Asymptotic properties of mean survival estimate based on the Kaplan–Meier curve with an extrapolated tail

Qi Gong[a] and Liang Fang[b]*

Asymptotic distribution of the mean survival time based on the Kaplan–Meier curve with an extrapolated 'tail' is derived. A closed formula of the variance estimate is provided. Asymptotic properties of the estimator were studied in a simulation study, which showed that this estimator was unbiased with proper coverage probability and followed a normal distribution. An example is used to demonstrate the application of this estimator. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: mean survival; restricted mean survival; extrapolation; variance estimate; quality-adjusted lifetime

1. INTRODUCTION

In many clinical trials, mean survival is difficult to estimate because of censoring or insufficient follow-up. As an alternative, Irwin [1] proposed restricted mean survival time to estimate expected lifetime restricted to a chosen time point T. Restricted mean survival time is defined as the integration of survival function from time 0 to a chosen time point T, which essentially is the area under the survival curve from time 0 to T. Test statistics based on the difference of two restricted mean survival times were derived to compare the distributions of survival time from two treatment groups as an alternative test under the situation of non-proportional hazard ratio [2–7]. Restricted mean survival is also widely used to estimate quality-adjusted survival in patient outcome research [8, 9].

The choice of restriction time T is often subjective and hard to justify. Karrison [2] discussed a method to choose T on the basis of the sample size. Literatures in quality-adjusted survival [8, 9] suggest choosing median follow-up time as T. However, in practice, the expected survival time (i.e., mean survival time) is much more meaningful for patients than the expected survival time up to a chosen time point T (i.e., restricted mean survival time). Estimating mean survival time is possible if the last observation on the study was an event and survival curve drops to 0 at the end, although the estimation could be unstable because of the small number of risk set at the tail of the survival curve. When the last observation on the survival curve is censored and the curve does not drop to 0 at the end, Efron [10] assumed 0 beyond the last observation. Moeschberger and Klein [11] and Klein *et al.* [12] investigated ways to extrapolate the Kaplan–Meier survival curve to 0. The extrapolation methods add a 'tail' to the Kaplan–Meier curve beyond the last observed time point. The 'tail' is estimated by a parametric function fit to the data. In other words, the tail is imputed on the basis of the parametric survival function estimated from the data. Gelber *et al.* [13] proposed estimating mean survival on the basis of the extrapolated Kaplan–Meier survival curve. Commercial software Stata (Stata

Corp., College Station, TX, USA) provides an option to calculate the mean survival using this extrapolation method without providing the variance estimate (www.stata.com/help.cgi?stci). However, to our knowledge, the statistical inference of mean survival estimator based on the extrapolated Kaplan–Meier curve has not been derived. Especially, the variance estimate of such estimator was obtained using a bootstrap method because no closed formula was available. In this article, the estimator of the mean survival time based on extrapolated Kaplan–Meier survival curve is referred to as the hybrid estimator because it includes a non-parametric estimate of survival and a parametric extrapolation of the tail beyond the last observation of the survival curve. Statistical inferences of the hybrid estimator, including the closed form of the variance estimate, are derived, and asymptotic properties are evaluated.

The article is organized as follows. In Section 2, notations are defined, and the hybrid estimator and its statistical inferences are derived. A simulation study is reported in Section 3 to evaluate the asymptotic properties of the hybrid estimator. In Section 4, we apply our methods to estimate the mean survival time and treatment effects from an oncology clinical trial. A discussion and summary is presented in Section 5.

2. NOTATIONS AND METHODS

Let μ denote the mean survival $\mu = \int_0^\infty S(t)dt$, where $S(t)$ is the true survival function for event time. Also, define $\mu = \mu_{KM} + \mu_P$, where μ_{KM} is restricted mean survival from time 0 up to the biggest observation time τ, and μ_P denotes the integration

[a]*Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA*

[b]*Genentech, Inc., South San Francisco, CA, USA*

*Correspondence to: Liang Fang, 1 DNA Way, MS 442A, South San Francisco, CA, 94080, USA.
E-mail: fang.liang@gene.com

under the $S(t)$ from τ to infinity. That is, $\mu_{KM} = \int_0^\tau S(t)dt$ and $\mu_P = \int_\tau^\infty S(t)dt$. $\hat{S}_{KM}(t)$ is the Kaplan–Meier estimate of $S(t)$ in time interval $[0, \tau]$. $\hat{S}_P(t)$ is the survival estimate of $S(t)$ in interval $[\tau, \infty]$ based on a parametric distribution, $f_P(t)$, fit to the data. For example, if the survival data are assumed to follow an exponential function with parameter λ (i.e., $f_P(t) = \lambda e^{-\lambda t}$), the estimate of the survival function $\hat{S}_P(t) = e^{-\hat\lambda t}$, where $\hat\lambda$ is the maximum likelihood estimator of λ. When replacing $S(t)$ with its estimates $\hat{S}_{KM}(t)$ and $\hat{S}_P(t)$ on corresponding intervals, we obtain $\hat\mu_{KM} = \int_0^\tau \hat{S}_{KM}(t)dt$ and $\hat\mu_P = \int_\tau^\infty \hat{S}_P(t)dt$ naturally. Thus, the hybrid mean survival estimator of μ is defined as

$$\hat\mu_H = \hat\mu_{KM} + \hat\mu_P = \int_0^\tau \hat{S}_{KM}(t)dt + \int_\tau^\infty \hat{S}_p(t)dt. \qquad (2.1)$$

In fact, one could choose the adjacent time point to be a time point (t_1) earlier than the last observed survival time (τ, $t_1 < \tau$) and let the parametric extrapolation start at t_1 because the Kaplan–Meier tail estimate is unstable due to few available data. Nevertheless, the results in this section can be extended naturally to obtain the variance estimate.

Let T_i denote the survival time, and C_i is the censoring time for the ith subject. Assume C_i is independent of T_i. The observed time is given by $X_i = \min(T_i, C_i)$, and the censoring indicator $\Delta_i = I(T_i = X_i)$, $i = 1, \ldots, n$.

To estimate μ_{KM}, we use counting process notations. Let $Y(t)$ be the number of subjects at risk at time t and $N_i(t)$ be the count of events at time t for the ith subject. $\Lambda(t)$ denotes the accumulate baseline hazard function.

It can be shown that

$$\hat{S}_{KM}(t) - S(t) = -\sum_{i=1}^n S(t) \int_0^t \frac{\hat{S}_{KM}(u-)I(Y(u) > 0)}{S(u)Y(u)} dM_i(u)$$
$$= \sum_{i=1}^n \eta_i(t),$$

where $\eta_i(t) = -S(t) \int_0^t \frac{\hat{S}_{KM}(u-)I(Y(u) > 0)}{S(u)Y(u)} dM_i(u)$ and $dM_i(u) = dN_i(u) - Y_i(u)d\Lambda(u) = dN_i(u) - Y_i(u)dS(u)/S(u-)$ [14]. Thus, $\hat\mu_{KM} - \mu_{KM} = \int_0^\tau \left[\hat{S}_{KM}(t) - S(t)\right] dt = \sum_{i=1}^n \varphi_i$, where $\varphi_i = -\int_0^\tau S(t) \int_0^t \frac{\hat{S}_{KM}(u-)I(Y(u) > 0)}{S(u)Y(u)} dM_i(u)dt$.

Under some regular conditions [15], $\hat\mu_{KM} - \mu_{KM} \sim N(0, \sigma_{11})$ where $\sigma_{11} = \frac{1}{n} \int_0^\tau \left[\int_t^\infty S(u)du\right]^2 \frac{d\Lambda(t)}{S(t-)S^C(t-)}$, and $S^C(t)$ is the true survival function for the censoring time.

To estimate μ_P, we define β as the parameter vector of the distribution fit to data. To estimate β, we write the likelihood function as $L = \prod_{\Delta_i=1} f_P(X_i) \prod_{\Delta_i=0} S_P(X_i)$ [16]. Score function is $U(\beta) = \partial \log(L)/\partial\beta$, and observed Fisher information is $I(\beta) = -\partial^2 \log(L)/\partial\beta\partial\beta^T$. Thus, it can be written that $\hat\beta - \beta = I^{-1}(\beta)U(\hat\beta) = \sum_{i=1}^n \psi_i$, where $\psi_i = I^{-1}(\beta)U_i(\hat\beta)$, and $\hat\beta - \beta \sim N(0, \sigma_{22})$, where $\sigma_{22} = I^{-1}(\beta)$.

Because $\eta_i(t)$, ϕ_i, and ψ_i are all independent and identically distributed with zero expectation, the covariance coefficient for the Kaplan–Meier and parametric extrapolated tail can be derived as $\sigma_{12} = \text{cov}\left(\hat\mu_{KM} - \mu_{KM}, \hat\beta - \beta\right) = \text{cov}\left(\sum_{i=1}^n \varphi_i, \sum_{i=1}^n \psi_i\right) = \sum_{i=1}^n \varphi_i\psi_i$. Thus, we have

$$\begin{pmatrix} \hat\mu_{KM} - \mu_{KM} \\ \hat\beta - \beta \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12}^T & \sigma_{22} \end{pmatrix}\right]. \qquad (2.2)$$

Applying function delta method, $\hat\mu_H = \hat\mu_{KM} + \hat\mu_P = \hat\mu_{KM} + \int_\tau^\infty \hat{S}_P(t; \hat\beta) dt$. Because we assume as $\hat\beta \to \beta$, $\hat{S}_P(t; \hat\beta) \to S(t)$. We have $\hat\mu_P \to \mu_P$ and can obtain $\hat\mu_H - \mu \sim N(0, \sigma_H)$, where

$$\sigma_H = \sigma_{11} + 2\sigma_{12}\frac{\partial\hat\mu_P}{\partial\beta} + \frac{\partial\hat\mu_P^T}{\partial\beta}\sigma_{22}\frac{\partial\hat\mu_P}{\partial\beta}.$$

All the terms mentioned in the preceding text can be consistently estimated by their sample analogs. The hybrid mean survival hypothesis testing is based on Z-test.

However, as discussed in many literatures [8, 17], the parametric function estimated from the data usually has poor fitting at the tail of the survival curve. This is because the majority of the data occur at the early part of the curve and drive the overall fitting of the parametric function. For example, a Weibull distribution can fit the data well when judging the goodness-of-fit from the entire course of the survival curve. However, the local fitting at the tail may be poor.

In such cases, we define another hybrid estimator of mean survival μ

$$\hat\mu_H^* = \hat\mu_{KM} + \hat\mu_P^* = \hat\mu_{KM} + \int_\tau^\infty \hat{S}_P^*(t; t_0, \beta^*)dt, \qquad (2.3)$$

where $\hat{S}_P^*(t; t_0, \hat\beta^*) = \hat{S}_{KM}(t_0)\hat{S}_P(t; \hat\beta^*|t > t_0)$. This estimator differs from (2.1) in such a way that the extrapolated tail is estimated via data beyond time point t_0, conditional on $t > t_0$. In other words, the parametric tail is estimated 'locally' and thus results in a better fit to the tail of the survival curve. This parametric tail is then attached to the Kaplan–Meier estimate $\hat{S}_{KM}(t_0)$ at time t_0 to obtain the estimated survival function for the period of time $t > t_0$: $\hat{S}_P^*(t; t_0, \hat\beta^*) = \hat{S}_{KM}(t_0)\hat{S}_P(t; \hat\beta^*|t > t_0)$. We define β^* as the parameter vector of the distribution fit to local data. To estimate β^*, we write the likelihood function as $L^* = \prod_{\Delta_i=1} f_P(X_i|X_i > t_0)I(X_i > t_0) \prod_{\Delta_i=0} S_P(X_i|X_i > t_0)I(X_i > t_0)$ [16]. Score function is $U^*(\beta^*) = \partial \log(L^*)/\partial\beta^*$, and observed Fisher information is $I^*(\beta^*) = -\partial^2 \log(L^*)/\partial\beta^*\partial\beta^{*T}$. Thus, it can be written that $\hat\beta^* - \beta^* = I^{*-1}(\beta)U^*(\hat\beta^*) = \sum_{i=1}^n \psi_i^*$, where $\psi_i^* = I^{*-1}(\beta^*)U_i^*(\hat\beta^*)$, and $\hat\beta^* - \beta^* \sim N(0, \sigma_{22}^*)$, where $\sigma_{22}^* = I^{*-1}(\beta^*)$.

For example, if we fit the local data with exponential distribution $S(t) = \exp\{-\lambda t\}$, the parametric part in (2.3) is $\hat{S}_P^*(t; t_0, \hat\lambda^*) = \hat{S}_{KM}(t_0)\exp\{-\hat\lambda^*(t - t_0)\}$. When constructing $\hat{S}_P^*(t; t_0, \hat\beta^*)$, the parametric piece $\hat{S}_P(t; \hat\beta^*|t > t_0)$ is connected to the Kaplan–Meier estimate $\hat{S}_{KM}(t_0)$ at time t_0. However, when the extrapolated trail is added to the Kaplan–Meier curve, the tail does not necessary go through the Kaplan–Meier estimate $\hat{S}_{KM}(\tau)$ at time τ. In other words, the end of the Kaplan–Meier curve and the start of the extrapolated tail can, but are not required to, connect to each other.

This hybrid estimator has been proposed by Su and Fang [17] and Gelber *et al.* [8] without providing detailed asymptotic properties. We derive the asymptotic properties of this estimator here. Similar to (2.2), we have $\begin{pmatrix} \hat{\mu}_{KM} - \mu_{KM} \\ \hat{\beta}^* - \beta^* \\ \hat{S}_{KM}(t_0) - S_{KM}(t_0) \end{pmatrix} \sim$

$$N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^* & \sigma_{12}^* & \sigma_{13}^* \\ \sigma_{12}^{*T} & \sigma_{22}^* & \sigma_{23}^* \\ \sigma_{13}^{*T} & \sigma_{23}^{*T} & \sigma_{33}^* \end{pmatrix} \right], \text{ where}$$

$\sigma_{11}^* = \sigma_{11},$

$\sigma_{12}^* = \text{cov}\left(\hat{\mu}_{KM} - \mu_{KM}, \hat{\beta}^* - \beta^* \right) = \text{cov}\left(\sum_{i=1}^{n} \varphi_i, \sum_{i=1}^{n} \psi_i^* \right)$

$\quad = \sum_{i=1}^{n} \varphi_i \psi_i^*,$

$\sigma_{13}^* = \text{cov}\left(\hat{\mu}_{KM} - \mu_{KM}, \hat{S}_{KM}(t_0) - S_{KM}(t_0) \right)$

$\quad = \text{cov}\left(\sum_{i=1}^{n} \varphi_i, \sum_{i=1}^{n} \eta_i(t_0) \right) = \sum_{i=1}^{n} \varphi_i \eta_i(t_0),$

$\sigma_{22}^* = I^{*-1}(\beta^*),$

$\sigma_{23}^* = \text{cov}\left(\hat{\beta}^* - \beta^*, \hat{S}_{KM}(t_0) - S_{KM}(t_0) \right)$

$\quad = \text{cov}\left(\sum_{i=1}^{n} \psi_i^*, \sum_{i=1}^{n} \eta_i(t_0) \right) = \sum_{i=1}^{n} \psi_i^* \eta_i(t_0),$

$\sigma_{33}^* = S(t_0)^2 \sum_{i=1}^{n} \int_0^{t_0} dN_i(t)/\{Y_i(t)[Y_i(t) - dN_i(t)]\}.$

$\psi_i^* = I^{*-1}(\beta^*) U_i^*\left(\hat{\beta}^* \right).$

Applying function delta method, $\hat{\mu}_H^* = \hat{\mu}_{KM} + \hat{\mu}_P^* = \hat{\mu}_{KM} + \int_\tau^\infty \hat{S}_P^*(t; t_0, \hat{\beta}^*) dt$. As $\hat{\beta}^* \to \beta^*$, $\hat{S}_P^*(t; t_0, \hat{\beta}^*) \to S(t|t > t_0)$. We have $\hat{\mu}_P^* \to \mu_P$ and can obtain $\hat{\mu}_H^* - \mu \sim N(0, \sigma_H^*)$, where

$$\sigma_H^* = \sigma_{11}^* + 2\sigma_{12}^* \frac{\partial \hat{\mu}_P^*}{\partial \beta^*} + \frac{\partial \hat{\mu}_P^{*T}}{\partial \beta^*} \sigma_{22}^* \frac{\partial \hat{\mu}_P^*}{\partial \beta^*} + 2\sigma_{13}^* \frac{\partial \hat{\mu}_P^*}{\partial S_{KM}(t_0)}$$

$$+ 2\frac{\partial \hat{\mu}_P^{*T}}{\partial \beta^*} \sigma_{23}^* \frac{\partial \hat{\mu}_P^*}{\partial S_{KM}(t_0)} + \frac{\partial \hat{\mu}_P^{*T}}{\partial S_{KM}(t_0)} \sigma_{33}^* \frac{\partial \hat{\mu}_P^*}{\partial S_{KM}(t_0)}$$

Similarly, all terms mentioned in the preceding text can be estimated consistently by their sample counterparts. *Z*-test could be used for the hypothesis testing.

3. SIMULATION STUDY

We conducted a simulation study to assess the asymptotic properties of the hybrid estimator in Section 2. Data were generated randomly from four scenarios. In the first scenario, the data were generated from a Weibull distribution $S(t) = \exp\{-(t/e^\lambda)^{1/\gamma}\}$ with parameters $\lambda = 1.5$ and $\gamma = 1.2$. In the second and fourth scenarios, the data were created from a piecewise Weibull distribution with parameters $\lambda_1 = 2.2$ and $\gamma_1 = 2.0$ when $t < 5$, and $\lambda_2 = 2.5$ and $\gamma_2 = 0.4$ when $t \geq 5$. In scenario 3, data were from an exponential distribution $S(t) = \exp\{-\lambda t\}$ with $\lambda = 0.2$. Event time T_i ($n = 100, 200, 300$) was generated from one of these distributions. Independent censoring C_i was generated from uniform distribution between (a, b), where a and b values vary (Table I) to achieve 20% and 30% censoring. The observed survival time was defined as $X_i = \min(T_i, C_i)$. Any T_i greater than b was censored because $C_i \in (a, b)$. For each sample, if $T_i < b$ for all i, then the adjacent time $\tau = \max(X_i) < b$. Otherwise, if one or more than one $T_i \geq b$, then $\tau = \max(X_i) = b$. As a result, the adjacent

time τ may vary among replications but would always be less than or equal to b. t_0 and τ are different. Time t_0 is the cutoff beyond which a chosen parametric form is fit in (2.3). Time τ is the maximum observation time, from which the extrapolated tail starts and goes to infinity. One thousand replications were generated for each scenario.

For all four scenarios, the true distribution forms were used as $S_P(t)$ to fit the curve. For scenarios 1 and 2, $S_P(t)$ were estimated on the basis of all data. In other words, hybrid estimator (2.1) was used to estimate mean survival time. Note that, for scenario 2, we fit a piecewise Weibull distribution to all data to obtain $S_P(t)$, and we did not force the tail to be adjacent to the Kaplan–Meier curve. Therefore, the mean survival was still estimated using estimator (2.1). For scenarios 3 and 4, only the data beyond time $t_0 = 5$ were used to fit $S_P(t; \beta^*|t > t_0)$, and then the parametric form was attached to the Kaplan–Meier estimate at time t_0, $S_{KM}(t_0)$, to form the final function $S_P^*(t; t_0, \beta^*) = S_{KM}(t_0)S_P(t; \beta^*|t > t_0)$. In other words, hybrid estimator (2.3) was used to estimate mean survival time. For all four scenarios, the adjacent time point τ is chosen to be the maximum survival time observed. The percent bias, coverage probability, empirical standard error (ESE), and asymptotic standard error (ASE) of the mean survival estimate were calculated (Table I). ESE is the sample standard error of mean lifetime estimate across all replications. ASE is the average of the derived asymptotic standard error based on our method from each of the replications.

In all four scenarios, the biases of the hybrid estimator were close to 0 ($< \pm 3\%$). The ESE were closed to the ASE. The coverage probabilities were close to the desired level of 95%. As the sample size went up or the censoring rate went down, the bias and the standard error became smaller and smaller. The choices of 75 percentile and 95 percentile time as the adjacent time point τ resulted in similar findings (data not shown). Therefore, the asymptotic properties derived in Section 2 were verified.

To study the robustness of the hybrid estimator to mis-specification of the underlying data distribution, we generated data from Weibull (scenarios 5 and 6, Table II), log-logistic (scenario 7), and log-normal (scenario 8) distributions and estimated mean survival time by using the hybrid estimator (2.3) with exponential distribution fit to the local data at the tail. Log-logistic survival function was expressed as $S(t) = 1/\{1 + \exp(-\lambda/\gamma)t^{1/\gamma}\}$. Log-normal density function was shown as $f(t) = \exp\{-[\log(t) - \lambda]^2/2\gamma\}/\sqrt{2\pi\gamma}$. It was already shown in Su and Fang [17] that the hybrid estimator (2.1) was not robust to mis-specification of the underlying data distribution and therefore not investigated here.

For all four scenarios, the percent biases of the mean survival were small ($<5.6\%$). The ESE was close to ASE, which indicated that the derived formulas of the asymptotic of the hybrid estimator were valid. However, the coverage probabilities were less than 95% for log-logistic and log-normal distributions and associated with larger biases (-3.21% and -5.56%, respectively). These results suggest that the hybrid estimator (2.3) is fairly robust to mis-specification of the underlying data distribution, although the biases in parameter estimates are inevitable.

4. EXAMPLE

As an illustration, we estimated the mean survival time with the hybrid estimator for data generated from a randomized, placebo-controlled, multi-center, phase III clinical trial. Patients in the control arm received drug A. Patients in the treatment arm received drug B. The primary end point in the trials was a time-to-event

Table I. Bias and coverage probability of the hybrid mean survival estimator in the simulation study.

True distribution	n	Censoring %	Censoring interval (a, b)	Bias/true %	ESE	ASE	CP
Scenario 1:	100	20	(5, 12)	0.55	0.751	0.743	0.94
Weibull		30	(2, 12)	0.68	0.813	0.824	0.95
($\lambda = 1.5, \gamma = 1.2$)	200	20	(5, 12)	0.38	0.518	0.516	0.95
		30	(2, 12)	0.57	0.564	0.563	0.95
	300	20	(5, 12)	0.88	0.425	0.423	0.95
		30	(2, 12)	−0.18	0.456	0.453	0.94
Scenario 2:	100	20	(8, 16)	1.70	0.818	0.869	0.96
Weibull		30	(5, 16)	2.21	0.871	0.917	0.96
($\lambda = 2.2, \gamma = 2.0$)+	200	20	(8, 16)	0.97	0.477	0.525	0.96
Weibull		30	(5, 16)	1.37	0.568	0.556	0.96
($\lambda = 2.5, \gamma = 0.4$)	300	20	(8, 16)	0.82	0.396	0.421	0.96
joint at $t = 5$		30	(5, 16)	0.59	0.422	0.423	0.95
Scenario 3:	100	20	(6, 10)	1.61	0.658	0.662	0.95
Exponential		30	(4, 10)	2.54	0.803	0.773	0.94
($\lambda = 0.2$)	200	20	(6, 10)	−0.98	0.446	0.453	0.96
$t_0 = 5$		30	(4, 10)	1.81	0.531	0.521	0.95
	300	20	(6, 10)	1.05	0.351	0.367	0.96
		30	(4, 10)	0.93	0.409	0.412	0.96
Scenario 4:	100	20	(8, 16)	0.01	0.671	0.656	0.95
Weibull		30	(5, 16)	0.40	0.756	0.688	0.94
($\lambda = 2.2, \gamma = 2.0$)+	200	20	(8, 16)	−0.75	0.470	0.448	0.94
Weibull		30	(5, 16)	−0.60	0.507	0.470	0.93
($\lambda = 2.5, \gamma = 0.4$)	300	20	(8, 16)	−0.74	0.390	0.364	0.94
joint at $t = 5$		30	(5, 16)	−0.71	0.410	0.379	0.93

Hybrid estimator (2.1) was used in scenarios 1 and 2, and hybrid estimator (2.3) was used in scenarios 3 and 4.
ESE, empirical standard error; ASE, asymptotic standard error; CP, coverage probability.

Table II. Robustness of the hybrid mean survival estimator in the simulation study.

True distribution	t_0	n	Censoring %	Censoring interval (a, b)	Bias/true %	ESE	ASE	CP
Scenario 5: Weibull ($\lambda = 2.5, \gamma = 0.4$)	12	200	20	(12, 18)	2.41	0.410	0.422	0.94
Scenario 6: Weibull ($\lambda = 0.8, \gamma = 1.4$)	5	200	20	(2, 10)	0.05	0.372	0.392	0.96
Scenario 7: Log-logistic ($\lambda = 0.8, \gamma = 0.4$)	4	200	20	(2, 7)	−3.21	0.239	0.230	0.84
Scenario 8: Log-normal ($\lambda = 0.3, \gamma = 1.2$)	4	200	20	(2, 7)	−5.56	0.455	0.437	0.81

Hybrid estimator (2.3) with exponential distribution fit to the tail was used in this table.
ESE, empirical standard error; ASE, asymptotic standard error.

variable, progression-free survival (PFS), defined as the duration from randomization to disease progression or death, whichever occurred first. The primary objective was to compare the PFS time between the two arms.

The difference in median PFS time was 1.1 months with a confidence interval of (0.4, 2.6) from Kaplan–Meier estimate, which indicates a small benefit of adding B to A. However, the hazard ratio estimated from a Cox regression model was 0.68, and the p-value was 0.0002 from the log-rank test. The hazard ratio indicates a reasonably large magnitude of risk reduction (32%), and the p-value suggested that the survival functions were significantly different for the two groups, which were not

reflected by the difference in median survival. Further evaluation of the Kaplan–Meier plot revealed that the curves came together at 50% of survival probability although separated elsewhere.

Above phenomenal was not uncommon in survival analysis. Another example can be found in the prescription information of Vectibix®, a drug for metastatic colorectal cancer [18]. The Kaplan–Meier curves for the primary end point (progression-free survival, PFS) in the registration trial of Vectibix® showed an interesting pattern where the survival curve for the treatment arm was well separated from the control arm, whereas two curves were very close at the median time. The log-rank test indicated a significant improvement in PFS ($p < 0.0001$), whereas the difference in medians was close to 0. In such case, researchers used mean survival time to measure the magnitude of treatment benefit. Calculating mean survival time was possible for that trial as the follow-up period was long enough for the survival curves in both arms to drop to 0 at the end.

For our example, the last observation of the control arm on the Kaplan–Meier curve was censored, and the curve did not drop to 0. Therefore, only restricted mean survival could be estimated. Instead, we estimated the mean survival with the hybrid estimator. A wide choice of distributions including Weibull, exponential, log-logistic, logistic, and log-normal were explored to fit the data overall or by method (2.1). We evaluated the model fitting through Cox–Snell residual plots [16]. However, none of them gave satisfying fitting, which meant that the hybrid mean survival estimator (2.1) cannot be applied, and we had to try the estimator (2.3). It was observed that the hazard rate seems to change after month 6. Thus, an exponential distribution was fit to the curve beyond $t_0 = 6$ months. The adjacent time point τ was chosen to be 15 months because sample survival was only 5% beyond that time point. In other words, the extrapolated tail was used to estimate the survival curve beyond τ to obtain a hybrid survival curve for mean survival calculation. Figure 1 (I) presents the Kaplan–Meier estimate and the fitted distribution function that was used to extrapolate the survival curve beyond τ. Figure 1 (II) shows the whole hybrid curve used to estimate mean lifetime. Fitted distributions and parameters were listed in Table III along with the hybrid mean survival estimated for each arm and difference between the two arms. The estimated difference in hybrid mean survival is 2.2 months. This magnitude better reflects the observed benefit from the estimated hazard ratio and the log-rank test.

5. DISCUSSION

In this article, we estimated mean survival time with a hybrid estimator in two forms of (2.1) and (2.3) on the basis of an extrapolated Kaplan–Meier survival curve beyond the last observation. This hybrid estimator of mean survival was shown to be unbiased and followed a normal distribution. A closed formula of the variance estimate for the hybrid estimator was derived and shown to provide the right coverage probability in the simulation study. Estimating the variance of a hybrid estimator such as ours has always been a challenge in the literatures and is usually obtained using re-sampling method [13]. The closed formula derived here will benefit researchers and software developers.

Traditionally, median survival is used as the standard measure of the center of data distribution, whereas mean survival is rarely acceptable because of censoring and skewed distribution of survival data. In some situations, such as adjuvant oncology trials, the event rate is so low that the median survival time cannot be
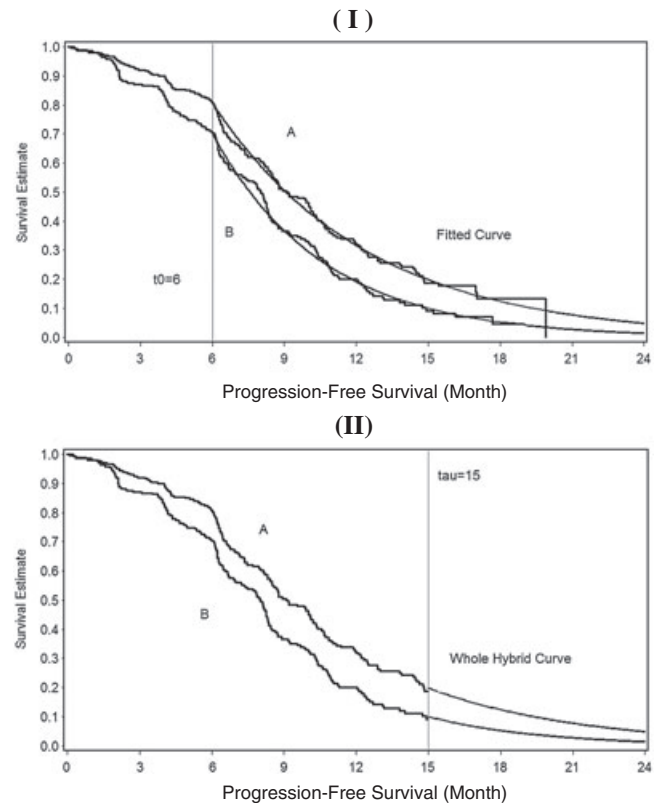
Figure 1. Kaplan–Meier estimate and fitted parametric curve for a phase III clinical study (see Table III for distribution and parameter values).

estimated. Mean survival is used instead, for example, calculating quality-adjusted lifetime and assessing relative cost-effectiveness in health economics [19]. One of the reasons that mean survival has not been widely used in the medical community is that it is difficult to estimate because of censoring and insufficient follow-up. Restricted mean survival was an alternative. However, choosing the restriction time is hard to justify and explain. The hybrid estimator provides an option to estimate mean survival without having to choose a restricted time, which will make the concept of mean survival easier to understand for patients. Using difference in mean survival to quantify the treatment benefit presents certain advantages in some situations as shown in the example.

The statistical inferences derived in the article can be easily extended to the calculation of quality-adjusted survival, which is essentially a weighted mean survival time. It is a major challenge in patient-reported outcome research that the quality-adjusted survival calculated on the basis of restricted mean survival cannot be extended to the point beyond restriction time T. Gelber *et al.* [13] investigated the quality-adjusted survival based on the extrapolated survival curves and obtained the confidence interval from bootstrap. Their estimator is essentially the same as our estimator (2.3).

The main limitation of the hybrid estimator is that the performance of the hybrid estimator depends on the goodness of fit of the parametric model. Certain model diagnosis has to be done [16]. The hybrid estimator (2.3) has certain advantages and flexibility that could achieve better model fitting. It is more important to obtain a good fitting at the tail part of the survival curve. If

Table III. Hybrid mean survival estimation for progression-free survival in the example.

Treatment	Exponential parameter λ beyond $t_0 = 6$	Estimated KM value at $t_0 = 6$	Survival function of the extrapolated tail	Median (month)	Difference in median (95% CI, month)	Hybrid mean survival (SE, month)	Difference (95% CI, p-value month)
A+PL	0.216	0.702	$0.702 * \exp(-0.216\,t)$	8.0		8.5 (0.37)	
					1.1 (0.4, 2.6)		2.2 (0.9, 3.5)
A+B	0.156	0.808	$0.808 * \exp(-0.156\,t)$	9.1		10.7 (0.56)	0.0005

The difference in medians is obtained via bootstrap method.
PL: placebo.

there is no appropriate parametric model that provides good fitting of the survival curve, especially at the tail part, one should avoid using the hybrid estimator and pursue restricted mean survival or median survival time instead. It is important to point out that all extrapolations beyond observed range of data are associated with untestable assumptions. Should the true distribution beyond last observed time change significantly, the hybrid mean survival estimator would provide biased estimate. In addition, the formulas in this article cannot be used without individual patient-level data. In the absence of individual patient-level data, one must turn to other methods, such as method of moments or regression analysis.

It is important to point out that it is not acceptable to use a hybrid estimator as an alternative to designing for sufficient follow-up period. In situations where it is difficult to have a sufficient follow-up period, such as adjuvant oncology trials, the hybrid estimator of mean survival could be considered as an alternative. Future research in this area includes exploring the possibility of deriving a robust variance estimator of the hybrid mean survival.

Acknowledgements

REFERENCES

[1] Irwin JO. The standard error of an estimate of expectational life. *Journal of Hygiene* 1949; **47**:188–189.
[2] Karrison T. Use of Irwin's restricted mean as an index for comparing survival in different treatment groups—interpretation and power considerations. *Controlled Clinical Trials* 1997; **18**:151–167.
[3] Pepe MS, Fleming TR. Weighted Kaplan–Meier statistics: a class of distance tests for censored survival data. *Biometrics* 1989; **45**:497–507.
[4] Pepe MS, Fleming TR. Weighted Kaplan–Meier statistics: large sample and optimality considerations. *Journal of the Royal Statistical Society, Series B* 1991; **53**:341–352.
[5] Chen P, Tsiatis AA. Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* 2001; **57**:1030–1038.
[6] Zucker DM. Restricted mean life with covariates: modification and extension of a useful survival analysis method. *Journal of the American Statistical Association* 1998; **93**:702–709.
[7] Shen Y, Cai J. Maximum of the weighted Kaplan–Meier tests with application to cancer prevention and screening trials. *Biometrics* 2001; **57**:837–843.
[8] Gelber RD, Cole BF, Gelber S, Goldhirsch A. Comparing treatment using quality adjusted survival: the Q-TWiST method. *American Statistician* 1995; **49**:161–169.
[9] Goldhirsch A, Gelber RD, Simes RJ, Glasziou P, Coates AS. Costs and benefits of adjuvant therapy in breast cancer: a quality-adjusted survival analysis. *Journal of Clinical Oncology* 1989; **7**(1):36–44.
[10] Efron B. The two sample problem with censored data. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* 1967; **4**:831–852.
[11] Moeschberger ML, Klein JP. A comparison of several methods of estimating the survival function when there is extreme right censoring. *Biometrics* 1985; **41**:253–259.
[12] Klein JP, Lee S, Moeschberger ML. A partially parametric estimator of survival in the presence of randomly censored data. *Biometrics* 1990; **46**:795–811.
[13] Gelber RD, Goldhirsch A, Cole BF. Parametric extrapolation of survival estimates with applications to quality of life evaluation of treatments. *Controlled Clinical Trials* 1993; **14**:485–499.
[14] Gill RD. Censoring and stochastic integrals. *Mathematical Centre Tracts*, 124. Mathematisch Centrum: Amsterdam, 1980.
[15] Gill R. Large sample behaviour of the product-limit estimator on the whole line. *The Annals of Statistics* 1983; **11**(1):49–58.
[16] Klein JP, Moeschberger ML. *Survival analysis: techniques for censored and truncated data*, 2nd ed. Springer: New York, 2003.
[17] Su Z, Fang L. A novel method to calculate mean survival time for time-to-event data. *Communications in Statistics-Simulation and Computation* 2011, accepted for publication.
[18] Cutsem EV, Peeters M, Siena S, Humblet Y, Hendlisz A, Neyns B, Canon J, Laethem JV, Maurel J, Richardson G, Wolf M, Amado RG. Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. *Journal of Clinical Oncology* 2007; **25**:1658–1664.
[19] Neymark N, Adriaenssen I, Gorlia T, Caleo S, Bolla M. Estimating survival gain for economic evaluations with survival time as principle endpoint: a cost-effectiveness analysis of adding early hormonal therapy to radiotherapy in patients with locally advanced prostate cancer. *Health Economics* 2002; **11**(3):233–48.
[20] Shen Y, Fleming TR. Weighted mean survival test statistics: a class of distance tests for censored survival data. *Journal of Royal Statistical Society, Series B* 1997; **59**:269–280.