Multivariate meta-analysis model for the difference in restricted mean survival times

ISABELLE R. WEIR*

Department of Biostatistics, Boston University School of Public Health, Boston, 801 Massachusetts Avenue, MA 02118, USA

iweir@bu.edu

LU TIAN

Department of Health Research and Policy, Stanford University, 150 Governor's Lane, Palo Alto, CA 94305, USA

LUDOVIC TRINQUART

Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Avenue, Boston, MA 02118, USA

SUMMARY

In randomized controlled trials (RCTs) with time-to-event outcomes, the difference in restricted mean survival times (RMSTD) offers an absolute measure of the treatment effect on the time scale. Computation of the RMSTD relies on the choice of a time horizon, τ . In a meta-analysis, varying follow-up durations may lead to the exclusion of RCTs with follow-up shorter than τ . We introduce an individual patient data multivariate meta-analysis model for RMSTD estimated at multiple time horizons. We derived the within-trial covariance for the RMSTD enabling the synthesis of all data by borrowing strength from multiple time points. In a simulation study covering 60 scenarios, we compared the statistical performance of the proposed method to that of two univariate meta-analysis models, based on available data at each time point and based on predictions from flexible parametric models. Our multivariate model yields smaller mean squared error over univariate methods at all time points. We illustrate the method with a meta-analysis of five RCTs comparing transcatheter aortic valve replacement (TAVR) with surgical replacement in patients with aortic stenosis. Over 12, 24, and 36 months of follow-up, those treated by TAVR live 0.28 [95% confidence interval (CI) 0.01 to 0.56], 0.46 (95% CI -0.08 to 1.01), and 0.79 (95% CI -0.43 to 2.02) months longer on average compared to those treated by surgery, respectively.

Keywords: Meta-analysis; Randomized controlled trial; Survival analysis.

1. INTRODUCTION

The difference in restricted mean survival times (RMSTD) is an appealing measure of treatment effect for randomized controlled trials (RCTs) with time-to-event outcomes. The RMSTD compares the mean

*To whom correspondence should be addressed.

survival times between the experimental and control groups up to a pre-specified, clinically relevant time point of interest, τ . We can estimate the RMSTD, $\delta(\tau)$, by taking the difference between the areas under the Kaplan–Meier curves from time 0 to τ in the experimental and comparator groups, respectively: $\hat{\delta}(\tau) = \int_0^\tau \hat{S}_E(u)du - \int_0^\tau \hat{S}_C(u)du$, where $\hat{S}_C$ and $\hat{S}_E$ are the Kaplan–Meier estimators of the survival functions for the experimental and comparator groups. We can interpret the RMSTD over time τ as the expected gain (or loss) in time to the event of interest in the experimental group over the comparator group. Some advantages include its clinically meaningful interpretation as an absolute measure of treatment effect in the time domain and lack of model assumptions such as proportional hazards (Royston and Parmar, 2013; Uno and others, 2014; Trinquart and others, 2016).

Meta-analyses of RCTs with time-to-event outcomes rely primarily on hazard ratios. Presenting the RMSTD would give additional insights and may influence the interpretation of treatment effects (Weir and others, 2018a). There is a limited number of methods described for meta-analyses using RMSTD as a measure of treatment effect (Wei and others, 2015; Lueza and others, 2016a). A potential issue in conducting a meta-analysis of RMSTD is that the selected RCTs may not necessarily share a common follow-up duration. Thus, conducting a univariate meta-analysis using a single fixed time horizon, τ , may lead to exclusion of available data because some RCTs may end before the defined time horizon while others may extend beyond. A second univariate method circumvents this issue by using estimated treatment effects from fitted flexible parametric models. However, this method may not be preferred because the estimates are based on extrapolation (Wei and others, 2015).

In this article, we introduce a multivariate method for meta-analysis of RMSTD with individual patient data. We extend the multivariate meta-analysis framework for synthesizing multiple outcomes to the setting of multiple time points. Our multivariate random-effects model for RMSTD incorporates between-time point covariances, allowing the meta-analysis to make use of and borrow strength from all available follow-up times across all RCTs. Our method gives combined treatment effect estimates of the RMSTD at each time point of interest. Unlike previous methods, it incorporates all observed data and does not rely on predictions from fitted models.

In Section 2, we introduce our multivariate model. In Section 3, we describe previous univariate methods. In Sections 4 and 5, we describe the methods and results of a simulation study. In Section 6, we report an illustrative example on RCTs of transcatheter aortic valve replacement (TAVR) in patients with aortic stenosis.

2. MULTIVARIATE META-ANALYSIS MODEL

Suppose, that we are interested in estimating the combined RMSTD at times $\tau_1, \dots, \tau_K$ based on a meta-analysis of individual patient data from J RCTs. To this end, we propose the following multivariate random-effects model. We denote the vector of RMSTD estimates for RCT j by $\hat{\delta}_j$, the entries of which we anticipate to be correlated. It is assumed that within each RCT:

$$\hat{\delta}_j \mid \delta_j = \begin{bmatrix} \hat{\delta}_{j,\tau_1} \\ \hat{\delta}_{j,\tau_2} \\ \vdots \\ \hat{\delta}_{j,\tau_{K_j}} \end{bmatrix} \mid \begin{bmatrix} \delta_{j,\tau_1} \\ \delta_{j,\tau_2} \\ \vdots \\ \delta_{j,\tau_{K_j}} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \delta_{j,\tau_1} \\ \delta_{j,\tau_2} \\ \vdots \\ \delta_{j,\tau_{K_j}} \end{bmatrix}, \hat{\Omega}_j \right), \quad (2.1)$$

where $\mathcal{N}$ denotes a multivariate normal distribution, δ_j is the vector of true underlying RMSTDs for RCT j at time horizons $\tau_1, \dots, \tau_{K_j}$, and $\hat{\Omega}_j$ denotes the within-trial covariance matrix of dimension $\tau_{K_j} \times \tau_{K_j}$ associated with $\hat{\delta}_j$.

Our primary interest is the mean vector of RMSTDs, $(\delta_{\tau_1}, \dots, \delta_{\tau_K})'$ for our set of time horizons. The multivariate random-effects model allows for the true RMSTD to vary from trial to trial:

$$\delta_j \sim N \left(\begin{bmatrix} \delta_{\tau_1} \\ \delta_{\tau_2} \\ \vdots \\ \delta_{\tau_K} \end{bmatrix}, \Sigma \right), \quad (2.2)$$

where Σ is the between-trial covariance matrix. This random-effects model accommodates differences in the availability of RMSTD due to varying follow-up durations across RCTs. In RCT j , the RMSTD can be estimated until t^* , the minimum of the maximum observed times in each randomization group (Uno and others, 2017). Thus, the RMSTD is identifiable at times $\tau_1, \dots, \tau_{K_j}$ with $K_j = \arg \max_{k \in [1, K]} \tau_k \leq t^*$.

The multivariate model makes use of all available data. Through the correlation between RMSTD at different time horizons, it borrows strength across RCTs to yield estimates with better statistical properties (Jackson and others, 2017).

The random-effects multivariate model requires estimation of both the within-trial and between-trial covariance, Ω_j and Σ , respectively. The between-trial covariance accounts for the correlation between the RMSTD values at different time horizons across trials. We regard Σ as unstructured and estimate it using the extension of the DerSimonian and Laird method-of-moments estimator to the multivariate situation (Chen and others, 2012; Jackson and others, 2010). The within-trial covariance accounts for the fact that in a given RCT, estimation of the RMSTD at different time horizons uses the same set of patient data. It specifies the covariance between the RMSTD at each pair of time horizons, τ_{k-1} and τ_k , for trial j . We derive the within-trial covariance matrix $\hat{\Omega}_j$. We also describe a bootstrap method in the next section.

2.1. Analytic derivation of within-trial covariance

By using the properties of martingale integrals, we show that for $\tau_k \leq \tau_l$, the covariance for the RMSTD at τ_k and τ_l reduces to the variance of the RMSTD at time τ_k (Text A1 of [supplementary material](#) available at *Biostatistics* online). The resulting covariance matrix is symmetrical, containing only variance terms.

That is, for RCT j , $\text{cov}(\delta_{j,\tau_k}, \delta_{j,\tau_l}) = \sigma^2(\delta_{j,\tau_k})$.

Specifically, we estimate σ_{kl}^2 , $k \leq l$, the (k, l) th element of the matrix, by

$$\hat{\sigma}^2(\delta_{j,\tau_k}) = \int_0^{\tau_k} \left(\int_t^{\tau_k} \hat{S}_{j1}(u) du \right)^2 \frac{d\hat{\Lambda}_{j1}(t)}{Y_{j1}(t)} + \int_0^{\tau_k} \left(\int_t^{\tau_k} \hat{S}_{j0}(u) du \right)^2 \frac{d\hat{\Lambda}_{j0}(t)}{Y_{j0}(t)}, \quad (2.3)$$

where $\hat{S}_{ji}(u)$ is the Kaplan–Meier estimator of the survival function of group i of RCT j , $\hat{\Lambda}_{ji}(t) = -\log\{\hat{S}_{ji}(t)\}$, the Nelson–Aalen estimator, and $Y_{ji}(t)$ is the number of patients at risk in the i th group of RCT j (Tian and others, 2017; Pepe and Fleming, 1989, 1991). Thus,

$$\hat{\Omega}_j = \begin{pmatrix} \hat{\sigma}^2(\delta_{j,\tau_1}) & \hat{\sigma}^2(\delta_{j,\tau_1}) & \hat{\sigma}^2(\delta_{j,\tau_1}) & \dots & \hat{\sigma}^2(\delta_{j,\tau_1}) \\ \hat{\sigma}^2(\delta_{j,\tau_1}) & \hat{\sigma}^2(\delta_{j,\tau_2}) & \hat{\sigma}^2(\delta_{j,\tau_2}) & \dots & \hat{\sigma}^2(\delta_{j,\tau_2}) \\ \hat{\sigma}^2(\delta_{j,\tau_1}) & \hat{\sigma}^2(\delta_{j,\tau_2}) & \hat{\sigma}^2(\delta_{j,\tau_3}) & \dots & \hat{\sigma}^2(\delta_{j,\tau_3}) \\ \dots & \dots & \dots & \dots & \dots \\ \hat{\sigma}^2(\delta_{j,\tau_1}) & \hat{\sigma}^2(\delta_{j,\tau_2}) & \hat{\sigma}^2(\delta_{j,\tau_3}) & \dots & \hat{\sigma}^2(\delta_{j,\tau_{K_j}}) \end{pmatrix} \quad (2.4)$$

2.2. Bootstrap derivation of within-trial covariance

In each of $B = 500$ bootstrap iterations, we sample each randomization group with replacement (Efron and Tibshirani, 1993). We then compute the RMSTD at the time horizons of interest using each of the re-sampled datasets. We compute the mean RMSTD at each time horizon over all B samples, $\bar{\delta}$. Finally, the bootstrap estimator of the within-trial covariance matrix is given by:

$$\hat{\Omega}_j = \frac{1}{B-1} \sum_{b=1}^B (\hat{\delta}_{j,b} - \bar{\delta}_j) * (\hat{\delta}_{j,b} - \bar{\delta}_j)^T \quad (2.5)$$

3. UNIVARIATE META-ANALYSIS METHODS

3.1. Univariate meta-analyses using only available data

A simple approach for a meta-analysis of the RMSTD reported at various follow-up times is to conduct a univariate meta-analysis at each time horizon of interest using only the RCTs with available RMSTD. Thus, only RCTs with t^* beyond each specified time horizon contribute to the respective overall estimated treatment effect at each time horizon, τ_k . We use the DerSimonian and Laird random-effects model for consistency with the multivariate model.

3.2. Univariate meta-analyses using predictions from flexible parametric models

A second univariate method with individual patient data uses predicted treatment effects at each time horizon for all RCTs (Wei and others, 2015). Royston and Parmar (2002) have developed a flexible parametric model to approximate the baseline log cumulative hazard function by fitting a restricted cubic spline function using the log of time. By fitting the flexible parametric model in each RCT, we can estimate the RMSTD at every time horizon of interest along with their standard errors, which are based on the delta method using numerical differentiation. These computations are implemented with the `rstpm2` package in R (Clements and Liu, 2018). We use these estimates and standard errors to conduct a univariate meta-analysis at each time horizon, again using the DerSimonian and Laird random-effects model for consistency (Wei and others, 2015). Thus, at each time horizon, estimates for all J RCTs contribute to the meta-analysis. This method removes the issue of missing data; however, it relies on estimates from a model with no guarantee of adequate fit.

In each RCT, the approximate log cumulative baseline hazard is given by $\ln H_0(t) = \gamma_0 + \gamma_1 \ln(t) + \gamma_2 \eta_1(\ln(t)) + \dots + \gamma_{K_0+1} \eta_{K_0}(\ln(t))$ where, γ_m ($m = 0, 1, \dots, K_0+1$) and η_m ($m = 1, 2, \dots, K_0$) represent the regression parameters and spline basis functions, respectively. K_0 is the number of distinct internal knots. Setting $\ln H_0(t)$ equal to the baseline spline function, $s(\ln t | \gamma, K_0)$, we can obtain the log cumulative hazard function by $\ln H(t) = s(\ln t | \gamma, K_0) + s(\ln t | \delta, K_1) x_{trt} + \beta x_{trt}$, where x_{trt} is an indicator of the treatment group and the interaction term, $s(\ln t | \delta, K_1) x_{trt}$ accounts for potential non-proportional hazards.

We use a model with three degrees of freedom (two knots) for the baseline distribution and one degree of freedom for the time-dependent effect to estimate the RMST (Wei and others, 2015). We can use maximum likelihood for parameter estimation since the model is fully parametric. Thus, the full model used throughout this article is given by $\ln H(t) = \gamma_0 + \gamma_1 \ln(t) + \gamma_2 \eta_1(\ln(t)) + \gamma_3 \eta_2(\ln(t)) + \gamma_1 (\ln(t)) x_{trt} + \beta x_{trt}$.

4. SIMULATION STUDY

We used a simulation study to assess the statistical performances of the proposed multivariate model and compare them to the two univariate methods. We simulated meta-analyses of 5, 10, and 30 RCTs with time-to-event outcomes in 60 scenarios. We generated the individual patient data for each RCT in

the meta-analysis according to either a random or non-random follow-up pattern. Briefly, we assumed a Weibull or piecewise Weibull survival model for each RCT and generated event times according to treatment group. In the case of non-random follow-up duration, RCTs showing statistically significant effects at intermediate time horizons were more likely to continue. For a given simulated meta-analysis, we introduced between-trial heterogeneity by incorporating variation on the baseline hazard and treatment effect.

4.1. Scenarios and performance measures

We defined 48 scenarios by the number of RCTs in the meta-analysis, the amount of between-trial heterogeneity, the true underlying hazard ratio, and whether we generated the data with proportional hazards or not (Table 1). In each scenario, we computed the combined effect at time horizons of $\tau_1 = 1$, $\tau_2 = 2$, and $\tau_3 = 3$ years. We set the pattern of follow-up availability to have control over the number of RCTs with available RMSTD at each time horizon. This was (100%, 80%, 60%) at (τ_1, τ_2, τ_3) , respectively. We generated the data according to a Weibull distribution with scale $\lambda = 12$ and shape $\nu = 0.7$ for proportional hazards. We used scale $\lambda = 12$ and shape $\nu_0 = 0.7$ for the comparator group and $\nu_1 = 0.9$ for the experimental group for non-proportional hazards. These parameters ensured for all scenarios that the survival distribution would extend beyond the largest time horizon, $\tau_3 = 3$ years. We varied the number of RCTs in the meta-analysis to be either 5, 10, or 30 (Davey and others, 2011). For each scenario investigated, we assessed the average bias, mean squared error (MSE), and coverage probability of each method. We simulated 5000 meta-analyses for each scenario. We summarized results by using nested loop plots (Rücker and Schwarzer, 2014).

We explored 12 additional scenarios by modifying four features, one at a time: a different distribution of trial sample sizes, a different pattern of follow-up availability, a piecewise Weibull model for data

Table 1. Scenarios for the Monte Carlo simulation study

Number of RCTs	N available RCTs at 1, 2, 3 years [†]	Proportional hazards [‡]	True HR	Between-trial variance
Random follow-up pattern				
5	(5, 4, 3)	Yes	0.7, 0.9	0.005, 0.02
5	(5, 4, 3)	No	0.7, 0.9	0.005, 0.02
10	(10, 8, 6)	Yes	0.7, 0.9	0.005, 0.02
10	(10, 8, 6)	No	0.7, 0.9	0.005, 0.02
30	(30, 24, 18)	Yes	0.7, 0.9	0.005, 0.02
30	(30, 24, 18)	No	0.7, 0.9	0.005, 0.02
Non-random follow-up pattern				
5	(5, 4, 3)	Yes	0.7, 0.9	0.005, 0.02
5	(5, 4, 3)	No	0.7, 0.9	0.005, 0.02
10	(10, 8, 6)	Yes	0.7, 0.9	0.005, 0.02
10	(10, 8, 6)	No	0.7, 0.9	0.005, 0.02
30	(30, 24, 18)	Yes	0.7, 0.9	0.005, 0.02
30	(30, 24, 18)	No	0.7, 0.9	0.005, 0.02

[†]For five RCTs, we generated data so that the RMSTD is identifiable for five, four, and three RCTs at $\tau = 1, 2$, and 3 years, respectively. In the non-random setting, RCTs showing positive results at 1 and 2 years are more likely to yield RMSTD identifiable at 2 and 3 years, respectively.

[‡]In the case of proportional hazards, we set $\lambda = 12$, $\nu = 0.7$. In the case of non-proportional hazards, we set $\lambda = 12$, $\nu_0 = 0.7$, $\nu_1 = 0.9$.

generation, and the use of maximum likelihood estimation (Table A1 of [supplementary material](#) available at *Biostatistics* online.). We aimed to show that the performance of each method was consistent for each modified feature. The additional distribution of trial sample sizes enabled larger trials with mean 1400 and standard deviation 600. The additional pattern of follow-up availability was (100%, 40%, 40%) for time horizons (τ_1, τ_2, τ_3) , respectively, which represented a meta-analysis in which the number of trials with available RMSTD dropped most dramatically. We included the piecewise Weibull model to induce poorer fit of the flexible parametric models and thus penalize the univariate model based on the flexible parametric model predictions (Figure A1 of [supplementary material](#) available at *Biostatistics* online). We used scale $\lambda = 12$ and shape $\nu = 0.7$, but changed the hazard ratio at 1.5 years from hazard ratio (HR) = 1 to HR = 0.7. Finally, we included maximum likelihood estimation to show that the results were consistent with method of moments.

4.2. Data generation

We simulate the survival times for N patients from J RCTs, each of size n_j with $\sum_{j=1}^J n_j = N$. We draw n_j from one of two log-normal distributions either with mean 6.47 and standard deviation 0.41 in order to obtain a mean sample size of 700 with standard deviation 300 or with mean 7.24 and standard deviation 0.02 in order to obtain a mean sample size of 1400 with standard deviation 600 ([Trinquart and others, 2014](#)). These distributions of trial sample size are consistent with reports from systematic reviews of cancer RCTs ([Pestine and others, 2018](#); [Bariani and others, 2015](#)). We achieve an approximate 1:1 allocation ratio in each trial, j , by drawing from a binomial distribution with probability of assignment to experimental treatment $p = 0.5$.

We assume the survival pattern of the data follow a Weibull distribution with shape and scale parameters $\nu > 0, \lambda > 0$, respectively ([Bender and others, 2005](#)). Thus, we define the hazard function for RCT $j \in \{1, \dots, J\}$ at time t by $h_j(t, x | \alpha_j, \phi_j) = \frac{\nu}{\lambda} \exp(\alpha_j) \exp((\theta + \phi_j)x) \left(\frac{t}{\lambda}\right)^{\nu-1}$, where $x = \frac{1}{2}$ for the experimental group and $x = -\frac{1}{2}$ for the comparator group. These coding values impose equal between-trial variances in the log HR for the two treatment groups ([Smith and others, 2005](#)). Let θ be the true underlying log hazard ratio.

To procure between-trial heterogeneity, we use trial-specific values α_j and ϕ_j to vary the baseline hazard $\frac{\nu}{\lambda} \exp(\alpha_j)$ and hazard ratio $\exp(\theta + \phi_j)$, respectively. We let α_j and ϕ_j be transformed Binomial random variables centered and scaled to be approximately Normal with the desired variances, σ^2 and ρ^2 , respectively: $\alpha_j \sim \frac{[Bin(n, p) - np]}{\sqrt{np(1-p)}} * \sigma$ and $\phi_j \sim \frac{[Bin(n, p) - np]}{\sqrt{np(1-p)}} * \rho$, with $n = 50$ ([Lueza and others, 2016b](#)).

We generate independent observations from a uniform distribution for each patient, i , in each RCT, j , given by $U_{ij} \sim Unif(0, 1)$. The survival times for individuals in the experimental group and comparator group are given by $T_{ij}^E = \lambda * \left(\frac{-\log(U_{ij})}{\exp(\alpha_j) \exp(\frac{1}{2}(\theta + \phi_j))} \right)^{\frac{1}{\nu}}$ and $T_{ij}^C = \lambda * \left(\frac{-\log(U_{ij})}{\exp(\alpha_j) \exp(-\frac{1}{2}(\theta + \phi_j))} \right)^{\frac{1}{\nu}}$, respectively.

4.3. Random and non-random follow-up patterns

We first generated data under a random follow-up pattern, as described in Table 1 and Table A1 of [supplementary material](#) available at *Biostatistics* online. We either let the percentage of trials with available RMSTD at each time horizon be (100%, 80%, 60%) or (100%, 40%, 40%) at (τ_1, τ_2, τ_3) , respectively. For example in the former case with five RCTs, we retained the RMSTD estimates at 1, 2, and 3 years for the three first generated RCTs; we retained the RMSTD estimates at 1 and 2 years for the fourth generated RCT; and we retained only the RMSTD estimate at 1 year for the fifth generated RCT.

We then consider that RCTs showing statistically significant results at intermediate time horizons are more likely to continue. Conversely, RCTs without evidence of difference between groups may be terminated early. Under this assumption, the RMSTD are identifiable not at random. We induced the following selection mechanism. First, we generate data for an RCT and we sequentially test whether there is a statistically significant treatment effect, at 1 and 2 years, by a log-rank test with significance level 0.05. If there is a significant difference, we continue to the next time horizon. If there is no statistically significant difference, we continue to the next time horizon only with 30% probability according to a Bernoulli experiment. Otherwise, we retain the RMSTD estimates up to the current time horizon. We repeat this process until the generated trials have a follow-up pattern that matches the preset pattern, as specified in Table 1 and Table A1 of [supplementary material](#) available at *Biostatistics* online.

4.4. True marginal difference in restricted mean survival times

We derived the true RMSTD at time horizon, τ , by:

$$\begin{aligned}\delta(\tau) &= \int_0^\tau S\left(t, x_{\text{trt}} = \frac{1}{2}\right) dt - \int_0^\tau S\left(t, x_{\text{trt}} = -\frac{1}{2}\right) dt \\ &= \int_0^\tau \left[\int_\kappa S\left(t, x_{\text{trt}} = \frac{1}{2} | \alpha_j, \phi_j\right) dF(\alpha_j, \phi_j) \right] dt - \int_0^\tau \left[\int_\kappa S\left(t, x_{\text{trt}} = -\frac{1}{2} | \alpha_j, \phi_j\right) dF(\alpha_j, \phi_j) \right] dt,\end{aligned}\quad (4.6)$$

where $\kappa = (\alpha_k, \phi_k)_{k=1,2,\dots,K}$ is the set of all possible pairs of instances of α_k and ϕ_k ([Lueza and others, 2016b](#)). The joint distribution, $F(\alpha_j, \phi_j)$ is defined as the probability of a particular pair instance, p_κ . Because $F(\alpha_j, \phi_j)$ is a discrete probability distribution, (4.6) is equivalent to $\delta(\tau) = \sum_{k=1}^K p_k \delta_k(\tau)$, where the restricted mean survival time for each pair, (α_k, ϕ_k) is given by:

$$\delta_k(\tau) = \int_0^\tau \exp\left(-\left(\frac{t}{\lambda}\right)^v \exp\left(\alpha_j + \frac{1}{2}(\theta + \phi_j)\right)\right) dt - \int_0^\tau \exp\left(-\left(\frac{t}{\lambda}\right)^v \exp\left(\alpha_j - \frac{1}{2}(\theta + \phi_j)\right)\right) dt. \quad (4.7)$$

Note that this calculation is the marginal effect assuming that availability pattern of RMSTD at a particular time horizon is non-informative. Secondly, in the case of the piecewise Weibull data generation, the true RMSTD was computed separately over the two time intervals (0–1.5 years, 1.5–3 years) according to the respective values for θ and then summed together.

5. SIMULATION STUDY RESULTS

The nested loop plots in Figures 1 and 2 show the bias and mean squared error of the four methods for the 48 primary scenarios. Nested loop plots for coverage probability are provided in Figure A2 of [supplementary material](#) available at *Biostatistics* online. The simulated trial sample sizes ranged from 155 to 2935 with mean 702 and standard deviation 302. Results for the 12 additional scenarios are shown in Tables A2–A4 of [supplementary material](#) available at *Biostatistics* online.

5.1. Random follow-up pattern

For data generated under the random follow-up mechanism, at a time horizon of $\tau_1 = 1$ year, the average bias is very small, with estimates within half a day of the true RMSTD for all methods. The MSE is also extremely small for all scenarios and methods at $\tau_1 = 1$ year. For the time horizons of $\tau_2 = 2$ and $\tau_3 = 3$

Fig. 1. Simulation results for random follow-up pattern

The mean bias and mean squared error are shown in the left and right panels, respectively. At the top of each graph, we define the 24 scenarios and 3 time horizons (1, 2, and 3 years). At 1 year, the RMSTD is identifiable in all RCTs. At 2 years, it is identifiable in 80% of RCTs. At 3 years, it is identifiable in 60% of RCTs.

Fig. 2. Simulation results for non-random follow-up pattern

The mean bias and mean squared error are shown in the left and right panels, respectively. At the top of each graph, we define the 24 scenarios and 3 time horizons (1, 2, and 3 years). At 1 year, the RMSTD is identifiable in all RCTs. At 2 years, it is identifiable in 80% of RCTs. At 3 years, it is identifiable in 60% of RCTs.

years, there is a slight increase in MSE for all scenarios and methods but the bias remains very small, within 1 day of the true effect. Over the 48 primary scenarios and three time horizons, the univariate method has higher MSE than the other methods. Overall, there are no considerable differences, although the multivariate model with analytically derived covariance and the univariate model with predictions from flexible parametric models have a slight advantage in terms of bias. All methods have coverage probabilities between 91% and 98% for all scenarios (Figure A2 of [supplementary material](#) available at *Biostatistics* online).

As the availability of RCT data decreased (or τ increased), the bias and MSE of the estimates increased. Generation of data under non-proportional hazards lead to underestimation of the true effect while with proportional hazards, the methods overestimated the true effects but were generally less biased. Simulating data under an HR of 0.9 lead to slightly less biased estimates as compared with an HR of 0.7 but the MSE remained unchanged. Decreasing the number of RCTs from 30 to 10 to 5 did not have a large impact on the bias and MSE. Shifting between low and high between-trial variances resulted in only small shifts in the estimates.

5.2. *Non-random follow-up pattern*

There are more apparent differences in performance for data generated according to the non-random follow-up pattern. For a time horizon of $\tau_1 = 1$ year, the average bias of the combined effect estimates is very similar for all methods and remains low together with the MSE for all scenarios. At the time horizon of $\tau_2 = 2$ years, there are small differences between the methods in terms of bias, but very similar MSEs. In some scenarios, the univariate model using only available data appears to be the least biased but this may be explained by the high degree of availability of RMSTD across RCTs as well as a balance of over- and underestimation of the combined effect resulting in an average bias close to zero. For a time horizon of $\tau_3 = 3$ years, we observe some departures among the methods. Here, the multivariate model with analytic covariance has the advantage with respect to bias and MSE, in particular when the underlying hazard ratio is 0.7. In most scenarios, the four methods are in agreement with respect to over or underestimation of the true effect. For all scenarios and time horizons, the coverage probability is always highest for the multivariate approach with analytic covariance. For all scenarios, the multivariate model with bootstrap-derived covariance performs similarly to the univariate model with estimates from the flexible parametric models.

5.3. *Additional scenarios*

Using a larger sample size distribution does not alter the results. The mean bias and MSE remains small for all four methods and coverage probabilities hover around 95% and remain highest for the multivariate method with analytically derived covariance. Modifying the number of RCTs with available RMSTD at each time horizon results in slight increases in bias, although still negligible with respect to the respective marginal effect. When using maximum likelihood estimation instead of method of moments, we see slight increases in mean bias, up to 0.0004 years higher as compared with method of moments. Finally, when generating data according to a piecewise Weibull distribution, the mean bias increases substantially for all methods. For a meta-analysis of five trials at time horizon τ_3 , the univariate model based on flexible parametric model estimates has the highest mean bias, -0.0096 , while the multivariate model with analytically derived covariance has mean bias -0.0074 . This can be explained by the fact that these models may not have good fit due to the change in hazard function at 1.5 years. Additionally, in these scenarios the coverage probability decreases dramatically for a time horizon of $\tau_3 = 3$ years across all methods. The coverage probability remains highest for the multivariate model with analytically derived covariance. For example, in a meta-analysis of five trials at time horizon τ_3 , the coverage probabilities

for the method based on flexible parametric models and the multivariate method with analytically derived covariance are 89% and 96%, respectively.

6. ILLUSTRATIVE EXAMPLE

We illustrate the methods on a meta-analysis of RCTs comparing TAVR with surgical replacement in patients with aortic stenosis. Aortic stenosis is growing increasingly common especially due to changing population demographics (Thaden and others, 2014). According to recent models, the estimated number of high-risk aortic stenosis patients across the United States and Europe combined exceeds 300 000 (Osnabrugge and others, 2013).

STACCATO, the first RCT of TAVR vs. surgery, was terminated early after the inclusion of 70 patients because of an excess of events at 30 days in the TAVR group. Subsequently, five other RCTs have compared TAVR to surgery. All RCTs except one tested for non-inferiority of TAVR.

In a prior meta-analysis, Gargiulo and others (2016) found a combined odds ratio for all-cause death of 0.80 [95% confidence interval (CI) from 0.51 to 1.25] in favor of TAVR. Siemieniuk and others (2016) found a combined hazard ratio of 0.86 (95% CI from 0.74 to 1.01). Finally, Siontis and others (2016) found a combined hazard ratio of 0.87 (95% CI from 0.76 to 0.99). A limitation is that these syntheses combined trials with variable follow-up, resulting in a combined estimate that is difficult to interpret.

6.1. Methods

We performed a meta-analysis of five RCTs described in Table 2, by using the RMSTD as a measure of treatment effect on all-cause mortality. We reconstructed individual patient data for all RCTs using the published Kaplan–Meier curves (Guyot and others, 2012).

The NOTION (Søndergaard and others, 2016), Partner2 (Leon and others, 2016), and SURTAVI (Reardon and others, 2017) RCTs have identifiable RMSTD up to 24 months while in PARTNER (Mack and others, 2015) and USCOREValve (Deeb and others, 2016), the RMSTD is identifiable to at least 36 months. We use $\tau_1 = 12$, $\tau_2 = 24$, and $\tau_3 = 36$ months as clinically relevant time horizons of interest.

Using the reconstructed data for each RCT, we tested for non-proportional hazards using the Grambsch–Therneau test with the Kaplan–Meier transformation in each RCT ($\alpha = 0.10$) and we used the Fisher method of combined p-values to test for non-proportional hazards across all RCTs ($\alpha = 0.05$).

We estimated the RMSTD, by integrating the Kaplan–Meier estimator, and the associated standard error at each available time horizon of interest. For each RCT, we also fit the flexible parametric model of

Table 2. Trial characteristics

	PARTNER	NOTION	US CoreValve	PARTNER2	SURTAVI
Trial characteristics					
<i>n</i> randomized	699	280	750	2032	1746
Intention-to-treat analysis	Yes	No	No	Yes	No
Follow-up (years)	5	2	3	2	2
Surgical risk group	High risk	All-comers	High risk	Intermediate risk	Intermediate risk
Valve type	Sapien	CoreValve	CoreValve	Sapien XT	CoreValve, Evolut R
Access site [†]	Femoral, thoracic	Femoral	Femoral	Femoral, thoracic	Femoral
Age, mean (SD)	84 (7)	79 (5)	83 (7)	82 (7)	80 (6)
Male (%)	57	53	53	55	57

[†]Femoral corresponds to transfemoral approach, thoracic corresponds to transapical or transaortic approach.

Fig. 3. Differences in restricted mean survival times over time in the aortic stenosis trials. We show the RMSTD for each trial from 0 to 36 months. The solid lines represent the true calculated RMSTD while dashed lines show the RMSTD as predicted by the Royston–Parmar flexible parametric model. At each time horizon (12, 24, and 36 months), we include the meta-analytic estimate and CI of the combined treatment effect for all four methods.

Royston and Parmar with 3 d.f. for the baseline distribution and 1 d.f. for the time-dependent coefficient accounting for the possibility of non-proportional hazards. We used these models to estimate the RMSTD and SE at all time horizons for all RCTs. We analyzed the RCTs using the four meta-analytic methods described previously: univariate with available data, univariate with predictions from flexible parametric models, multivariate with analytical covariance, and multivariate with bootstrap covariance.

6.2. Results

According to the Grambsch–Therneau test, there is indication of non-proportional hazards in the PARTNER trial ($p < 0.001$) and the Fisher method to combine p-values (Grambsch–Therneau) found evidence of non-proportional hazards in at least one trial ($p = 0.0229$).

We report the observed RMSTD estimates and their meta-analysis at each time point in Figure A3 of [supplementary material](#) available at *Biostatistics* online. Figure 3 shows the RMSTD for each RCT over time, as well as the predictions from the flexible parametric models. At 12, 24, and 36 months, Figure 3 also shows the combined effect estimates from each of the meta-analytic methods.

At $\tau_1 = 12$ months, RMSTD is identifiable in all five RCTs. The multivariate meta-analysis with analytically derived covariance gives a combined RMSTD estimate of 0.28 months (95% CI from 0.01 to 0.56 months), indicating that over 1 year of follow-up, those treated by TAVR lived 0.28 months longer on average as compared with those treated by surgery. The result is statistically significant ($p = 0.05$). The univariate meta-analysis of the RMSTD predicted from the flexible parametric models gives a combined effect estimate of 0.21 months (95% CI from 0.03 to 0.44) but it is not statistically significant ($p = 0.08$).

Of note, the predicted RMSTD is smaller than the observed RMSTD by the fitted models in four of the five RCTs.

For a time horizon of $\tau_2 = 24$ months, the RMSTD is again estimable in all five RCTs. The combined effect from the multivariate meta-analysis with analytically derived covariance is 0.46 months (95% CI from -0.08 to 1.01 , $p = 0.09$).

At $\tau_3 = 36$ months, the RMSTD is identifiable in only two RCTs, PARTNER and USCorevalve. The multivariate meta-analysis with analytically derived covariance gives an estimated combined effect of 0.79 months (95% CI from the flexible parametric models gives an estimate of 0.54 months (95% CI from -0.05 to 1.12 , $p = 0.07$). The univariate method gives an estimate of 1.45 months (95% CI from 0.13 to 2.76 , $p = 0.03$). This larger combined estimate can be explained by the fact that at 36 months, only two RCTs contribute data to this approach and both have the greatest RMSTD estimates over the 36 months.

In general, the results based on the flexible parametric models may be biased due to poor precision on the RMSTD estimation, especially in the Partner trial. At 36 months and for a significance level of 0.05, only the univariate method shows evidence of a statistically significant result.

7. DISCUSSION

We have introduced a new model for individual patient data meta-analyses of RCTs using the RMSTD as a measure of treatment effect. We showed that our multivariate random-effect model lead to estimates with comparable bias and MSE as other methods. However, our model offers a key advantage in that it uses all available evidence at all time horizons of interest. Previously described univariate meta-analysis methods rely on predictions from a parametric model or do not make full use of available data.

The multivariate model is based upon the specification of a set of clinically relevant time horizons of interest. Including additional time horizons may improve the model. However, this would come with the requirement to estimate more values for the within-trial covariance.

This work has several limitations. First, our model, like previously described methods, requires individual patient data. In particular, it is an essential part of our approach to estimate the covariance matrix. However, as shown in our illustrative example, numerical methods have been validated to reconstruct individual survival data from the Kaplan–Meier curves. These methods have high levels of accuracy and reproducibility (Guyot and others, 2012; Trinquart and others, 2016). Moreover, we use the same reconstructed data for all four meta-analytic methods in our illustrative example and therefore, any discrepancy would remain consistent. Second, we pursued the DerSimonian and Laird random-effects model for the univariate and multivariate models. We explored additional scenarios with the maximum likelihood estimator and we found very similar results (Tables A2–A4 of [supplementary material](#) available at *Biostatistics* online). While the method of moments has been criticized in the univariate setting, its use in multivariate setting has not been as fully evaluated (Brockwell and Gordon, 2001; Hardy and Thompson, 1996). Future work could consider other estimators for the between-trial covariance. We implemented a bootstrap method to derive the within-trial covariance. It is computationally more intensive than the other methods, especially for a high number of samples. On the other hand, the multivariate meta-analysis with analytically derived covariance requires less computation time load.

Our simulation study also has limitations. First, we generated data according to a Weibull model, which favors the fit of the flexible parametric models described by Royston and Parmar. The univariate meta-analysis of estimates from these models may have performed differently if the models did not have a good fit. However, we explore additional scenarios (A7–A9) with data generated according to a piecewise Weibull for which the flexible parametric model has poorer fit. The results showed that the multivariate model with analytically derived covariance was preferred with respect to bias and coverage. Secondly, our calculation of the true marginal RMSTD relied implicitly on the assumption that missingness was non-informative. Thus, there is no direct comparison of a true effect for data generated according to our

non-random follow-up mechanism. Instead, we could only evaluate how well the models estimated the truth according to the marginal RMSTD assuming random follow-up patterns.

To our knowledge, our illustrative example is the first meta-analysis of RCTs comparing TAVR and surgery in patients with aortic stenosis to use RMSTD as the measure of treatment effect. Previous meta-analyses combined trials with variable follow-up by using the hazard ratio. However, our analyses suggest evidence of non-proportional hazards, which makes the combined hazard ratio difficult to understand. On the contrary, we provided quantification of the treatment effect on the time scale. Two previous meta-analyses suggested no difference between TAVR and surgery and another one concluded that TAVR was superior to surgery for mortality. We have previously shown the value of the RMSTD for non-inferiority trials (Weir and Trinquart, 2018). Our meta-analysis did not find evidence that TAVR is superior to surgery for all-cause mortality, with trivial gains in life-expectancy.

8. SOFTWARE

We developed an R package, metaRMST (available from <https://CRAN.R-project.org/package=metaRMST> (date last accessed 20 May 2019)), for meta-analysis of RMSTD that includes each of the described methods and the data for the aortic stenosis RCTs (Weir and others, 2018b). In addition, we provide R scripts for reproducibility of the illustrative example results (available from https://github.com/iweir/MVMA_RMST_aortic_stenosis_analysis (date last accessed 20 May 2019)).

SUPPLEMENTARY MATERIAL

Supplementary material is available at <http://biostatistics.oxfordjournals.org>.

ACKNOWLEDGMENTS

We thank Katia Oleinik (Boston University) for her help with the BU Shared Computing Cluster, and Mark Clements (Karolinska Institutet) for implementation of RMST calculations in his R package, rstpm2. We thank Martin Leon and Maria Alu for providing us with a Kaplan–Meier curve for the Partner2 trial.

Conflict of Interest: None declared.

FUNDING

This work is funded by an NIH grant, National Institute of General Medical Sciences (NIGMS) Interdisciplinary Training Grant for Biostatisticians (T32 GM-074905).

REFERENCES

- BARIANI, G. M., DE CELIS, F., ANEZKA, C. R., PRECIVALE, M., ARAI, R., SAAD, E. D. AND RIECHELMANN, R. P. (2015). Sample size calculation in oncology trials. *American Journal of Clinical Oncology* **38**, 570–574.
- BENDER, R., AUGUSTIN, T. AND BLETNER, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine* **24**, 1713–1723.
- BROCKWELL, S. E. AND GORDON, I. R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in Medicine* **20**, 825–840.
- CHEN, H. M., ALISA, K. AND DUPUIS, J. (2012). A method of moments estimator for random effect multivariate meta-analysis. *Biometrics* **68**, 1278–1284.
- CLEMENTS, M. AND LIU, X. R. (2018). *rstpm2: Generalized Survival Models*. R Package Version 1.4.2.

- DAVEY, J., TURNER, R. M., CLARKE, M. J. AND HIGGINS, J. P. T. (2011). Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: a cross-sectional, descriptive analysis. *BMC Medical Research Methodology* **11**, 160.
- DEEB, G. M., REARDON, M. J., CHETCUTI, S., PATEL, H. J., GROSSMAN, P. M., YAKUBOV, S. J., KLEIMAN, N. S., COSELLI, J. S., GLEASON, T. G., LEE, J. S. *and others*. (2016). 3-year outcomes in high-risk patients who underwent surgical or transcatheter aortic valve replacement. *Journal of the American College of Cardiology* **67**, 2565–2574.
- EFRON, B. AND TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*, 1st edition. Boca Raton, FL: CRC Press.
- GARGIULO, G., SANNINO, A., CAPODANNO, D., BARBANTI, M., BUCCHERI, S., PERRINO, C., CAPRANZANO, P., INDOLFI, C., TRIMARCO, B., TAMBURINO, C. *and others*. (2016). Transcatheter aortic valve implantation versus surgical aortic valve replacement a systematic review and meta-analysis systematic review and meta-analysis of TAVI versus SAVR. *Annals of Internal Medicine* **165**, 334–344.
- GUYOT, P., ADES, A.E., OUWENS, M. J. N. M. AND WELTON, N. J. (2012). Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan–Meier survival curves. *BMC Medical Research Methodology* **12**, 9.
- HARDY, R. J. AND THOMPSON, S. G. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* **15**, 619–629.
- JACKSON, D., WHITE, I. R., PRICE, M., COPAS, J. AND RILEY, R. D. (2017). Borrowing of strength and study weights in multivariate and network meta-analysis. *Statistical Methods in Medical Research* **26**, 2853–2868.
- JACKSON, D., WHITE, I. R. AND THOMPSON, S. G. (2010). Extending DerSimonian and Laird’s methodology to perform multivariate random effects meta-analyses. *Statistics in Medicine* **29**, 1282–1297.
- LEON, M. B., SMITH, C. R., MACK, M. J., MAKKAR, R. R., SVENSSON, L. G., KODALI, S. K., THOURANI, V. H., TUZCU, E. M., MILLER, D. C., HERRMANN, H. C. *and others*. (2016). Transcatheter or surgical aortic-valve replacement in intermediate-risk patients. *New England Journal of Medicine* **2016**, 1609–1620.
- LUEZA, B., MAUGUEN, A., PIGNON, J., RIVERO-ARIAS, O., BONASTRE, J., MAR-LC Collaborative Group. (2016a). Difference in restricted mean survival time for cost-effectiveness analysis using individual patient data meta-analysis: evidence from a case study. *PLoS One* **11**, e0150032.
- LUEZA, B., ROTOLO, F., BONASTRE, J., PIGNON, J. AND MICHIELS, S. (2016b). Bias and precision of methods for estimating the difference in restricted mean survival time from an individual patient data meta-analysis. *BMC Medical Research Methodology* **16**, 37.
- MACK, M. J., LEON, M. B., SMITH, C. R., MILLER, D. C., MOSES, J. W., TUZCU, E. M., WEBB, J. G., DOUGLAS, P. S., ANDERSON, W. N., BLACKSTONE, E. H. *and others*. (2015). 5-year outcomes of transcatheter aortic valve replacement or surgical aortic valve replacement for high surgical risk patients with aortic stenosis (partner 1): a randomised controlled trial. *Lancet* **385**, 2477–2484.
- OSNABRUGGE, R. L. J., MYLOTTE, D., HEAD, S. J., VAN MIEGHEM, N. M., NKOMO, V. T., LEREUN, C. M., BOGERS, A. J. J. C., PIAZZA, N. AND KAPPETEIN, A. P. (2013). Aortic stenosis in the elderly: disease prevalence and number of candidates for transcatheter aortic valve replacement: a meta-analysis and modeling study. *Journal of the American College of Cardiology* **62**, 1002–1012.
- PEPE, M. S. AND FLEMING, T. R. (1989). Weighted Kaplan–Meier statistics: a class of distance tests for censored survival data. *Biometrics* **45**, 497–507.
- PEPE, M. S. AND FLEMING, T. R. (1991). Weighted Kaplan–Meier statistics: large sample and optimality considerations. *Journal of the Royal Statistical Society: Series B (Methodological)* **53**, 341–352.
- PESTINE, E., STOKES, A. AND TRINQUART, L. (2018). Representation of obese participants in obesity-related cancer randomized trials. *Annals of Oncology* **29**, 1582–1587.
- REARDON, M. J., VAN MIEGHEM, N. M., POPMA, J. J., KLEIMAN, N. S., SØNDERGAARD, L., MUMTAZ, M., ADAMS, D. H., DEEB, G. M., MAINI, B., GADA, H. *and others*. (2017). Surgical or transcatheter aortic-valve replacement in intermediate-risk patients. *New England Journal of Medicine* **376**, 1321–1331.

- ROYSTON, P. AND PARMAR, M. K. B. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* **21**, 2175–2197.
- ROYSTON, P. AND PARMAR, M. K. B. (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.
- RÜCKER, G. AND SCHWARZER, G. (2014). Presenting simulation results in a nested loop plot. *BMC Medical Research Methodology* **14**, 129.
- SIEMIENIUK, R. A., AGORITSAS, T., MANJA, V., DEVJI, T., CHANG, Y., BALA, M. M., THABANE, L. AND GUYATT, G. H. (2016). Transcatheter versus surgical aortic valve replacement in patients with severe aortic stenosis at low and intermediate risk: systematic review and meta-analysis. *BMJ* **354**, i5130.
- SIONTIS, G. C. M., PRAZ, F., PILGRIM, T., MAVRIDIS, D., VERMA, S., SALANTI, G., SØNDERGAARD, L., JÜNI, P. AND WINDECKER, S. (2016). Transcatheter aortic valve implantation vs. surgical aortic valve replacement for treatment of severe aortic stenosis: a meta-analysis of randomized trials. *European Heart Journal* **37**, 3503–3512.
- SMITH, C. T., WILLIAMSON, P. R. AND MARSON, A. G. (2005). Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Statistics in Medicine* **24**, 1307–1319.
- SØNDERGAARD, L., STEINBRÜCHEL, D. A., IHLEMANN, N., NISSEN, H., KJELDSSEN, B. J., PETURSSON, P., NGO, A. T., OLSEN, N. T., CHANG, Y., FRANZEN, O. W. and others. (2016). Two-year outcomes in patients with severe aortic valve stenosis randomized to transcatheter versus surgical aortic valve replacement. *Circulation: Cardiovascular Interventions* **6**, e003665.
- THADEN, J. J., NKOMO, V. T. AND ENRIQUEZ-SARANO, M. (2014). The global burden of aortic stenosis. *Progress in Cardiovascular Diseases* **56**, 565–571.
- TIAN, L., FU, H., RUBERG, S. J., UNO, H. AND WEI, L. J. (2017). Efficiency of two sample tests via the restricted mean survival time for analyzing event time observations. *Biometrics* **74**, 694–702.
- TRINQUART, L., IOANNIDIS, J. P. A., CHATELLIER, G. AND RAVAUD, P. (2014). A test for reporting bias in trial networks: simulation and case studies. *BMC Medical Research Methodology* **14**, 112.
- TRINQUART, L., JACOT, J., CONNER, S. C. AND PORCHER, R. (2016). Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. *Journal of Clinical Oncology* **34**, 1813–1819.
- UNO, H., CLAGGETT, B., TIAN, L., INOUE, E., GALLO, P., MIYATA, T., SCHRAG, D., TAKEUCHI, M., UYAMA, Y., ZHAO, L. and others. (2014). Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of Clinical Oncology* **32**, 2380.
- UNO, H., TIAN, L., CRONIN, A., BATTIOUI, C. AND HORIGUCHI, M. (2017). *Vignette for survRM2 Package: Comparing Restricted Mean Survival Time*. R Package Version 1.0-2.
- WEI, Y., ROYSTON, P., TIERNEY, J. F. AND PARMAR, M. K. B. (2015). Meta-analysis of time-to-event outcomes from randomized trials using restricted mean survival time: application to individual participant data. *Statistics in Medicine* **34**, 2881–2898.
- WEIR, I. R., MARSHALL, G., SCHNEIDER, J., SHERER, J., LORD, E. M., GYAWALI, B., PAASCHE-ORLOW, M., BENJAMIN, E. AND TRINQUART, L. (2018a). Interpretation of time-to-event outcomes in randomized trials: an online randomized experiment. *Annals of Oncology* **30**, 96–102.
- WEIR, I. R., TIAN, L. AND TRINQUART, L. (2018b). *metaRMST: Meta-Analysis of RMSTD*. R Package Version 1.0.0.
- WEIR, I. R. AND TRINQUART, L. (2018). Design of non-inferiority randomized trials using the difference in restricted mean survival times. *Clinical Trials* **15**, 499–508.