Therapeutic Innovation & Regulatory Science (2024) 58:721–729
https://doi.org/10.1007/s43441-024-00640-7

DIA

ORIGINAL RESEARCH

Improvement of Midpoint Imputation for Estimation of Median Survival Time for Interval-Censored Time-to-Event Data

Yuki Nakagawa[1,2] · **Takashi Sozu**[3]

Received: 12 May 2023 / Accepted: 8 March 2024 / Published online: 10 April 2024

Abstract

Background Progression-free survival (PFS) is used to evaluate treatment effects in cancer clinical trials. Disease progression (DP) in patients is typically determined by radiological testing at several scheduled tumor-assessment time points. This produces a discrepancy between the true progression time and the observed progression time. When the observed progression time is considered as the true progression time, a positively biased PFS is obtained for some patients, and the estimated survival function derived by the Kaplan–Meier method is also biased.

Methods While the midpoint imputation method is available and replaces interval-censored data with midpoint data, it unrealistically assumes that several DPs occur at the same time point when several DPs are observed within the same tumor-assessment interval. We enhanced the midpoint imputation method by replacing interval-censored data with equally spaced timepoint data based on the number of observed interval-censored data within the same tumor-assessment interval.

Results The root mean square error of the median of the enhanced method is almost always smaller than that of the midpoint imputation regardless of the tumor-assessment frequency. The coverage probability of the enhanced method is close to the nominal confidence level of 95% in most scenarios.

Conclusion We believe that the enhanced method, which builds upon the midpoint imputation method, is more effective than the midpoint imputation method itself.

Keywords Cancer clinical trial · Interval censoring · Median survival time · Progression-free survival · Survival analysis

Introduction

In late-stage cancer clinical trials, overall survival (OS), which is the time from the date of randomization in a clinical trial to the date of death due to any cause, and progression-free survival (PFS), which is the time from randomization to the time of the objective tumor progression or death due to any cause, are the generally obtained and tracked parameters [1]. The OS value is used as the primary endpoint to measure efficacy because it reflects the treatment benefits, and PFS is used as the secondary endpoint.

However, the use of OS as the primary endpoint involves increased sample size and trial duration due to fewer expected events. Additionally, the impact of subsequent therapy after the assigned treatment on survival time must be considered when evaluating the therapeutic effects of OS. Therefore, PFS, which eliminates the impact of subsequent therapy in cancer clinical trials or serves as the surrogate endpoint for OS in the areas where the surrogacy of PFS for OS is established, is primarily used as the endpoint for the evaluation of pure efficacy. Particularly, PFS is used as the primary endpoint in clinical trials to develop a therapy for conditions such as lung, breast, colon, or liver cancers [2–10].

Disease progression (DP) in patients is typically determined by radiological testing at several scheduled intervals under the criteria defined by guidelines such as RECIST v1.1 [11]. The true date of death can be obtained (i.e. death is not interval censored); however, the true date of

✉ Yuki Nakagawa
nakagawa.yuki99@chugai-pharm.co.jp

[1] Biometrics Department, Chugai Pharmaceutical Co., Ltd., 2-1-1 Nihonbashi-Muromachi, Chuo-Ku, Tokyo 103-8324, Japan

[2] Department of Management Science, Graduate School of Engineering, Tokyo University of Science, Tokyo, Japan

[3] Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science, Tokyo, Japan

DP cannot. The interval-censored data obtained after the advent of DP only indicates that it occurred between adjacent tumor assessments. The PFS is overestimated when deterministically estimating the survival function from the Kaplan–Meier method that considers the observed date of DP as the true date of DP. This procedure is known as the upper-limit imputation (UI) method. Practical approaches for interval-censored data are described by Bogaerts et al. [12].

This problem can be solved by increasing the tumor-assessment frequency and reducing the length of the tumor-assessment intervals. However, because the radiological testing for tumor assessment involves radiation exposure, it is highly invasive, and repeated tumor assessments may impair patients' health. Therefore, a method is required that appropriately estimates the survival function with a limited tumor-assessment frequency.

Law and Brookmeyer [13] evaluated the statistical performances of the Kaplan–Meier estimator based on the midpoint imputation (MI) method. Compared with the UI method, this method can significantly reduce bias. However, several DP dates are replaced with the same date when several DPs are observed in the same tumor-assessment interval. It is unrealistic to assume that several DPs occurred at the same time point. Furthermore, Nishikawa and Tango [14] reported that such a deterministic approach produces an underestimated standard error, and its coverage probability of the confidence interval (CI) of the median based on the Brookmeyer and Crowley method [15] cannot be ensured.

Turnbull [16] proposed a method for providing an estimator, referred to below as the Turnbull estimator, which is a nonparametric maximum likelihood estimator of the cumulative distribution function, using a self-consistency algorithm for the interval-censored data. Gentlemen and Greyer [17] provided a sufficient condition for the almost sure convergence of the maximum likelihood estimator to the true underlying distribution function. Wellner and Zhan [18] presented a hybrid algorithm that converged the Turnbull estimator rapidly using an expectation–maximization iterative convex minorant algorithm. The Turnbull estimator can be calculated using standard software. For example, Anderson-Bergman [19] proposed a computationally efficient algorithm for the nonparametric maximum likelihood estimator, which is implemented in the icenReg R package [20]. However, the Turnbull estimator is rarely used in practical clinical research [2–10]. The reason may be that non-statisticians, as well as some statisticians, have difficulty understanding the methodology.

This paper improves the MI method by retaining its statistical simplicity, and the statistical performances of the enhanced midpoint imputation (EMI) method are evaluated in comparison to the conventional methods. We focus on the estimation of median survival time, which is generally used as the representative value for survival time in practical clinical research [2–10].

The remainder of this paper is organized as follows. Sect. "Statistical Settings" describes the statistical settings and the EMI method. Sect. "Simulation Study" compares of the EMI method with conventional methods through simulation, and Sect. "Numerical Examples Based on Actual Clinical Trials" shows survival estimates using the EMI method. Sect. "Discussion" presents additional remarks and the conclusion.

Statistical Settings

Notation and Framework

Consider a randomized cancer clinical trial that uses PFS as one of the endpoints. For simplicity, only a single arm is considered, and it is assumed that the PFS events were exclusively DP with no deaths in this section unless specifically noted. Patients with solid tumors visit a hospital to evaluate the tumor status with a scheduled tumor assessment. For patient i, the time from the initial treatment to the actual DP is denoted as X_i. When the DP is observed as a PFS event, the interval-censored data, $(L_i, R_i]$ can be obtained, signifying that the exact DP date lies between two consecutive tumor assessments. R_i and L_i $(L_i \leq R_i)$ are the observed interval limits of DP.

Upper-Limit Imputation

Researchers generally replace X_i with R_i and estimate the distribution of the PFS using the Kaplan–Meier method because the exact DP date cannot be determined. The UI can be described as follows:

$$Y_i = R_i \tag{1}$$

where Y_i is defined as the corrected PFS for the ith patient.

Midpoint Imputation

The estimated distribution of the PFS based on the UI is frequently overestimated. The MI replaces the interval-censored data with the midpoint between L_i and R_i to reduce the overestimation [13]. The MI can be described as follows:

$$Y_i = (L_i + R_i)/2 \tag{2}$$

Enhanced Midpoint Imputation

The two conventional methods described above individually correct the interval-censored data of several patients. These

methods require the unrealistic assumption that several DP dates are replaced with the same date (i.e., tie data) when several DPs are observed in the same tumor-assessment interval. Though multiple imputation methods exist, they do not maintain the simplicity of the conventional methods, and the influence of random numbers is unsuitable for application in confirmatory clinical trials. The following intuitive approach retains the simplicity of the MI and considers the issue of tied data. Correcting the interval-censored data based on the number of patients, n_j, who exhibit DP in the same tumor-assessment interval j, can be considered an improvement of the MI method.

$$Y_{s_j} = L_j + \left(R_j - L_j\right) \times \frac{s_j}{n_j + 1} \tag{3}$$

R_j and L_j ($L_j \leq R_j$) are the observed interval limits of DP in the tumor-assessment interval, j. Here, $s_j = 1, 2, \ldots, n_j$. Patients who exhibit DP in the same tumor-assessment interval, j, are assigned an s_j with a different positive integer. The EMI involves data handling techniques to calculate n_j and s_j. When a patient has a unique combination of L_j and R_j, the EMI derives the same Y_{s_j} as Y_i of the MI. Therefore, the EMI can be considered a method that expands upon the MI.

The specifically corrected interval-censored data are considered using the EMI. Table 1 showcases numerical examples distinguishing the EMI method from conventional methods. Assume seven patients experience DP as in the PFS event. Patients 1–4 have the same combination of L_j and R_j. Thus, n_j is derived as 4, and the values of Y_{s_j} are 4.8, 5.6, 6.4, and 7.2. The EMI does not allocate a specific value to each patient, such as $Y_1 = 4.8$ and $Y_2 = 5.6$. In the actual data analysis, any Y_{s_j} can be assigned to any patient as needed. Patient 5 has a unique combination of L_j and R_j. Thus, n_j is derived as 1, and Y_5 is derived as 7, similar to the MI. Thus, several interval-censored data points will be corrected for each combination of L_j and R_j. Similar to other methods, the survival distribution is typically estimated using the Kaplan–Meier method with both the corrected interval-censored data and uncorrected data, such as death. The median PFS values are estimated as 8, 6, and 7 by the UI, MI, and EMI, respectively.

The actual data analysis procedures were as follows: 1) Sort patient-level data by L_j and R_j and calculate n_j, the number of patients having an identical combination of L_j and R_j. 2) Calculate Y_{s_j} where $s_j = 1, 2, \ldots, n_j$. 3) Assume Y_{s_j} as time-to-event data. 4) Perform time-to-event analysis using well-known software, such as the LIFETEST procedure of SAS.

In a clinical trial employing a parallel group design, the EMI must be used separately in each arm. When performing a covariate-adjusted analysis, for instance, a subgroup analysis by sex, the EMI must be used separately for each subgroup because the EMI introduces a bias in the estimate if the order of data records (as given by the indices s_j) is related to the group. For example, if the data in group A are recorded before those in group B (i.e., the s_j are allocated based on the order of data records), the corrected survival time given by the EMI may be shorter than the true value that would have been observed, had it not been censored in group A. Conversely, the corrected survival time may be longer than the true value in group B.

Simulation Study

This study considered a randomized cancer clinical trial with patients equally allocated to two interventions, using PFS as one of the endpoints to compare the experimental treatment with the control treatment for solid tumors. Simulation studies were performed, focusing on the median, often used as a representative value of survival time, to evaluate the statistical performances of the EMI in comparison with conventional methods based on practical clinical trial settings. To emphasize the difference between the statistical performances of the EMI and the MI, it was assumed that all patient visits consistently followed the scheduled tumor assessments and that death was not observed before the documented DP in this simulation study. This setting resembles the well-controlled clinical trials of first-line therapies against metastatic cancer.

Table 1 Numerical examples

i	L_i	R_i	n_j	s_j	UI	MI	EMI
1	4	8	4	1	8	6	4.8
2	4	8	4	2	8	6	5.6
3	4	8	4	3	8	6	6.4
4	4	8	4	4	8	6	7.2
5	4	10	1	1	10	7	7
6	8	12	2	1	12	10	9.3
7	8	12	2	2	12	10	10.7

Simulation Settings

Four patterns of tumor-assessment frequency, including every 6, 9, 12, and 20 weeks (corresponding to every 1.4, 2.1, 2.8, and 4.6 months), were analyzed. Notably, "every 20 weeks" is a reference setting for an unrealistic scenario. The simulation data were generated as follows:

Step 1: For the ith patient, the time from randomization to the potential DP, X_i, was generated from a specified Weibull distribution. The true PFS is defined as $S(t) = \exp(-\lambda t^{\gamma})$. γ was set either to 0.5, 1.0, or 2.0. The patients were randomly assigned to the control and experimental arms in a 1:1 ratio. In the control arm, λ was derived based on γ and the true median PFS in the control arm of 3, 6, 9, or 12 months. In the experimental arm, λ was derived based on γ and the true median PFS in the control arm and a hazard ratio (HR) of 0.6 or 0.75. The dropout date, Z_i, was exponentially distributed with a dropout rate of 5% per 12 months to account for loss to follow-up before the DP.

Step 2: The time from randomization to the tumor assessment,$V_j (j = 0, \ldots, K_i)$, was generated based on a tumor-assessment frequency of every 6, 9, 12, or 20 weeks, where V_0 represents the randomization date. If $V_{k_i-1} < X_i \leq V_{k_i}$, the L_i and R_i are equal to V_{k_i-1} and V_{k_i}, respectively.

Step 3: For patient $i = 1$ to N, where N of 100, 200, or 400 is the total number of patients, steps 1 and 2 were repeated.

Step 4: Based on the event-sample size proportion, defined as the proportion of the number of events in the analysis to the total sample size of 0.8, the analysis timing was determined as the time at which the planned number of events was reached. The patients with $R_i \leq Z_i$ were treated as those with the event.

Step 5: Patients, without a documented event at the analysis timing were censored at the last tumor-assessment date, $Z\prime_i$, before the analysis timing or the loss to follow-up. Patients without a post-baseline tumor assessment before the analysis timing were censored at the randomization date. The survival time, $T_i = \min\left(Y_i, Z\prime_i\right)$, and the censoring indicator, $\delta_i = I\left(R_i > Z\prime_i\right)$, where $I(\bullet)$ is an indicator function, were set for the UI, MI, and EMI. Y_i is derived from (1), (2), and (3). For the Turnbull estimator, T_i equals $\left(L_i, R_i\right]$ or $\left(Z\prime_i, \infty\right]$ if $R_i \leq Z\prime_i$ or $R_i > Z\prime_i$, respectively.

Patients were assumed to have been uniformly recruited over 12 months. The values of the simulation parameters were determined based on practical clinical trials. For example, a scenario with 400 patients, a HR of 0.75, a median PFS of 3 months in the control arm, and a tumor-assessment frequency of every 6 weeks, resembles the IMpower133 trial [2]. Similarly, a scenario with 100 patients, a HR of 0.60, a median PFS of 3 months in the control arm, and a tumor-assessment frequency of every 9 or 12 weeks is similar to Group F in the GO30140 trial [9].

The number of simulations, S, was set to 5000 for each scenario, defined by all combinations of the simulation parameters. Using SAS 9.4 for the Monte Carlo simulations, the median PFS estimate, $\widehat{M}_s$, for the sth simulation, was calculated using either the Kaplan–Meier or Turnbull estimators. Given that the Turnbull estimator is indifferent to where the probability mass is placed inside the Turnbull interval, this renders the Turnbull estimator non-unique. A detailed description of the Turnbull estimator was provided by Bogaerts et al. [12] Although the appropriate handling of non-unique estimates of the Turnbull estimator remains unclear, we addressed non-unique estimates of the median as the upper limit of the Turnbull interval that includes the median. This approach is implemented in the ICLIFETEST procedure in SAS. The bias and root mean square error (RMSE) of the median were defined in (4) and (5), respectively.

$$bias\ in\ median = \sum_{s=1}^{S} \frac{\widehat{M}_s - M}{S} \tag{4}$$

$$RMSE\ of\ median = \sqrt{\sum_{s=1}^{S} \frac{\left(\widehat{M}_s - M\right)^2}{S}} \tag{5}$$

Here, M is defined as the true median PFS. In the sth simulation, the 95% CI for the median PFS based on the Brookmeyer and Crowley method was calculated with a double logarithmic transformation. The standard errors of the Kaplan–Meier and Turnbull estimators were calculated using Greenwood's formula and simple bootstrap imputation, respectively. The estimable proportion, the proportion of obtained estimations of both the upper and lower limits of the 95% CI to the total number of simulations, and the coverage probability were used. Defined as the proportion of instances where the 95% CI includes the true median to the total number of simulations, these are presented for each scenario, differentiated by a unique set of simulation parameters. If either the lower or upper limit of the 95% CI was not obtained, the true median is not included in the 95% CI. Due to space limitations, only a subset of the results from the scenarios, defined by a unique set of simulation parameters, is presented in this paper. The SAS and all the other SAS Institute, Inc. product or service names are registered trademarks or trademarks of the SAS Institute, Inc. (Cary, NC, USA).

Table 2 Bias and RMSE of median with 200 patients, γ of 1.0, and HR of 0.6

True Median* [months]	Tumor assessment frequency** [months]	Bias (RMSE) [months]							
		Control arm				Experimental arm			
		UI	Turnbull	MI	EMI	UI	Turnbull	MI	EMI
3 versus 5	1.4	0.8 (1.0)	0.8 (1.0)	0.1 (0.6)	0.1 (0.4)	0.7 (1.1)	0.7 (1.1)	0.0 (0.8)	0.0 (0.7)
	2.1	1.1 (1.2)	1.1 (1.2)	0.1 (0.3)	0.1 (0.4)	1.1 (1.4)	1.1 (1.4)	0.1 (0.8)	0.1 (0.7)
	2.8	1.7 (2.1)	1.7 (2.1)	0.3 (1.3)	0.1 (0.5)	1.2 (1.7)	1.2 (1.7)	−0.2 (1.2)	0.1 (0.7)
	4.6	1.6 (1.6)	1.6 (1.6)	−0.7 (0.7)	0.5 (0.6)	2.8 (3.5)	2.8 (3.5)	0.5 (2.1)	0.2 (0.8)
6 versus 10	1.4	0.7 (1.2)	0.7 (1.2)	0.0 (0.9)	0.0 (0.8)	0.7 (1.7)	0.7 (1.7)	0.1 (1.5)	0.1 (1.5)
	2.1	1.0 (1.5)	1.0 (1.5)	0.0 (1.1)	0.1 (0.8)	1.1 (1.9)	1.1 (1.9)	0.1 (1.6)	0.1 (1.5)
	2.8	1.5 (2.0)	1.5 (2.0)	0.1 (1.3)	0.1 (0.8)	1.4 (2.2)	1.4 (2.2)	0.0 (1.7)	0.1 (1.4)
	4.6	3.0 (3.1)	3.0 (3.1)	0.7 (1.1)	0.2 (0.9)	2.4 (3.3)	2.4 (3.3)	0.1 (2.2)	0.1 (1.4)
9 versus 15	1.4	0.7 (1.5)	0.7 (1.5)	0.0 (1.3)	0.0 (1.3)	0.8 (2.4)	0.8 (2.4)	0.1 (2.2)	0.1 (2.2)
	2.1	1.0 (1.8)	1.0 (1.8)	0.0 (1.4)	0.0 (1.3)	1.1 (2.5)	1.1 (2.5)	0.0 (2.3)	0.1 (2.2)
	2.8	1.4 (2.1)	1.4 (2.1)	0.0 (1.6)	0.1 (1.3)	1.4 (2.8)	1.4 (2.8)	0.1 (2.4)	0.1 (2.2)
	4.6	2.1 (3.1)	2.2 (3.1)	-0.2 (2.3)	0.1 (1.3)	2.3 (3.5)	2.3 (3.5)	0.0 (2.6)	0.1 (2.2)
12 versus 20	1.4	0.8 (2.0)	0.8 (2.0)	0.1 (1.8)	0.1 (1.7)	0.8 (3.1)	0.8 (3.1)	0.2 (3.0)	0.2 (3.0)
	2.1	1.0 (2.2)	1.0 (2.2)	0.0 (1.9)	0.0 (1.8)	1.1 (3.2)	1.1 (3.2)	0.1 (3.0)	0.1 (2.9)
	2.8	1.4 (2.4)	1.4 (2.4)	0.0 (1.9)	0.0 (1.7)	1.5 (3.4)	1.5 (3.4)	0.1 (3.0)	0.1 (2.9)
	4.6	2.3 (3.0)	2.3 (3.0)	0.0 (2.0)	0.1 (1.7)	2.2 (3.9)	2.2 (3.9)	−0.1 (3.3)	−0.1 (2.9)

*Control arm versus Experimental arm

**The first, second, third, and fourth lines correspond to every 6, 9, 12, and 20 weeks

Simulation Results

Table 2 lists the bias and RMSE of the median with 200 patients, a γ of 1.0, and an HR of 0.6. The UI overestimated the median in all scenarios. These overestimations increased as the tumor-assessment frequency decreased. The Turnbull estimator produced results similar to that of the UI. Both biases of the median of the MI and EMI were close to zero in all scenarios except for the unrealistic scenarios with a tumor-assessment frequency of 20 weeks. In the realistic extreme scenarios, where the tumor-assessment frequency was every 12 weeks regardless of the median PFS, the bias of the median of the EMI was approximately 0.2 months. Even in the most extreme scenarios, where the median PFS was three months, and the tumor-assessment frequency was every 20 weeks, the bias of the median of the EMI was approximately 0.5 months. The UI and Turnbull estimator provided a larger RMSE of the median compared to the MI. The EMI consistently provided a smaller RMSE of the median compared to the MI regardless of the tumor-assessment frequency. In scenarios with the same true median, the RMSE of the median of the MI was affected by the tumor-assessment frequency, while that of the EMI remained consistent.

In the scenario with a γ of 1.0 and HR of 0.6, the EMI provided a bias of the median similar to the MI and a smaller RMSE of the median compared to the MI, irrespective of the number of patients (Tables A1 and A2). In the scenario with 200 patients and an HR of 0.6, these trends were observed irrespective of the value of γ (Tables A3 and A4). The bias and RMSE of the median of the EMI decreased with an increase in the number of patients or γ. As the number of events around the tumor-assessment interval that included the true median increased with an increase in the number of patients or γ, it can be inferred that the bias and RMSE of the median of the EMI decreased accordingly. In the scenario with an extremely small number of events around the tumor-assessment interval that included the true median, such as the one with 200 patients, a γ of 0.5, HR of 0.6, and a true median of 33.3 months, the EMI produced results similar to the MI. In the scenario with the longest true median of 33.3 months, both the EMI and the MI had a bias in the median of around 1 month. However, considering the absolute value of the true median, this bias is likely negligible.

Table 3 shows the coverage probability and the estimable proportion of the 95% CI with 200 patients, a γ of 1.0, and an HR of 0.6. The UI, Turnbull estimator, and MI often yielded low estimable proportions, and their coverage probability did not meet the nominal confidence level of 0.95. In contrast, the EMI consistently provided a high estimable proportion and a coverage probability that met the nominal confidence level in nearly all scenarios, excluding the unrealistic scenario with a true median of three months and a tumor-assessment frequency of

Table 3 Coverage probability and estimable proportion of 95%CI of median with 200 patients, γ of 1.0, and HR of 0.6

True Median* [months]	Tumor assessment frequency** [months]	Coverage probability (Estimable proportion)							
		Control arm				Experimental arm			
		UI	Turnbull	MI	EMI	UI	Turnbull	MI	EMI
3 versus 5	1.4	0.92 (0.99)	0.92 (0.99)	0.92 (0.99)	0.96 (1.00)	0.79 (1.00)	0.79 (1.00)	0.88 (1.00)	0.96 (1.00)
	2.1	0.39 (0.67)	0.38 (0.67)	0.39 (0.67)	0.97 (1.00)	0.79 (0.99)	0.78 (0.99)	0.79 (0.99)	0.96 (1.00)
	2.8	0.93 (0.93)	0.93 (0.93)	0.93 (0.93)	0.96 (1.00)	0.06 (0.92)	0.05 (0.91)	0.88 (0.92)	0.96 (1.00)
	4.6	0.00 (0.10)	0.00 (0.10)	0.10 (0.10)	0.86 (1.00)	0.93 (0.93)	0.93 (0.93)	0.93 (0.93)	0.96 (1.00)
6 versus 10	1.4	0.93 (1.00)	0.93 (1.00)	0.93 (1.00)	0.96 (1.00)	0.94 (1.00)	0.94 (1.00)	0.94 (1.00)	0.95 (1.00)
	2.1	0.38 (1.00)	0.37 (1.00)	0.94 (1.00)	0.95 (1.00)	0.80 (1.00)	0.79 (1.00)	0.94 (1.00)	0.95 (1.00)
	2.8	0.92 (0.99)	0.92 (0.98)	0.92 (0.99)	0.96 (1.00)	0.80 (1.00)	0.79 (1.00)	0.87 (1.00)	0.96 (1.00)
	4.6	0.64 (0.72)	0.63 (0.71)	0.64 (0.72)	0.97 (1.00)	0.93 (0.99)	0.93 (0.99)	0.93 (0.99)	0.96 (1.00)
9 versus 15	1.4	0.93 (1.00)	0.93 (1.00)	0.91 (1.00)	0.96 (1.00)	0.93 (1.00)	0.92 (1.00)	0.95 (1.00)	0.95 (1.00)
	2.1	0.93 (1.00)	0.93 (1.00)	0.93 (1.00)	0.96 (1.00)	0.95 (1.00)	0.95 (1.00)	0.95 (1.00)	0.95 (1.00)
	2.8	0.93 (1.00)	0.93 (1.00)	0.93 (1.00)	0.96 (1.00)	0.93 (1.00)	0.92 (1.00)	0.93 (1.00)	0.95 (1.00)
	4.6	0.02 (0.96)	0.02 (0.95)	0.94 (0.96)	0.95 (1.00)	0.93 (1.00)	0.92 (1.00)	0.93 (1.00)	0.96 (1.00)
12 versus 20	1.4	0.93 (1.00)	0.92 (1.00)	0.94 (1.00)	0.95 (1.00)	0.95 (1.00)	0.94 (1.00)	0.95 (1.00)	0.95 (1.00)
	2.1	0.87 (1.00)	0.86 (1.00)	0.94 (1.00)	0.95 (1.00)	0.94 (1.00)	0.93 (1.00)	0.94 (1.00)	0.96 (1.00)
	2.8	0.93 (1.00)	0.93 (1.00)	0.93 (1.00)	0.96 (1.00)	0.95 (1.00)	0.95 (1.00)	0.95 (1.00)	0.96 (1.00)
	4.6	0.65 (0.99)	0.64 (0.99)	0.80 (0.99)	0.96 (1.00)	0.94 (1.00)	0.93 (1.00)	0.94 (1.00)	0.96 (1.00)

*Control arm versus Experimental arm

**The first, second, third, and fourth lines correspond to every 6, 9, 12, and 20 weeks

20 weeks. Given the nature of the Brookmeyer and Crowley method wherein the upper and lower limits of the 95% CI were selected from the event dates, the EMI flexibly provided the 95% CI by using more candidate time points than the other methods.

These patterns were consistent regardless of the number of patients and γ (Tables A5, A6, A7 and A8). The EMI rarely gave a conservative coverage probability, for example, 0.99 in a scenario with 200 patients, a γ of 2.0, HR of 0.6, a true median of three months, and a tumor-assessment frequency of nine weeks.

These findings suggest that the EMI offers similar accuracy and greater precision than the MI. The HR does not remarkable influence the aforementioned statistical performances, as observed in scenarios with an HR of 0.75. The UI, MI, and EMI all demonstrated high performance concerning the bias and RMSE of the HR (Table A9).

Additional supplementary analyses were provided in Section A10, detailing differences in assumed and estimated survival functions (Tables A.10–1, A.10–2, and A.10–3), the type I error and power of the log-rank test (Tables A.10–4 and A.10–5), the bias, RMSE, and coverage probability when patients were noncompliant with tumor-assessment schedules (Tables A.10–6 and A.10–7), and consistency and unbiasedness (Tables A.10–8 and A.10–9).

Numerical Examples Based on Actual Clinical Trials

The EMI, MI, and UI were applied to the numerical example data, which was generated based on the PFS data of the actual randomized phase III trial from Project Data Sphere [21]. The original data was derived from the randomized phase III trial [6], which compared the effect of panitumumab with best supportive care (BSC) to that of BSC alone in patients with metastatic colorectal cancer. For simplicity, the BSC arm data was removed. In the example data, assuming a tumor-assessment frequency of six weeks, the date of DP in the original data was deemed to be observed between the assumed tumor-assessment interval to emphasize the difference between the statistical performances of the three methods. The censoring and death event dates were retained from the original data. Figure 1 shows the survival estimates based on the exemplary data. The medians using the EMI, MI, UI, and Turnbull were 2.4, 2.1, 2.8, and 1.5 months, respectively. The EMI provided a smooth survival function compared with the other Kaplan–Meier estimators.

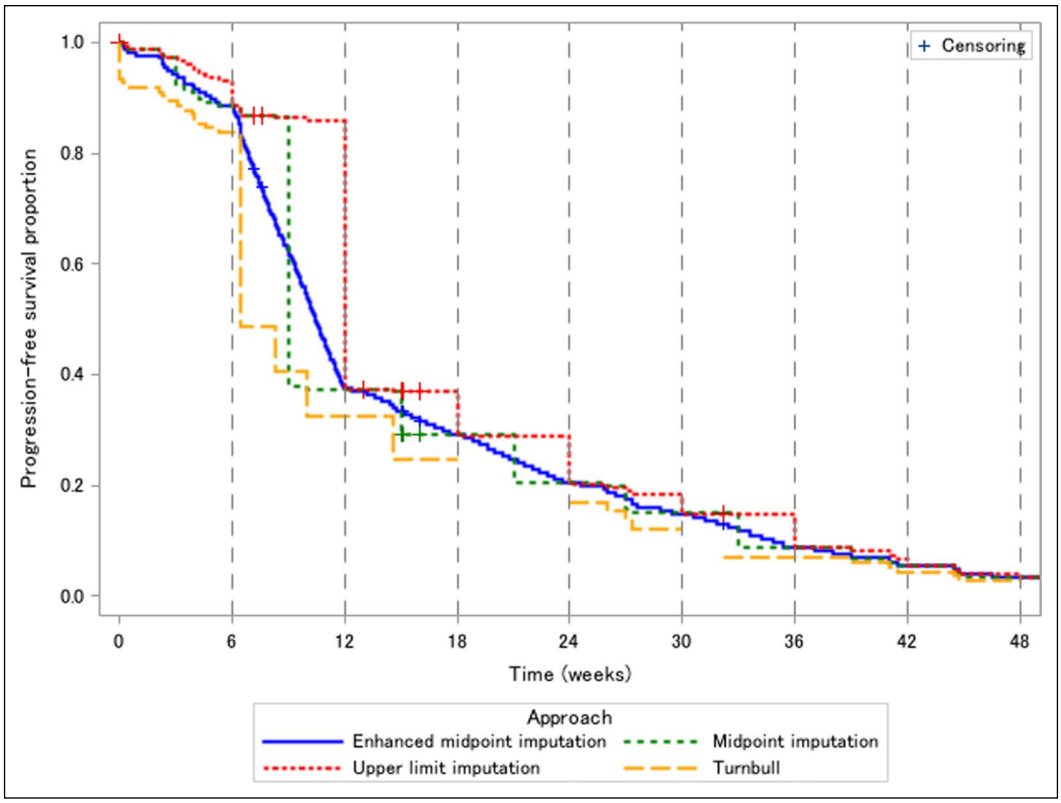

Figure 1 Survival estimates based on example data.

Discussion

This study followed an effective approach, which is an improvement of the MI method without its statistical complexity, to estimate PFS. The statistical performances of the EMI were evaluated through simulation studies under various scenarios based on realistic assumptions.

The EMI provided precision higher than that of the MI. The EMI rarely provided a conservative coverage probability. However, considering that it provided the CI with a nominal confidence level in almost all the scenarios and a conservative CI would be more acceptable in comparison with a liberal CI in medical research, the EMI's CI can be used. Moreover, sensitivity analysis, such as the CI based on the bootstrap method, can be useful.

The Turnbull estimator's results aligned closely with those of the UI. This correspondence can be attributed to two primary factors: (1) instances where the Turnbull estimates for the median were non-unique were substituted with the upper bounds of the Turnbull interval containing the median; (2) the consistent adherence to tumor assessments at pre-scheduled time points, where deaths occurring prior to the DP were not taken into account. Consequently, the Turnbull intervals largely mirrored the preset tumor-assessment intervals.

The simulations encompass a broad range of scenarios representative of clinical trials for solid metastatic cancer. Thus, the findings of this study are potentially applicable in practical clinical trials.

It is worth highlighting that the simulation studies operated under the premise that any death occurring prior to a documented DP is not recorded. In clinical trials focusing on first-line therapies against metastatic cancers, death is a relatively uncommon occurrence. However, it becomes more consequential in clinical trials of later-line therapies. The EMI remains viable for real-world trials, even in cases with patient noncompliance related to tumor-assessment schedules or when deaths precede a documented DP. In these circumstances, the EMI's performance surpasses that of the MI, if marginal DPs are identified within the same tumor-assessment interval.

Conclusion

Based on the preceding discussion, we believe that the EMI is more useful than the MI, particularly in well-controlled clinical trials of first-line therapies against metastatic cancer, in which the incidence of noncompliance with the tumor-assessment schedule and documented death before DP are

negligible. In future work, the proposed method will be compared with others, which include smoothing methods.

Acknowledgements
The authors thank the editor and referees for their very helpful suggestions that substantially improved the article. The dataset (PDS UID: Colorec_Amgen_2004_310) used as the numerical example of this article is available in the Project Data Sphere repository: https://data.projectdatasphere.org/projectdatasphere/html/home.

Author Contributions
All authors contributed to the study's conception and design. The analysis was performed by YN. The first draft of the manuscript was written by YN, and TS commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding
This research received no specific Grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Conflict of interest
Yuki Nakagawa is an employee of Chugai Pharmaceutical Co., Ltd.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1007/s43441-024-00640-7.

References

1. Food and Drug Administration. Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics. Accessed 24 Apr 2023. In: Search for FDA guidance documents. U.S. Food and Drug Administration, 2018
2. Horn L, Mansfield AS, Szczęsna A, Havel L, Krzakowski M, Hochmair MJ, et al. First-line atezolizumab plus chemotherapy in extensive-stage small-cell lung cancer. N Engl J Med. 2018;379:2220–9. https://doi.org/10.1056/NEJMoa1809064.
3. Peters S, Camidge DR, Shaw AT, Gadgeel S, Ahn JS, Kim DW, et al. Alectinib versus crizotinib in untreated ALK-positive non–small-cell lung cancer. N Engl J Med. 2017;377:829–38. https://doi.org/10.1056/NEJMoa1704795.
4. Saura C, Oliveira M, Feng YH, Dai MS, Chen SW, Hurvitz SA, et al. Neratinib plus capecitabine versus lapatinib plus capecitabine in HER2-positive metastatic breast cancer previously treated with ≥ 2 HER2-directed regimens: phase III NALA trial. J Clin Oncol. 2020;38:3138–49. https://doi.org/10.1200/JCO.20.00147.
5. Robson M, Im SA, Senkus E, Xu B, Domchek SM, Masuda N, et al. Olaparib for metastatic breast cancer in patients with a germline BRCA mutation. N Engl J Med. 2017;377:523–33. https://doi.org/10.1056/NEJMoa1706450.
6. Van Cutsem E, Peeters M, Siena S, Humblet Y, Hendlisz A, Neyns B, et al. 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. J Clin Oncol. 2007;25:1658–64. https://doi.org/10.1200/JCO.2006.08.1620.
7. Qin S, Li J, Wang L, Xu J, Cheng Y, Bai Y, et al. Efficacy and tolerability of first-line cetuximab plus leucovorin, fluorouracil, and oxaliplatin (FOLFOX-4) versus FOLFOX-4 in patients with RAS wild-type metastatic colorectal cancer: the open-label, randomized, phase III TAILOR trial. J Clin Oncol. 2018;36:3031–9. https://doi.org/10.1200/JCO.2018.78.3183.
8. Saltz LB, Clarke S, Díaz-Rubio E, Scheithauer W, Figer A, Wong R, et al. Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study. J Clin Oncol. 2008;26:2013–9. https://doi.org/10.1200/JCO.2007.14.9930.
9. Lee MS, Ryoo BY, Hsu CH, Numata K, Stein S, Verret W, et al. Atezolizumab with or without bevacizumab in unresectable hepatocellular carcinoma (GO30140): an open-label, multicentre, phase 1b study. Lancet Oncol. 2020;21:808–20. https://doi.org/10.1016/S1470-2045(20)30156-X.
10. Finn RS, Qin S, Ikeda M, Galle PR, Ducreux M, Kim TY, et al. Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. N Engl J Med. 2020;382:1894–905. https://doi.org/10.1056/NEJMoa1915745.
11. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45:228–47. https://doi.org/10.1016/j.ejca.2008.10.026.
12. Bogaerts K, Komarek A, Lesaffre E. Survival analysis with interval-censored data: A practical approach with examples in R, SAS, and BUGS. Boco Raton: CRC Press; 2017.
13. Law CG, Brookmeyer R. Effects of mid-point imputation on the analysis of doubly censored data. Stat Med. 1992;11:1569–78. https://doi.org/10.1002/sim.4780111204.
14. Nishikawa M, Tango T. Behavior of the Kaplan-Meier estimator for deterministic imputations to interval-censored data and the Turnbull estimator. Jpn J Biom. 2003;24:71–94. https://doi.org/10.5691/jjb.24.71.
15. Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics. 1982;38:29–41. https://doi.org/10.2307/2530286.
16. Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. J R Stat Soc B (Methodol). 1976;38:290–5. https://doi.org/10.1111/j.2517-6161.1976.tb01597.x.
17. Gentlemen R, Greyer CJ. Maximum likelihood for interval censored data: consistency and computation. Biometrika. 1994;81:618–23. https://doi.org/10.2307/2337135.
18. Wellner JA, Zhan Y. A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. J Am Stat Assoc. 1997;92:945–59. https://doi.org/10.1080/01621459.1997.10474049.
19. Anderson-Bergman C. An efficient implementation of the EMICM algorithm for the interval censored NPMLE. J Comp Graph Stat. 2017;26:463–7. https://doi.org/10.1080/10618600.2016.1208616.
20. Anderson-Bergman C. icenReg: regression models for interval censored. Data. Version 2.0; 8. https://CRAN.R-project.org/package=icenReg
21. Project data sphere; 2022. https://data.projectdatasphere.org/projectdatasphere/html/home. Accessed 24 Apr 2023.