{-2}\widehat{\mathrm{Var}}\left(\widehat{\mu}_1\right) + \widehat{\mu}_0^{-2}\widehat{\mathrm{Var}}\left(\widehat{\mu}_0\right)\end{aligned}$$

Thus we construct a confidence interval for $\log(D_2)$ by $\log(\widehat{D}_2)$ and $\widehat{\mathrm{Var}}\{\log(\widehat{D}_2)\}$ and then transform it back to the original ratio scale.

2.3 Between-group comparison with covariate adjustment

The new command `strmst2` accommodates an analysis of covariance-type adjusted analysis proposed by Tian, Zhao, and Wei (2014), so that users can consider covariates in the between-group comparison.

To describe Tian, Zhao, and Wei's (2014) method, we need some more notation. For a given truncation time τ, let $Y = \min(T, \tau)$ be the restricted survival time, and let Z be the treatment indicator. Also, let V denote a q-dimensional vector for baseline covariates. Tian, Zhao, and Wei's (2014) method considers the following generalized linear regression model,

$$E(Y \mid Z, V) = g^{-1}\left(\alpha + \beta Z + \gamma' V\right)$$

where $g(\cdot)$ is a given smooth and strictly increasing link function and $\theta = (\alpha, \beta, \gamma')'$ is a $(q+2)$-dimension unknown parameter vector with β being the parameter of the primary interest measuring the group difference in RMST adjusted for baseline covariates V. Tian, Zhao, and Wei (2014) use the following inverse-probability censoring weighted estimation equation to fit the model parameters,

$$S_n(\theta) = n^{-1}\sum_{i=1}^{n}\frac{\widetilde{\Delta}_i}{\widehat{G}\left\{\min\left(Y_i, \tau\right)\right\}}(1, Z_i, V_i')'\left\{Y_i - g^{-1}\left(\alpha + \beta Z_i + \gamma' V_i\right)\right\} = 0$$

where $\widehat{G}(\cdot)$ is the KM estimator of the censoring time C and $\widetilde{\Delta}_i = I\{\min(Y_i, \tau) \le C_i\}$. Let $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}')'$ be the solution to $S_n(\theta) = 0$. Tian, Zhao, and Wei (2014) show that as n goes to ∞, $\widehat{\theta}$ converges to the solution to

$$S(\theta) = E\left(1, Z, V'\right)'\left\{Y - g^{-1}\left(\alpha + \beta Z + \gamma' V\right)\right\} = 0$$

even when the regression model is misspecified. Also, they show that $W = n^{1/2}(\widehat{\theta} - \theta_0)$, where θ_0 is the solution to $S(\theta)$, is asymptotically normal and provide a procedure to estimate the variance of W.

The choice of the link function $g(\cdot)$ corresponds to the choice of measure. Specifically, `strmst2` automatically chooses the identity function as $g(\cdot)$ for the inference of the

difference in RMST, and it chooses $\log(\cdot)$ for the inference of the ratio of RMST. For the inference of the ratio of RMTL, $\log(\cdot)$ is used as the link function, and Y's in the above estimating equations are all replaced by $\tau - Y$.

Note that several other methods for the covariate adjustment have been studied (Zucker 1998; Andersen, Hansen, and Klein 2004). For example, Andersen, Hansen, and Klein (2004) studied a pseudovalue technique to handle censored observations and applied it to a regression model for the RMST. Program codes for their pseudovalue approach are available on the three major platforms (Stata, R, and SAS) with detailed documentation (Klein et al. 2008; Parner and Andersen 2010). In contrast to Andersen, Hansen, and Klein's (2004) method, Tian, Zhao, and Wei's (2014) method uses an inverse-probability censoring weighting technique to handle censored observations instead of using pseudovalues.

3 The strmst2 command

3.1 Syntax

`strmst2` *groupvar* [*if*] [*in*] [`, tau(`#`) covariates(`*varlist*`) level(`#`) reference(`#`) rmtl`]

As with all `st` commands, the data must be declared as survival time using `stset` before using `strmst2`. A numeric variable for the group should always be specified. For example, this numeric variable can be an indicator variable with values of `1` (typically indicating subjects in the active treatment group) or `0` (typically indicating subjects in the control group).

3.2 Options

`tau(`#`)` specifies the truncation time point for the RMST calculation as a scalar value. `tau()` needs to be smaller than the minimum of the largest observed time (either event or censor) in each of the groups. The default is to use the minimum of the largest observed event time among all groups.

`covariates(`*varlist*`)` specifies covariates to be used for the adjusted analyses. The default is to perform unadjusted analyses. When `covariates()` is specified, the analysis of covariance-type adjusted analyses are performed using those variables passed as covariates. This can be one variable or more than one variable.

`level(`#`)` specifies the confidence level; the default is `level(95)`, which will produce 95% confidence intervals.

`reference(`#`)` specifies the reference category. The default is the smallest value.

`rmtl` displays between-group contrasts for the ratio of the RMTL, in addition to the metrics for the RMST. The default is to show between-group contrasts only for the RMST.

3.3 Stored results

`strmst2` stores the following in `r()`:

Scalars

`r(tau)`	truncation time used in the analyses
`r(reference)`	value of the reference category used in the analyses

Matrices

`r(unadjustedresult)`	between-group contrasts from the unadjusted analyses
`r(rmstarm0)`	RMST results in arm 0 for the unadjusted analyses
`r(rmstarm1)`	RMST results in arm 1 for the unadjusted analyses
`r(adjustedresult)`	between-group contrasts from the adjusted analyses
`r(rmstdiffadj)`	results of the parameter estimates with the model to derive an adjusted difference in RMST
`r(rmstratioadj)`	results of the parameter estimates with the model to derive an adjusted ratio in RMST
`r(rmtlratioadj)`	results of the parameter estimates with the model to derive an adjusted ratio in RMTL

4 The strmst2pw command

4.1 Syntax

`strmst2pw` *indicator1* [`, rmtl`] (Syntax 1)

`strmst2pw` *indicator1 reference_indicator* [`, rmtl`] (Syntax 2)

`strmst2pw` can be used after `strmst2` with the `covariates()` option. With syntax 1, a summary of between-group contrasts is displayed for the group identified by *indicator1* versus the reference category used in the previously fit model. With syntax 2, a summary of between-group contrasts is displayed for the group identified by *indicator1* versus the group identified by *reference_indicator*, which represents a different group from that used as the reference category in the previously fit model.

4.2 Option

`rmtl` displays between-group contrasts for the ratio of the RMTL, in addition to the metrics for the RMST. The default is to show between-group contrasts only for the RMST.

5 Examples

To illustrate the new command, we use data from the PBC study conducted by the Mayo Clinic. The list of raw data are seen in Fleming and Harrington (2011, 359, *Appendix D*). Here we use 312 cases who participated in the randomized phase—158 cases randomized to the D-penicillamine group and 154 cases randomized to the placebo group. The edited dataset includes a variable for time on study (in years) and survival status, which takes a value of `1` for subjects who died. The `treatment` variable is coded

as `1` (for the active treatment arm) or `0` (for the control arm). Figure 1 displays the KM estimate for time to death within each treatment group.

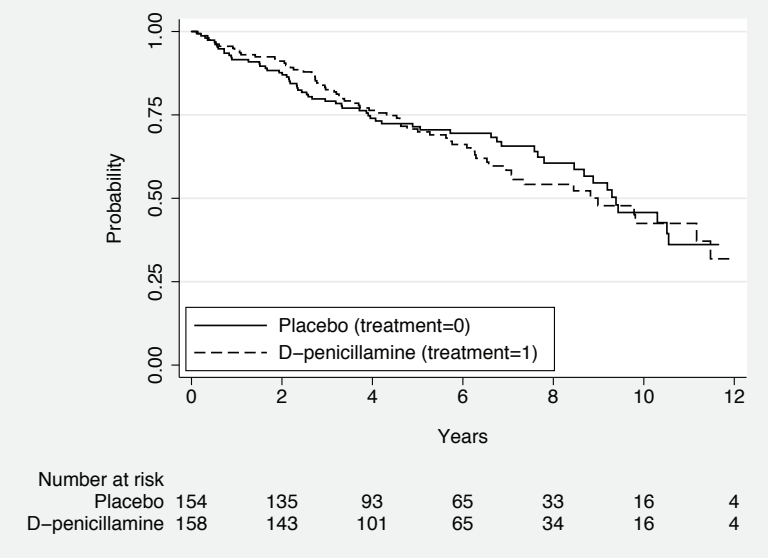

Figure 1. KM estimate for time to death within each treatment group

The example data are then loaded into memory and declared as survival time.

```
. version 13.1
. set seed 1234
. use pbc
(PBC data (Fleming and Harrington, Appendix D, 1991, Wiley))
. stset time, f(status)
```

(*output omitted*)

5.1 Unadjusted analysis

We first examine the association between the treatment group and the RMST measures with unadjusted analysis. Below is an example syntax for unadjusted analysis and output when a truncation time of 10 is specified. Because the `time` variable is given in years, the estimates will be interpreted for a follow-up time of 10 years. The `reference()` option is not specified, so results will be presented versus the default reference group. In this example, the control arm is coded with the smallest value for the `treatment` variable; thus the control arm is the default reference group. The `rmtl` option is specified to obtain results for the RMTL metrics as well as the RMST metrics.

```
. strmst2 treatment, tau(10) rmtl

Number of observations for analysis = 312

The truncation time: tau = 10 was specified.
Restricted Mean Survival Time (RMST) by arm
```

Group	Estimate	Std. Err.	[95% Conf.	Interval]
arm 1	7.146	0.284	6.589	7.704
arm 0	7.283	0.297	6.700	7.866

```
Restricted Mean Time Lost (RMTL) by arm
```

Group	Estimate	Std. Err.	[95% Conf.	Interval]
arm 1	2.854	0.284	2.296	3.411
arm 0	2.717	0.297	2.134	3.300

```
Between-group contrast (arm 1 versus arm 0)
```

Contrast	Estimate	[95% Conf.	Interval]	P>|z|
RMST (arm 1 - arm 0)	-0.137	-0.943	0.669	0.739
RMST (arm 1 / arm 0)	0.981	0.877	1.097	0.739
RMTL (arm 1 / arm 0)	1.050	0.786	1.404	0.740

The `strmst2` command returns the RMST and RMTL on each group and the results of the between-group contrast measures listed above. Here the difference in RMST (the first row of the `Between-group contrast` in the output) was -0.137 years. The point estimate indicated that on average, patients on the active treatment survive 0.137 years shorter than those on the placebo when patients were followed up for 10 years. While no statistically significant difference was observed ($p = 0.739$), the 95% confidence interval (-0.943 to 0.669) was relatively tight around 0, suggesting that the difference in RMST would be at most plus or minus one year. The ratio of RMST was 0.981 and the ratio of RMTL was 1.050, with p-values very similar to that for the difference in RMST.

5.2 Adjusted analysis

For implementation of Tian, Zhao, and Wei's (2014) adjusted analysis for the RMST, the only difference is for the user to specify the `covariates()` option to the command. For illustration, let's use three baseline variables in the PBC data (`age`, `bili`, and `albumin`) as the covariates for adjustment. Below is the corresponding syntax and output of the adjusted analyses, again specifying a truncation time of 10 years. As before, the `reference()` option is not specified, so results will be presented using the default reference group.

```
. strmst2 treatment, tau(10) covariates(age bili albumin) rmtl

Number of observations for analysis = 312

The truncation time: tau = 10 was specified.
Note: adjusted analysis may take a few minutes to run...
```

Model summary (difference of RMST)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
intercept	2.743	2.134	1.29	0.199	-1.440	6.926
_Itreatment_1	-0.210	0.343	-0.61	0.540	-0.883	0.463
age	-0.069	0.018	-3.90	0.000	-0.103	-0.034
bili	-0.325	0.039	-8.39	0.000	-0.401	-0.249
albumin	2.550	0.472	5.40	0.000	1.624	3.475

Model summary (ratio of RMST)

	Coef.	Std. Err.	z	P>|z|	exp(Coef.)	[95% Conf.	Interval]
intercept	1.369	0.356	3.84	0.000	3.930	1.955	7.899
_Itreatment_1	-0.033	0.050	-0.65	0.514	0.968	0.877	1.068
age	-0.009	0.003	-3.41	0.001	0.991	0.985	0.996
bili	-0.087	0.013	-6.52	0.000	0.917	0.893	0.941
albumin	0.360	0.080	4.49	0.000	1.434	1.225	1.678

Model summary (ratio of time-lost)

	Coef.	Std. Err.	z	P>|z|	exp(Coef.)	[95% Conf.	Interval]
intercept	1.992	0.695	2.86	0.004	7.332	1.876	28.655
_Itreatment_1	0.035	0.127	0.27	0.786	1.035	0.806	1.329
age	0.025	0.007	3.81	0.000	1.026	1.012	1.039
bili	0.063	0.008	8.33	0.000	1.065	1.049	1.080
albumin	-0.750	0.149	-5.03	0.000	0.472	0.353	0.633

We provide a summary for each of the three models. With adjustment for the three specified covariates, the difference in RMST was -0.210, and this difference was not statistically significant ($p = 0.540$). Note that with the adjusted analyses, indicator variables are created to represent the treatment assignment. In this scenario, an indicator variable called `_Itreatment_1` was created and is coded as `1` for the active treatment arm and `0` otherwise.

5.3 Extension to more than two comparison groups

`strmst2` can accommodate comparisons of more than two treatment groups. For illustration, we first create a third treatment arm using the PBC data. We then run the unadjusted and adjusted analyses by specifying the `reference()` option so that patients coded with `treatment` = 2 will be the reference group.

```
. replace treatment=2 if runiform()<0.3
(97 real changes made)
. strmst2 treatment, tau(10) reference(2) rmtl

Number of observations for analysis = 312

The truncation time: tau = 10 was specified.
Restricted Mean Survival Time (RMST) by arm
```

Group	Estimate	Std. Err.	[95% Conf.	Interval]
arm 2	7.485	0.351	6.797	8.173
arm 1	7.165	0.338	6.502	7.828
arm 0	7.019	0.379	6.277	7.761

Restricted Mean Time Lost (RMTL) by arm

Group	Estimate	Std. Err.	[95% Conf.	Interval]
arm 2	2.515	0.351	1.827	3.203
arm 1	2.835	0.338	2.172	3.498
arm 0	2.981	0.379	2.239	3.723

Between-group contrast (arm 1 versus arm 2)

Contrast	Estimate	[95% Conf.	Interval]	P>|z|
RMST (arm 1 - arm 2)	-0.320	-1.275	0.636	0.512
RMST (arm 1 / arm 2)	0.957	0.840	1.091	0.512
RMTL (arm 1 / arm 2)	1.127	0.787	1.615	0.514

Between-group contrast (arm 0 versus arm 2)

Contrast	Estimate	[95% Conf.	Interval]	P>|z|
RMST (arm 0 - arm 2)	-0.466	-1.478	0.546	0.367
RMST (arm 0 / arm 2)	0.938	0.815	1.079	0.368
RMTL (arm 0 / arm 2)	1.185	0.819	1.716	0.367

```
. strmst2 treatment, tau(10) covariates(age bili albumin) reference(2) rmtl

Number of observations for analysis = 312

The truncation time: tau = 10 was specified.
Note: adjusted analysis may take a few minutes to run...
```

Model summary (difference of RMST)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
intercept	2.539	2.132	1.19	0.234	-1.639	6.717
_Itreatment_0	-0.169	0.416	-0.41	0.685	-0.985	0.647
_Itreatment_1	-0.202	0.398	-0.51	0.612	-0.982	0.579
age	-0.068	0.018	-3.82	0.000	-0.102	-0.033
bili	-0.321	0.039	-8.32	0.000	-0.397	-0.246
albumin	2.596	0.476	5.45	0.000	1.662	3.530

Model summary (ratio of RMST)

	Coef.	Std. Err.	z	P>|z|	exp(Coef.)	[95% Conf.	Interval]
intercept	1.317	0.359	3.67	0.000	3.733	1.846	7.550
_Itreatment_0	-0.002	0.060	-0.03	0.978	0.998	0.887	1.123
_Itreatment_1	-0.024	0.060	-0.39	0.694	0.977	0.868	1.099
age	-0.009	0.003	-3.26	0.001	0.991	0.986	0.996
bili	-0.087	0.013	-6.65	0.000	0.917	0.894	0.941
albumin	0.368	0.081	4.55	0.000	1.445	1.233	1.693

Model summary (ratio of time-lost)

	Coef.	Std. Err.	z	P>|z|	exp(Coef.)	[95% Conf.	Interval]
intercept	2.024	0.684	2.96	0.003	7.569	1.979	28.949
_Itreatment_0	0.168	0.160	1.05	0.292	1.183	0.865	1.617
_Itreatment_1	0.055	0.145	0.38	0.704	1.057	0.795	1.405
age	0.026	0.007	3.92	0.000	1.026	1.013	1.039
bili	0.060	0.008	8.00	0.000	1.062	1.047	1.078
albumin	-0.783	0.154	-5.08	0.000	0.457	0.338	0.618

For the unadjusted results, two sets of between-group contrasts are shown: one set for patients coded with `treatment` = 1 versus the reference group (`treatment` = 2) and the second set for patients coded with `treatment` = 0 versus the reference group. For the adjusted results, two indicator variables were created and included in the model: `_Itreatment_0` is coded as 1 for patients with `treatment` = 0 and as 0 otherwise; `_Itreatment_1` is coded as 1 for patients with `treatment` = 1 and as 0 otherwise. In comparison to treatment group 2, the unadjusted difference in RMST was −0.320 for the control group and −0.466 for treatment group 1; neither of these differences was statistically significant ($p = 0.512$ and $p = 0.367$, respectively). With adjustment for the three covariates, the difference in RMST was −0.169 for the control group versus treatment group 2 ($p = 0.685$) and −0.202 for treatment group 1 versus treatment group 2 ($p = 0.612$).

If we are interested in reporting the adjusted difference in treatment group 1 versus the control group, we can use syntax 2 of the postestimation command `strmst2pw` with the indicator variables created from the previous analysis, as shown below.

```
. strmst2pw _Itreatment_1 _Itreatment_0, rmtl
Summary of between-group contrast (adjusted for the covariates)
```

	Estimate	[95% Conf.	Interval]	P>|z|
RMST (arm 1 - arm 0)	-0.033	-0.892	0.827	0.940
RMST (arm 1 / arm 0)	0.978	0.862	1.110	0.732
RMTL (arm 1 / arm 0)	0.893	0.654	1.220	0.478

Lastly, if we wish to see these between-group contrasts for treatment group 1 versus treatment group 2 in this same format, we would use syntax 1 of `strmst2pw` because treatment group 2 was the reference group in the previously fit model.

```
. strmst2pw _Itreatment_1, rmtl
Summary of between-group contrast (adjusted for the covariates)
```

	Estimate	[95% Conf.	Interval]	P>|z|
RMST (arm 1 - arm 2)	-0.202	-0.982	0.579	0.612
RMST (arm 1 / arm 2)	0.977	0.868	1.099	0.694
RMTL (arm 1 / arm 2)	1.057	0.795	1.405	0.704

The results summarized above are identical to those obtained from `strmst2` with the `reference()` option specified such that treatment group 2 is the reference group.

6 Conclusion

We introduced the new commands `strmst2` and `strmst2pw`, which implement inference of the difference in the RMST, the ratio of the RMST, and the ratio of the RMTL to summarize the difference between survival distributions. These are robust and clinically interpretable summary metrics; these are useful alternatives to the hazard ratio. Similar program code is available for both R (`survRM2` package; https://cran.r-project.org/web/packages/survRM2/) and SAS (http://bcb.dfci.harvard.edu/∼huno).

7 References

Andersen, P. K., and R. D. Gill. 1982. Cox's regression model for counting processes: A large sample study. *Annals of Statistics* 10: 1100–1120.

Andersen, P. K., M. G. Hansen, and J. P. Klein. 2004. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Analysis* 10: 335–350.

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

Fleming, T. R., and D. P. Harrington. 2011. *Counting Processes and Survival Analysis*. 2nd ed. Hoboken, NJ: Wiley.

Gill, R. D. 1984. Understanding Cox's regression model: A Martingale approach. *Journal of the American Statistical Association* 79: 441–447.

Hernán, M. A. 2010. The hazards of hazard ratios. *Epidemiology* 21: 13–15.

Hjort, N. L. 1992. On inference in parametric survival data models. *International Statistical Review* 60: 355–387.

Klein, J. P., M. Gerster, P. K. Andersen, S. Tarima, and M. Pohar Perme. 2008. SAS and R functions to compute pseudo-values for censored data regression. *Computer Methods and Programs in Biomedicine* 89: 289–300.

Lin, D. Y., and L. J. Wei. 1989. The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* 84: 1074–1078.

Miller, R. G., Jr. 1981. *Survival Analysis*. New York: Wiley.

Parner, E. T., and P. K. Andersen. 2010. Regression analysis of censored data using pseudo-observations. *Stata Journal* 10: 408–422.

Royston, P., and M. K. B. Parmar. 2011. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine* 30: 2409–2421.

———. 2013. 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 Medical Research Methodology* 13: 152.

Tian, L., L. Zhao, and L. J. Wei. 2014. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. *Biostatistics* 15: 222–233.

Uno, H., B. Claggett, L. Tian, E. Inoue, P. Gallo, T. Miyata, D. Schrag, M. Takeuchi, Y. Uyama, L. Zhao, H. Skali, S. Solomon, S. Jacobus, M. Hughes, M. Packer, and L.-J. Wei. 2014. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of Clinical Oncology* 32: 2380–2385.

Uno, H., J. Wittes, H. Fu, S. D. Solomon, B. Claggett, L. Tian, T. Cai, M. A. Pfeffer, S. R. Evans, and L.-J. Wei. 2015. Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies. *Annals of Internal Medicine* 163: 127–134.

Zucker, D. M. 1998. Restricted mean life with covariates: Modification and extension of a useful survival analysis method. *Journal of the American Statistical Association* 93: 702–709.

About the authors

Angel Cronin is a statistician in the Division of Population Sciences at the Dana–Farber Cancer Institute in Boston.

Lu Tian is an associate professor in the Department of Biomedical Data Science at the Stanford University School of Medicine.

Hajime Uno is an assistant professor in the Division of Population Sciences at the Dana–Farber Cancer Institute in Boston.