



Published in final edited form as:

Stat Med. 2020 June 30; 39(14): 1952–1964. doi:10.1002/sim.8523.

Nonparametric Estimation of Broad Sense Agreement Between Ordinal and Censored Continuous Outcomes

Tian Dai¹, Ying Guo¹, Limin Peng¹, Amita Manatunga¹

¹Department of Biostatistics and Bioinformatics, Emory University, 1518 Clifton Rd, NE., Atlanta, GA 30322, USA.

Abstract

The concept of broad sense agreement (BSA) has recently been proposed for studying the relationship between a continuous measurement and an ordinal measurement (Peng et al. [1]). They developed a non-parametric procedure for estimating the BSA index, which is only applicable to completely observed data. In this work, we consider the problem of evaluating BSA index when the continuous measurement is subject to censoring. We propose a nonparametric estimation method built upon a derivation of a new functional representation of the BSA index, which allows for accommodating censoring by plugging in the nonparametric survival function estimators. We establish the consistency and asymptotic normality for the proposed BSA estimator. We also investigate an alternative approach based on the strategy of multiple imputation, which is shown to have better empirical performance with small sample sizes than the plug-in method. Extensive simulation studies are conducted to evaluate our proposals. We illustrate our methods via an application to a Surgical Intensive Care Unit study.

Keywords

Nonparametric Estimation; Agreement; Censored Observations; Multiple Imputation

1 | INTRODUCTION

In biomedical research, agreement studies are often carried out to evaluate the similarity of measurements obtained by different raters, or to assess whether a new instrument can adequately reproduce the result of a “gold standard” instrument. Various methods have been developed for studying the agreement between categorical measurements ([2],[3],[4], [5],[6], [7],[8] among others) and the agreement between continuous outcomes ([9],[10],[11],[12], [13],[14],[15] among others). These methods are confined to the applications of comparing measurements on the same scale.

Peng et al. [1] proposed the concept of broad sense agreement (BSA), which lays the foundation for a new framework for assessing the correspondence between a continuous scale and an ordinal scale. Peng et al. [1] proposed a sensible BSA index/measure and

developed a nonparametric estimation procedure for it. The BSA index is a chance-corrected agreement measure that lies between -1 and 1 . A higher value of BSA indicates stronger alignment between the ordinal scale and continuous scale. A motivating example from Peng et al. [1] is the Melanoma and Depression study Musselman et al. [16], in which the depression was measured by the clinician-administered Hamilton Depression Scale (HAM-D) and self-reported dimensional scale (Carroll-D). The clinician-administered HAM-D provided a well-defined depression grade: no depression, mild depression, and severe depression, while the self-reported Carroll-D provided a continuous score. The BSA was applied to assess whether the less time-consuming Carroll-D could provide consistent results as compared with HAM-D to determine the grade of depression. The estimated BSA from the study was 0.941 , which was close to the upbound of 1 . This high value of BSA indicated a high capability of finding interpretable cut-off points of Carroll-D that would lead to highly consistent ordinal categorical depression grades based on the HAM-D. A more detailed review of BSA is presented in Section 2.1.

In this work, we aim to extending the BSA framework by developing BSA estimators which can accommodate censored continuous measures. For example, in intensive care unit studies, disease severity scores are often used by clinicians for classifying patients to different risk groups. A number of studies have been conducted to evaluate the relationship between the disease severity scoring and the risk of mortality ([17],[18]). However, it remains unknown to which extent the risk grouping method based on disease severity scores (ordinal) is concordant with disease-related survival times. One major challenge in this example is that disease-related survival times are often subject to censoring due to ICU discharge.

Simulation results in section 3 show that directly applying Peng et al. [1]'s estimation procedure with censored observations will result in considerably biased estimates. To the best of our knowledge, this work is the first to address censored data in the BSA framework.

The presence of censoring to the continuous measurement can greatly complicate the estimation and inference of the BSA index. To address this challenge, we first derive a new functional representation of the BSA index, which delineates the dependency of the BSA index on the the distributions of ordinal and continuous measurements. Such a new representation facilitates the development of a new plug-in type estimator of BSA index which handles censoring via plugging in existing nonparametric survival distribution estimators for censored data. We show the proposed estimator is consistent and asymptotically normal and works well with moderate to large sample sizes. The proposed estimator also provides computational advantages over Peng et al. [1]'s estimation procedure, and can be used as an alternative approach for cases without censoring.

One limitation of the new plug-in estimator is that its empirical performance may be less satisfactory when the sample size is small. To achieve improved estimation accuracy with small datasets, we investigate an alternative nonparametric estimator based on multiple imputation techniques. The imputation-based estimator demonstrate good small sample performance at the cost of additional computational intensity.

The remainder of this paper is organized as follows. In Section 2, we first provide a short review of existing estimation method for the BSA measure and then present the proposed plug-in and imputation-based estimation methods. We establish the asymptotic properties and develop inference procedures for the proposed methods. In Section 3, we conduct simulation studies to evaluate the performance of our proposed estimators in comparison with the naive methods that either exclude censored observations or treat censored observations as observed events. The results show that the proposed estimators successfully correct the bias of the naive methods in the presence of censoring. We then illustrate our proposed methods via an application to a surgical ICU data in Section 4. Finally, we conclude with some remarks in Section 5.

1.1 | Methods

In this section, we provide a brief review of the broad sense agreement (BSA) concept, the BSA index and its estimation procedure in Peng et al. [1]. Let X and Y denote a continuous and an ordinal measure of a common outcome from the same subject. Let \mathfrak{D}_X and \mathfrak{D}_Y be the domain of X and Y , respectively. Peng et al. [1] provided the definition of perfect broad sense agreement (disagreement) between X and Y if and only if there exists an increasing (decreasing) step function ψ from \mathfrak{D}_X to \mathfrak{D}_Y such that $Y = \psi(X)$ with probability 1. The definition of perfect broad sense agreement (disagreement) implies that one is able to identify a set of cut-points for the continuous X such that the discretized X is in perfect concordance (discordance) with the ordinal Y . Therefore, the BSA framework defined upon the above concept provides useful information on the degree of concordance between a continuous measure and an ordinal measure.

the perfect broad sense agreement entails a scenario where, for randomly selected X with $Y = l$, denoted by $X_{(*l)}$, it must be satisfied that $X_{(*1)} \cdots < X_{(*L)}$

Based on the above definition, Peng et al. [1] further introduced the definition of a BSA index, denoted as ρ_{bsa} , to characterize the extent to which the relationship between X and Y departs from the perfect BSA scenario. Denote Y as an ordinal variable that takes values $1 < \cdots < L$, where X is a continuous variable. The perfect broad sense agreement entails a scenario where, for randomly selected X with $Y = l$, denoted by $X_{(*l)}$, it must be satisfied that $X_{(*1)} < \cdots < X_{(*L)}$, i.e. the ordering of the continuous measurements should perfectly match the ordering of their corresponding ordinal measurements. The BSA measure is defined based on the mean square distance between the observed ranks of $\{X_{(*1)} < \cdots < X_{(*L)}\}$ and their expected ranks under the scenario of perfect BSA. Specifically, we denote the observed ranks of $\{X_{(*1)} < \cdots < X_{(*L)}\}$ by (R_1, \cdots, R_L) . Based on the previous description, the expected rank of $\{X_{(*1)} < \cdots < X_{(*L)}\}$ under perfect BSA is $(1, \cdots, L)$. Therefore, the BSA measure takes the form,

$$\rho_{bsa} = 1 - \frac{E\left\{\sum_{l=1}^L (l - R_l)^2\right\}}{E\left\{\sum_{l=1}^L (l - R_l)^2 \mid X \perp Y\right\}}. \tag{1}$$

The ρ_{bsa} is a scaled global measure that takes values from -1 to 1 , with larger value indicates better BSA and a value of $1(-1)$ indicates perfect broad sense agreement

(disagreement). To demonstrate these extreme cases, recall that under the perfect broad sense agreement, it must be satisfied that $X_{(*1)} < \dots < X_{(*L)}$. In the contrary case with perfect broad sense disagreement, the ranks of $X_{(*1)} < \dots < X_{(*L)}$ are reversed; that is, $X_{(*1)} > \dots > X_{(*L)}$. In other words, when X and Y are in perfect broad sense agreement (or disagreement), $(R_1, \dots, R_L) = (1, \dots, L)$ (or $(L, \dots, 1)$) with probability 1 across random samples of X . Therefore, when there is perfect broad sense agreement, the BSA index $\rho_{bsa} = 1$ given that $E\left\{\sum_{l=1}^L (l - R_l)^2\right\}$ equals 0. For the perfect broad sense disagreement scenario, it can be shown that $\rho_{bsa} = -1$ through some derivation. The details are presented in Appendix A in Peng et al. [1] paper.

Peng et al. [1] developed a nonparametric estimation procedure for the proposed BSA index, i.e. ρ_{bsa} . Suppose the observed data consist of n complete random samples of (X, Y) , denoted by $\{x_i, Y_i\}_{i=1}^n$, and can be arranged based on the ordering of Y as follows,

$$\begin{aligned} &(X_1, Y_1 = 1), \dots, (X_{n_1}, Y_{n_1} = 1), \\ &(X_{n_1 + 1}, Y_{n_1 + 1} = 2), \dots, (X_{n_1 + n_2}, Y_{n_1 + n_2} = 2), \\ &\vdots \\ &(X_{\sum_{l=1}^{L-1} n_l + 1}, Y_{\sum_{l=1}^{L-1} n_l + 1} = L), \dots, (X_{\sum_{l=1}^L n_l}, Y_{\sum_{l=1}^L n_l} = L), \end{aligned}$$

where $n_l = \sum_{i=1}^n I(Y_i = l)$ is the number of subjects in the l th level of Y and $\sum_{l=1}^L n_l = n$. Peng et al. [1] showed that under the independence assumption between X and Y , the expected mean square distance can be determined based on the number of Y levels, i.e., $E\left\{\sum_{l=1}^L (l - R_l)^2 \mid X \perp Y\right\} = C_L = (L^3 - L)/6$.

Adopting the idea of stratified sampling without replacement, Peng et al. [1] proposed to estimate $E\left\{\sum_{l=1}^L (l - R_l)^2\right\}$ by the sample mean of the mean square distance based on stratified samples without replacement, i.e.

$$w_n = \left(\prod_{l=1}^L n_l\right)^{-1} \sum_{j_1=1}^{n_1} \dots \sum_{j_L=1}^{n_L} \sum_{l=1}^L \left\{l - \sum_{r=1}^L I(\tilde{X}_{l, j_l} \geq \tilde{X}_{r, j_r})\right\}^2.$$

Note that $(\tilde{X}_{1, j_1}, \dots, \tilde{X}_{L, j_L})$ is a random realization of $X_{(*1)}, \dots, X_{(*L)}$ and $\sum_{r=1}^L I(\tilde{X}_{l, j_l} \geq \tilde{X}_{r, j_r})$ is the rank of \tilde{X}_{l, j_l} among $(\tilde{X}_{1, j_1}, \dots, \tilde{X}_{L, j_L})$ with $\tilde{X}_{l, j_l} = X_{\sum_{r=1}^{l-1} n_r + j_l}$ where $1 \leq j_l \leq n_l$ for $1 \leq l \leq L$.

Therefore, Peng et al. [1]'s estimator of ρ_{bsa} is given by

$$\hat{\rho}_{bsa} = \frac{W_n}{C_L} = 1 - \frac{\left(\prod_{l=1}^L n_l\right)^{-1} \sum_{j_1=1}^{n_1} \cdots \sum_{j_L=1}^{n_L} \sum_{l=1}^L \left\{l - \sum_{r=1}^L I(\tilde{X}_{l,j_l} \geq \tilde{X}_{r,j_r})\right\}^2}{(L^3 - L)/6}. \quad (2)$$

The asymptotic variance of $\hat{\rho}_{bsa}$ is estimated using the jackknife method ([1]).

1.2 | A new functional representation of BSA

Peng et al. [1]'s estimation procedure for the BSA index required completely observed data. In biomedical data, the continuous measure may be censored due to many reasons, such as loss of follow-up. Simulation results in section 3 show that Peng et al. [1]'s estimator will result in considerably biased estimates if directly applied to censored observations. We note that the estimation of the denominator term of BSA index is fairly simple and does not require any information from the continuous X . The key of estimating BSA lies in the estimation of the numerator term $E\left\{\sum_{l=1}^L (l - R_l)^2\right\}$.

We first derive a new functional representation of the BSA index. Specifically, denote the conditional survival function of $[X|Y=l]$ as $S_l(x) = Pr(X > x | Y = l)$, $l = 1, \dots, L$. We can show that the BSA measure in Equation (1) can be written as a functional of the L conditional survival functions $\{S_1(\cdot), \dots, S_L(\cdot)\}$. That is,

$$\rho_{bsa} = \frac{2 \sum_{r=1}^{L-1} \sum_{l=r+1}^L (l-r) \int_0^\infty S_l(x) dS_r(x)}{(L^3 - L)/6} - 1. \quad (3)$$

Detailed derivation of Equation (3) can be found in [19]. This result holds for continuous measurements X with a finite lower bound. In the paper, without loss of generality, we assume the lower bound is zero. Let S_0 be the collection of a finite dimensional vector survival functions \vec{S} on \mathbb{R}^L with finite support $[0, \tau_1] \times \cdots \times [0, \tau_L]$. Let $\|\tilde{S}_1 - S_1\|_\infty \equiv \sup_x |\tilde{S}_1(x) - S_1(x)| \equiv \|\tilde{S}_1 - S_1\|_{\mathcal{S}}$ where $\mathcal{S} = \{1_{(-\infty, x]} : x \in \mathbb{R}\}$. Then the functional $T: S_0 \mapsto \mathbb{R}$

$$T(\vec{S}) = \frac{2 \sum_{r=1}^{L-1} \sum_{l=r+1}^L (l-r) \int_0^\infty S_l d(-S_r)}{(L^3 - L)/6} - 1.$$

We note that the new representation of ρ_{bsa} only consists of several conditional survival functions, which can be easily estimated with censored data based on existing software. Equation (3) naturally motivates us to construct an estimator of the BSA measure by plugging in the estimators of $\{S_1(\cdot), \dots, S_L(\cdot)\}$.

Suppose the observed data consist of n random samples $\{\tilde{X}_i, Y_i, \delta_i\}_{i=1}^n$, where \tilde{X}_i is the observed continuous measurement for subject i , which is defined as the minimum of the continuous measurement X_i and the censoring variable C_i , and δ_i is the censoring indicator,

which equals zero if the observation is censored, and one if not censored. C and X are independent conditionally on Y . Let survival function $G_I(x) = Pr(C > x | Y = I)$. For any $I \in \{1, \dots, L\}$, $S_I(x)$ can be consistently estimated by the Kaplan Meier estimator $\hat{S}_I(x)$ stratified for Y for any $x \in [0, \tau_I^*]$, where $\tau_I^* = \sup\{x: S_I(x)G_I(x) > 0\}$ (Breslow and Crowley, 1974; Wang, 1987; Cai, 1997). Given (1), we propose a nonparametric plug-in estimator for ρ_{bsa} :

$$\hat{\rho}_{bsa} = \frac{2 \sum_{r=1}^{L-1} \sum_{l=r+1}^L (l-r) \int_0^{\hat{\tau}_l \wedge \hat{\tau}_r} \hat{S}_l(x) d\hat{S}_r(x)}{(L^3 - L)/6} - 1, \tag{4}$$

where $\hat{\tau}_I = \max(\tilde{X}_{I,1}, \dots, \tilde{X}_{I,n_I})$, $a \wedge b = \min(a, b)$. Note that the Kaplan-Meier estimators can be readily obtained by standard software. Given that $\hat{S}_I(x)$ and $\hat{S}_r(x)$ are piecewise-constant functions, the integrals in (4) can be easily computed through a finite summation. Thus, $\hat{\rho}_{bsa}$ is a computationally simple nonparametric estimator. [19] shows the proposed nonparametric plug-in estimator for ρ_{bsa} in (4) is equivalent to the estimator presented in Peng et al. [1] when there is no censoring. Through simulation studies, [19] shows the nonparametric plug-in estimator is computationally much faster than Peng et al. [1]'s estimator.

In the following, we establish asymptotic properties for the proposed estimator $\hat{\rho}_{bsa}$ via the Hadamard differentiability of the functional T and the statistical properties of stratified Kaplan Meier estimators of the conditional survival functions. For simplicity, we denote $\vec{\hat{S}} = (\hat{S}_1, \dots, \hat{S}_L)'$ and $\vec{S} = (S_1, \dots, S_L)'$. First, we show the functional T is Hadamard differentiable in Lemma 1.

Lemma 1 *Let S_0 be the collection of a finite dimensional vector of survival functions \vec{S} on \mathbb{R}^L with finite support $[0, \tau_1] \times \dots \times [0, \tau_L]$. Let $\|\tilde{S}_1 - S_1\|_\infty \equiv \sup_x |\tilde{S}_1(x) - S_1(x)| \equiv \|\tilde{S}_1 - S_1\|_\mathcal{S}$ where $\mathcal{S} = \{1_{(-\infty, x]} : x \in \mathbb{R}\}$. Then the functional $T: S_0 \mapsto \mathbb{R}$*

$$T(\vec{S}) = \frac{2 \sum_{r=1}^{L-1} \sum_{l=r+1}^L (l-r) \int_0^\infty S_l d(-S_r)}{(L^3 - L)/6} - 1$$

is Hadamard differentiable at \vec{S} with respect to Kolmogorov distance $d_K = \|\cdot\|_\infty$.

The proof of Lemma 1 is provided in Appendix 6.1.

Next, we show the statistical properties of $\vec{\hat{S}}$. Breslow and Crowley (1974) shows that if both S_I and G_I are continuous distributions, on the support $[0, \tau_I^*]$, where $\tau_I^* = \sup\{x: S_I(x)G_I(x) > 0\}$, the Kaplan Meier estimator \hat{S}_I uniformly consistent and $\sqrt{n}(\hat{S}_I(x) - S_I(x))$, $x \in [0, \tau_I^*]$ converges weakly to a Gaussian process W with mean 0 and

covariance function $Cov(W_I(s), W_I(t)) = S_I(s)S_I(t)a_I(s)$, $s \leq t$ where $a_I(s) = \int_0^s \frac{-dS_I}{S_I^2 G_I}$. We further have $\sqrt{n} \left\{ \widehat{S}(x) - \vec{S}(x) \right\}$ weakly converges to a tight, zero mean Gaussian process $\vec{W}(x) = (W_{1(x)}, \dots, W_{L(x)})'$. Based on Lemma 1 and the statistical properties of \widehat{S} , we establish the asymptotic properties of the estimator $\widehat{\rho}_{bsa}$ in the following theorem.

Theorem 1 Assume $\widehat{\tau}_I \xrightarrow{p} \tau_I$, $\tau_I^* = \tau_I$, $\widehat{S}_I(\widehat{\tau}_I) \xrightarrow{p} S_I(\tau_I) = 0$, for any $l, r \in \{1, \dots, L\}$, the proposed estimator $\widehat{\rho}_{bsa}$ has the following asymptotic properties:

(i) The estimator $\widehat{\rho}_{bsa}$ is strongly consistent. That is, $|\widehat{\rho}_{bsa} - \rho_{bsa}| \rightarrow 0$ with probability 1.

(ii) The estimator $\widehat{\rho}_{bsa}$ has the following weak convergence result,

$$\sqrt{n}(\widehat{\rho}_{bsa} - \rho_{bsa}) \rightarrow^d T'_S(\vec{w}),$$

where \vec{w} is the zero-mean Gaussian process and $T'_S(\vec{W})$ follows a zero-mean normal distribution.

The proof of Theorem 1 is provided in Appendix 6.2. We note that the assumption $\tau_I^* = \tau_I$ implies the upper bounds of the survival time is less than or equal the upper bounds of the censoring time. This condition is required because the asymptotic properties of the Kaplan Meier estimator of the survival function are only valid within $[0, \tau_I^*]$. Our BSA plug-in estimator, which is derived based on the survival function estimators, inherits this limitation. When this assumption is not valid, that is the upper bound of the censoring time is less than that of the survival time, the proposed BSA estimator estimates the broad sense agreement between the survival outcome and the ordinal outcome within the upper bound of the censoring time instead of the entire range of the survival time distribution.

Since BSA estimator is a Hadamard differentiable function of survival function estimator \widehat{S} , it is theoretically possible to attempt an analytical form for the asymptotic variance of the BSA estimator from the asymptotic covariance of the \widehat{S} following functional delta method. However, the analytical expression for the variance of the BSA estimator is technically challenging since the covariance of \widehat{S} is already complicated and the variance of BSA estimator involves the covariance of \widehat{S} through a complicated function. Therefore, we propose to estimate the variance of BSA estimator using resampling method. The resampling method demonstrated good performance through the simulation studies in section 2.

1.3 | An alternative estimator based on multiple imputation

Multiple imputation (MI) is a popular technique for handling missing data. Censoring in survival data is a special form of missing data problem. Taylor et al. [20] proposed a multiple imputation method to handle missing event times for censored observations. The idea was to impute missing event time from the estimated distribution of event times amongst those at risk after the censoring time. Their study demonstrated that nonparametric multiple imputation successfully recovered the missing time information in the estimation of survival distributions. Hsu et al. [21] extended Taylor et al. [20]'s work by incorporating auxiliary variables to define the risk sets and performing multiple imputation only within the risk sets. They showed that with either time-independent or time-dependent auxiliary variables, the multiple imputation approach demonstrated similar results in terms of reduced bias due to dependent censoring and improved efficiency as an inverse probability of censoring weighted method.

In this section, we adopt the ideas from Taylor et al. [20] and Hsu et al. [21] by considering the censoring of the continuous X as a missing data problem and propose an alternative estimator for ρ_{bsa} based on imputing the censored observations of X . In the following, we present the algorithm to impute a censored observation with $\tilde{X} = x_j$, $Y = l$, $\delta = 0$.

Step 1: Identify the risk set for censored observation $(x_j, l, 0)$. Denote the risk set as $R(j^+|l) = \{i : \tilde{X}_i > x_j, Y_i = l, i = 1, \dots, n\}$, which includes all the observations whose Y takes the same value l and whose survival time \tilde{X} is longer than the censored time x_j . Note that observations with different Y values other than l are excluded from the risk set.

Step 2: Estimate the distribution of event times among those at risk at the censored time x_j using the Kaplan-Meier (KM) estimator based on the risk set $R(j^+|l)$. The KM estimator of the survival function of \tilde{X} given $\tilde{X} = x_j$ and $Y = l$ is denoted as $\hat{S}_{j^+|l}(x)$. It is easy to see that $\hat{S}_{j^+|l}(x)$ jumps only at observed X values in $R(j^+|l)$.

Step 3: Impute an event time for x_j by drawing random samples from the empirical distribution of event times estimated in step 2. To impute x_j , generate a random value from the uniform distribution $U(0, 1)$. Find two neighboring observations x_s, x_t from the sorted risk set $R(j^+|l)$, so that $(1 - \hat{S}_{j^+|l}(x_s), 1 - \hat{S}_{j^+|l}(x_t))$ includes α . The imputed value for x_j is defined as x_s .

Repeat the above algorithm for all censored observations until they have been imputed. This procedure imputes the censored observations with the observed values unless the largest observation is censored in which case some imputed values may include this largest value. With M sets of imputations, there are M enhanced data sets and hence M BSA estimates with associated jackknife variances, say $\hat{\rho}_{bsa}^m$ and \hat{U}_{bsa}^m , respectively. The imputation-based estimator of BSA is defined as the average of the M imputation-based BSA estimates:

$$\bar{\rho}_{bsa}^m = \sum_{m=1}^M \hat{\rho}_{bsa}^m / M.$$

The variance estimator of $\hat{\rho}_{bsa}^M$ is $W = \bar{U} + (1 + M^{-1})B$ ([22]), which consists of two components: the average within-imputation variance, which is the average of variance estimates from the imputed data sets, i.e., $\bar{U} = \sum_{m=1}^M \hat{U}_{bsa}^m / M$, and the between-imputation variance, which is the sample variance of the imputed-data BSA estimates, i.e.,

$B = \sum_{m=1}^M \{ \hat{\rho}_{bsa}^m - \bar{\rho}_{bsa} \}^2 / (M - 1)$. The confidence interval can be constructed using fisher's z transformation.

To incorporate the full uncertainty in the imputation, a bootstrapping stage can be added when estimating the survival distributions. For $Y = I$, consider the bootstrap sample

$\{ (X_1^{(B)}, I, \delta_1^{(B)}), \dots, (X_n^{(B)}, I, \delta_n^{(B)}) \}$ selected with replacement from the original data set. The

imputed risk set for the censored time $(x_j, I, 0)$ can be redefined as

$R^{(B)}(j^+ | I) = \{ i : x_i^{(B)} > x_j, Y_i = I, i = 1, \dots, n \}$ to include those observations that are at risk at time x_j in the bootstrap sample.

2 | SIMULATION RESULTS

We conducted Monte Carlo simulations to evaluate the finite sample performance of the proposed methods. Specifically, we compared empirical bias, standard deviation and coverage rates of 95% confidence intervals when sample size, censoring pattern and censoring rate varied in the simulation. We considered two sample sizes: $N=60, 120$ representing small and moderate sample sizes, respectively. We compared the performance of the proposed methods with two naive methods: Peng et al. [1]'s estimator based on a partial dataset that excludes censored subjects, i.e. $\{ (X_i, Y_i, \delta_i) : \delta_i = 1, i = 1, \dots, n \}$ and Peng et al. [1]'s estimator using the whole dataset while ignoring the censoring to X.

Simulated data was generated as follows. We assumed $L = 3$ and generated Y from $\{1, 2, 3\}$ with equal probability. Continuous variable X was generated from a non-normal distribution, $Y + Weibull(2, \xi)$, where the scale parameter ξ was chosen to simulate different levels of BSA between X and Y . Censoring time was generated independently of continuous variable X conditional on Y from a non-normal distribution, $\pi^C Y + Weibull(2, \xi^C)$, where ξ^C was selected to achieve different censoring percentages and π^C controlled the balance of censoring proportions among Y levels. If π^C takes the value of 1, censoring rates were the same for different Y . If $\pi^C = 0.5$, censoring rate was higher for larger Y values. Results in the following tables were based on 500 simulated datasets.

Tables 1–2 show that when homogeneous censoring proportions of X were presented across Y levels, naive methods that either exclude censored observations or treat censored observations as complete observations (referred to as complete obs. and all obs. respectively) produce similar bias for all censoring proportions and their bias increases dramatically as censoring proportion of X increases. For example, with moderate BSA, i.e., $\rho_{bsa} = 0.550$, the bias of naive estimators increases from 10% to 42% of the true value of BSA as censoring proportion increases from 27% to 70%. This leads to a considerably different interpretation of the magnitude of BSA, where $\rho_{bsa} = 0.550$ is considered moderate BSA and $\rho_{bsa} = 0.8$ indicates fairly strong BSA between X and Y . The results also demonstrate that both plug-in

method and MI-based methods significantly reduce the bias compared to naive methods. The plug-in method performs best when there is strong BSA and the MI-based methods outperform the plug-in method in moderate BSA scenarios.

When the censoring rates are heterogeneous across Y levels, the plug-in estimator doesn't perform very well in the small sample scenario with moderate BSA (Table 1) and has similar or slightly larger bias as compared to the naive methods. But its performance does improve significantly in the moderate sample size case where it generally shows lower bias than the naive methods. The MI methods always outperform the naive methods, demonstrating much lower bias.

For both homogeneous and heterogeneous censoring rate scenarios, as sample size increases, the bias of plug-in method decreases while the biases of naive methods remain the same or even increase. The proposed variance estimator of plug-in method seems to overestimate the variance when there is heavy censoring. The coverage probabilities of the estimated confidence intervals are close to the nominal level (95%) at most cases. The only exception is when there is heavy homogeneous censoring in the high BSA data, in which cases all methods fail. The variance estimator of MI method without the bootstrap procedure always underestimates the variance. After adding the bootstrap procedure, the variance estimation improves significantly and the constructed confidence intervals have reasonably good coverage probabilities especially for small to moderate censoring rates.

In Table 4, we compare the computation times of different estimation methods for estimating BSA index and jackknife standard error in one simulation. The table shows that the computing costs of Peng et al. [1]'s estimator and MI-based estimators increase dramatically with the increase of sample size. For a sample size of 240, the computing time for Peng et al. [1]'s estimator is 4077 seconds while the computing time for the proposed plug-in method is only 5 seconds. The computing time of both MI-based methods are very similar and is nine times (which is the number of imputations in each simulation) longer than that of Peng et al. [1]'s estimator. Given the results in Table 4, the plug-in method is more appropriate for large sample size datasets and the MI-based method may be used for small sample size data.

3 | AN APPLICATION TO A SURGICAL ICU PATIENTS STUDY

Acute Physiology and Chronic Health Evaluation II, often known as APACHE II, is a severity-of-disease classification system which has been extensively used in intensive care unit (ICU) to assess the morbidity of patients and stratify risk of death ([23]). Many literatures have shown a significant correlation between APACHE-II score and the probability of hospital-related mortality as well as hospital-acquired infections ([24],[25]). However, we sometimes observe that severely sick patients die or acquire infections shortly after ICU or hospital discharge. In this case, survival endpoints, such as progression free survival (PFS), can be adopted as an alternative way to assess the risk of hospital-related mortality and infections.

In this study, 150 patients requiring postoperative surgical ICU (SICU) care were enrolled. PFS was measured as the time from first day in SICU to death or first severe infection. PFS

was censored by the hospital discharge date if no events happened during the hospital stay. Six patients who had preconditioned bloodstream infections (BSI) and lower respiratory tract infections (LRI) were excluded from the analysis. For the rest 144 patients, 24 hospital-related deaths were observed, and 66 hospital-acquired infection incidences were observed, which included 30 BSI and 36 LRI. Eighty three patients were observed to be event-free during their hospital stays since SICU. Two risk groups are determined based on clinical guideline ([26], [17]) using APACHE-II score calculated upon admission to the SICU: APACHE II score 0–24 correspond to low risk group; and APACHE II 25 correspond to high risk group. Our main interest was to study the relationship between APACHE-II risk groups and PFS, in which PFS was defined as a composite endpoint of time to first infection/death.

Table 4 presents the estimated ρ_{bsa} and the associated standard errors and confidence intervals. We assume that we assume that hospital discharge is a random censoring event to PFS. Fisher's z-transformation is used to compute confidence intervals. Naive estimates for ρ_{bsa} directly using the whole dataset without adjusting for censoring or using only observed subjects, are close to 0, which suggests no beyond-chance agreement between risk group and PFS. However, estimates obtained using plug-in method or both multiple imputation methods are about 0.27–0.30, which indicates fair agreement. The associated confidence intervals exclude 0 for all three proposed methods. We conclude that there is significant broad sense agreement between risk group and PFS but the magnitude of the BSA is in the low range. Since the censoring is related to hospital discharge, there is a possibility that the upper bound of the censoring time may be less than the upper bound of the PFS time depending on the hospital discharge policy. If that is the case, the estimated BSA measures the broad sense agreement between PFS and the APACHE-II within the upper bound of the hospital discharge time. We also note that the independent censoring assumption may be questionable for the data example since censoring is related to hospital discharge. This may affect the accuracy of the proposed estimator and inference procedure. Deeper investigation is needed in future studies to fully study the impact when the independent censoring assumption is violated. Another limitation of the application is that APACHE was developed to mainly address in-hospital morbidity and mortality while the PFS measured in the current study captures both in hospital morbidity and mortality (before discharge) as well as potentially out of hospital events. For future studies with APACHE score, it will be helpful to define and measure a more meaningful endpoint that would restrict to events before discharge.

4 | CONCLUSION

In this paper, we propose novel estimation methods for broad sense agreement index when the continuous measure is subject to censoring. Specifically, we propose a plug-in method based on the conditional survival distributions, which is computationally efficient and has desirable theoretical properties. In addition, we propose another estimation method for smaller data sets using multiple imputation techniques. We develop inference procedures for the proposed estimators and demonstrate the small-sample performance of the proposed methods via simulation studies.

The application of the proposed plug-in estimator is not limited to the case when the continuous measurement is subject to censoring. In fact, the new plug-in estimator can be a useful alternative to Peng et al. [1]'s estimator when the sample size is large. In particular, the plug-in method can provide a much more computationally efficient estimation of the BSA by dramatically reducing the computation time required by Peng et al. [1]'s estimator.

5 | DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Funding information

NIH, Grant/Award Number:R01

MH118771,R01MH079448,R01MH105561,UL1TR002378

6 | APPENDIX

6.1 | Proof of Lemma 1

Proof First, we give the definition of Hadamard differentiability (Gill, 1989; Wellner, 1989).

A function $T: S \rightarrow R$ is Hadamard differentiable at S with respect to the Kolmogorov distance if there exists $\dot{T}(S; \cdot)$ continuous and linear satisfying

$$\frac{|T(S_X) - T(S) - \dot{T}(S; S_X - S)|}{\|S_X - S\|} = o(1) \text{ for all } S_X \text{ satisfying } \|S_X - S\|_{\infty} \rightarrow 0 \text{ for some}$$

function \cdot .

For any $S_{lx} \rightarrow S_l$, where $l = 1, \dots, L$, define $\alpha_{lx} \equiv (S_{lx} - S_l)/x$. For Hadamard differentiability we have $\alpha_{lx} \rightarrow \alpha_l$ with respect to $\|\cdot\|_{\infty}$ for some (bounded) function α_l . Denote $\vec{\alpha} = (\alpha_1, \dots, \alpha_L)$. Define

$$\dot{T}(S; \vec{\alpha}) \equiv \frac{2 \sum_{r=1}^{L-1} \sum_{l=r+1}^L (l-r) (\int \alpha_r dS_l - \int \alpha_l dS_r)}{(L^3 - L)/6}$$

We have

$$\begin{aligned} & \frac{T(S_{1x}, \dots, S_{Lx}) - T(S_1, \dots, S_L)}{x} - \dot{T}(\alpha_{1x}, \dots, \alpha_{Lx}) \\ &= \frac{2 \sum_{r=1}^{L-1} \sum_{l=r+1}^L (l-r) x \int \alpha_{lx} d\alpha_{rx}}{(L^3 - L)/6} \\ &= \frac{2 \sum_{r=1}^{L-1} \sum_{l=r+1}^L (l-r) [\int \alpha_l d(S_{rx} - S_r) + \int (\alpha_{lx} - \alpha_l) d(S_{rx} - S_r)]}{(L^3 - L)/6} \end{aligned}$$

Since \dot{T} is continuous, it suffices to show that the right side converges to 0. For any $l, r \in \{1, \dots, L\}$ and $l > r$,

$$\int (\alpha_{lx} - \alpha_l)d(S_{rx} - S_r) \leq \|\alpha_{lx} - \alpha_l\|_\infty \left\{ \int dS_{rx} + \int dS_r \right\} \leq 2\|\alpha_{lx} - \alpha_l\|_\infty \rightarrow 0$$

Fix $\epsilon > 0$, since the limit function α_l is right continuous with left limits, there is a step function with a finite number m of jumps, say $\tilde{\alpha}$, which satisfies $\|\alpha - \tilde{\alpha}\| < \epsilon$. Thus,

$$\begin{aligned} \left| \int \alpha_l d(S_{rx} - S_r) \right| &\leq \left| \int (\alpha_l - \tilde{\alpha}_l) d(S_{rx} - S_r) \right| + \left| \int \tilde{\alpha}_l d(S_{rx} - S_r) \right| \\ &\leq 2\|\alpha_l - \tilde{\alpha}_l\|_\infty + \sum_{j=1}^m |\tilde{\alpha}_l(x_j - 1)| \|(S_{rx} - S_r)[x_j - 1, x_j]\| \\ &\leq 2\epsilon + 2m\|\tilde{\alpha}_l\|_\infty \|S_{rx} - S_r\|_\infty \rightarrow 2\epsilon. \end{aligned}$$

Since ϵ is arbitrary, this completes the proof that T is Hadamard differentiable.

6.2 | Proof of theorem 1

Proof Before proving Theorem 1, we first show

$$\sqrt{n}|\hat{\rho}_{bsa} - T(\hat{S}_1, \dots, \hat{S}_L)| \xrightarrow{P} 0 \tag{5}$$

Consider the difference between $\sqrt{n}\hat{\rho}_{bsa}$ and $\sqrt{n}T(\hat{S}_1, \dots, \hat{S}_L)$,

$$\begin{aligned} &\sqrt{n}|\hat{\rho}_{bsa} - T(\hat{S}_1, \dots, \hat{S}_L)| \\ &= \sqrt{n} \left| \frac{2 \sum_{r=1}^{L-1} \sum_{l=r+1}^L - (l-r) \int_0^{\hat{\tau}_l \wedge \hat{\tau}_r} \hat{S}_l d\hat{S}_r}{(L^3 - L)/6} - 1 \right. \\ &\quad \left. - \frac{2 \sum_{r=1}^{L-1} \sum_{l=r+1}^L - (l-r) \int_0^\infty \hat{S}_l d\hat{S}_r}{(L^3 - L)/6} + 1 \right| \\ &= \frac{2\sqrt{n}}{(L^3 - L)/6} \left| \sum_{r=1}^{L-1} \sum_{l=r+1}^L - (l-r) \int_0^{\hat{\tau}_l \wedge \hat{\tau}_r} \hat{S}_l d\hat{S}_r - \sum_{r=1}^{L-1} \sum_{l=r+1}^L - (l-r) \right. \\ &\quad \left. \int_0^\infty \hat{S}_l d\hat{S}_r \right| \\ &\leq \frac{2\sqrt{n}}{(L^3 - L)/6} \sum_{r=1}^{L-1} \sum_{l=r+1}^L (1-r) \left| \int_{\hat{\tau}_l \wedge \hat{\tau}_r}^\infty \hat{S}_l d\hat{S}_r \right| \tag{6} \end{aligned}$$

Given the assumptions that $S_l(\hat{\tau}_l) \xrightarrow{P} S_l(\tau_l) = 0$ and $S_r(\hat{\tau}_r) \xrightarrow{P} S_r(\tau_r) = 0$ and the strong uniform consistency of the KM estimators \hat{S}_l and \hat{S}_r , the right-hand side of inequality (6) converges to zero in probability.

Next, we prove (i), (ii) of the theorem. The result of $T(\hat{S}_1, \dots, \hat{S}_L) \xrightarrow{P} \rho_{bsa} = T(S_1, \dots, S_L)$ follows the continuity of the Hadamard differentiable function T and the uniform strong consistency of Kaplan Meier estimators \widehat{S} . Then $\hat{\rho}_{bsa} \xrightarrow{P} \rho_{bsa}$ via Equation (5). Thus, the statement (i) is true.

According to Equation (5), to prove (ii) is equivalent to show that

$$\sqrt{n}(T(\hat{S}_1, \dots, \hat{S}_L) - T(S_1, \dots, S_L)) \xrightarrow{d} T'_{\{S_1, \dots, S_L\}}(\vec{W}), \quad \text{as } n \rightarrow \infty \quad (7)$$

It's shown that $\sqrt{n}\{\widehat{S} - \vec{S}\}$ weakly converges to a tight, zero mean Gaussian process \vec{W} . The functional T is proven to be Hadamard-differentiable. Then, statement (7) is true according to the functional delta method (van der Vaart and Wellner, 1996). Because \vec{W} is a tight Gaussian process, the derivative $T'_{\{S_1, \dots, S_L\}}(\vec{W})$ is normally distributed. Thus, the statement (ii) of the theorem is proven true.

references

- [1]. Peng L, Li R, Guo Y, Manatunga A. A Framework for Assessing Broad Sense Agreement Between Ordinal and Continuous Measurements. *Journal of the American Statistical Association* 2011;106(496):1592–1601. [PubMed: 23399559]
- [2]. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 1960;20(1):37–46.
- [3]. Fleiss JL. Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin* 1971;76(5):378—382.
- [4]. Kraemer HC. Extension of the Kappa Coefficient. *Biometrics* 1980;36(2):207—216. [PubMed: 7190852]
- [5]. Williamson JM, Lipsitz S, Manatunga AK. Modeling Kappa for Measuring Dependent Categorical Agreement Data. *Biostatistics* 2000;1(2):191–202. [PubMed: 12933519]
- [6]. Barnhart HX, Williamson JM. Weighted Least-Squares Approach for Comparing Correlated Kappa. *Biometrics* 2002;58(4):1012—1019. [PubMed: 12495157]
- [7]. Guo Y, Manatunga AK. Modeling the Agreement of Discrete Bivariate Survival Times using Kappa Coefficient. *Lifetime Data Analysis* 2005;11(3):309–332. [PubMed: 16133882]
- [8]. Guo Y, Manatunga AK. A note on assessing agreement for frailty models. *Statistics and Probability Letters* 2010;80(7–8):527–533. [PubMed: 20607108]
- [9]. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45(1):255–268. [PubMed: 2720055]
- [10]. Lin LI. Assay Validation Using the Concordance Correlation Coefficient. *Biometrics* 1992;48(2):599–604.
- [11]. Lin L, Hedayat A, Sinha B, Yang M. Statistical Methods in Assessing Agreement Models, Issues and Tools. *Journal of the American Statistical Association* 2002;97(457):257–270.
- [12]. Lin L, Hedayat A, Wenting W. A Unified Approach for Assessing Agreement for Continuous and Categorical Data. *Journal of Biopharmaceutical Statistics* 2007;17(4):629–652.
- [13]. Quiroz J. Assessment of Equivalence Using a Concordance Correlation Coefficient in a Repeated Measurement Design. *Journal of Biopharmaceutical Statistics* 2005;15(6):913–928. [PubMed: 16279351]
- [14]. King TS, Chinchilli VM, Carrasco JL, Wang K. A Class of Repeated Measures Concordance Correlation Coefficients. *Journal of Biopharmaceutical Statistics* 2007;17(4):653–672. [PubMed: 17613646]
- [15]. Janson H, Olsson U. A Measure of Agreement for Interval or Nominal Multivariate Observations. *Educational and Psychological Measurement* 2001;61(2):277–289.
- [16]. Musselman DL, Lawson DH, Gumnick JF, Manatunga AK, Penna S, Goodkin RS, et al. Paroxetine for the Prevention of Depression Induced by High-Dose Interferon Alfa. *The New England Journal of Medicine* 2001;344(13):961–966. [PubMed: 11274622]

- [17]. Bouch DC, Thompson JP. Severity scoring systems in the critically ill. *Continuing Education in Anaesthesia Critical Care* 2008;8(5):181–185.
- [18]. Saleh A, Ahmed M, Sultan I, Abdel-Lateif A. Comparison of the mortality prediction of different ICU scoring systems (APACHE II and III, SAPS II, and SOFA) in a single-center ICU subpopulation with acute respiratory distress syndrome. *Egyptian journal of chest diseases and tuberculosis* 2015;64(4):843–848.
- [19]. Wei B, Dai T, Peng L, Guo Y, Manatunga AK. An Alternative Representation of Broad Sense Agreement for Complete Data. Technical Report 2018;18–01.
- [20]. Taylor JM, Hsu CH, Murray S. Survival estimation and testing via multiple imputation. *Statistics & Probability Letters* 2002;58(3):221–232.
- [21]. Hsu CH, Taylor JM, Murray S, Commenges D. Survival analysis using auxiliary variables via non-parametric multiple imputation. *Statistics In Medicine* 2006;25(20):3503–3517. [PubMed: 16345047]
- [22]. Rubin DB, Schenker N. Multiple imputations in health-care database: an overview and some applications. *Statistics in Medicine* 1991;10(4):585–598. [PubMed: 2057657]
- [23]. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Critical Care Medicine* 1985;13(10):818–829. [PubMed: 3928249]
- [24]. Wisplinghoff H, Seifert H, Wenzel RP, Edmond MB. Inflammatory response and clinical course of adult patients with nosocomial bloodstream infections caused by *Candida* spp. *Clinical Microbiology and Infection* 2006;12(2):170–177. [PubMed: 16441456]
- [25]. Naved SA, Siddiqui S, Khan FH. APACHE-II Score Correlation With Mortality And Length Of Stay In An Intensive Care Unit. *Journal of the College of Physicians and Surgeons–Pakistan: JCPSP* 2011;21(1):4–8. [PubMed: 21276376]
- [26]. Greenwood B, Szumita PM, Levy H, Lilly CM. Error rates among clinical pharmacists in calculating the APACHE II score. *Pharmacotherapy* 2007;27(2):285–289. [PubMed: 17253918]

TABLE 1

Simulation results for estimating moderate broad sense agreement ($\rho_{bsa} = 0.550$) when $C \sim \text{Weibull}+ Y$: empirical biases (EmpBias), empirical standard deviation (EmpSD), estimated standard deviation (EstSD), and coverage probabilities of 95% confidence intervals (Cov95).

Equal censoring proportions across Y: C ~ Weibull+ Y									
Censoring	Methods	N=60				N=120			
		EmpBias	EmpSD	EstSD	Cov95	EmpBias	EmpSD	EstSD	Cov95
Low (27%,27%,27%)	Comp obs. ^a	0.054	0.115	0.115	89.8	0.061	0.077	0.880	88.0
	All obs. ^b	0.060	0.097	0.095	87.8	0.059	0.066	0.067	86.2
	Plug-in ^c	0.050	0.115	0.114	90.8	0.029	0.078	0.081	93.4
	MI ^d	-0.020	0.131	0.114	91.0	-0.005	0.083	0.080	94.0
	BMI ^e	0.022	0.132	0.120	93.0	-0.006	0.083	0.083	96.0
Moderate (50%,50%,50%)	Comp obs.	0.149	0.115	0.117	75.0	0.149	0.074	0.080	65.6
	All obs.	0.143	0.080	0.081	67.4	0.146	0.052	0.056	37.8
	Plug-in	0.074	0.115	0.144	94.0	0.047	0.092	0.105	93.2
	MI	-0.020	0.144	0.119	90.0	-0.006	0.110	0.083	87.2
	BMI	0.028	0.144	0.141	93.8	-0.017	0.109	0.100	91.9
Heavy (70%,70%,70%)	Comp obs.	0.232	0.125	0.129	72.3	0.246	0.075	0.078	35.8
	All obs.	0.239	0.060	0.061	21.0	0.243	0.041	0.042	1.0
	Plug-in	0.101	0.143	0.161	91.6	0.075	0.117	0.139	92.6
	MI	-0.010	0.207	0.116	75.8	-0.032	0.157	0.089	72.5
	BMI	-0.014	0.193	0.159	88.8	-0.049	0.155	0.137	90.6
Unequal censoring proportions across Y: C ~ Weibull+0.5Y									
Censoring	Methods	N=60				N=120			
		EmpBias	EmpSD	EstSD	Cov95	EmpBias	EmpSD	EstSD	Cov95
Low (25%, 32%, 38%)	Comp obs.	0.035	0.125	0.123	93.2	0.034	0.080	0.083	94.4
	All obs.	-0.049	0.114	0.113	91.8	-0.047	0.076	0.078	92.2
	Plug-in	0.060	0.106	0.114	92.6	0.031	0.076	0.082	94.2
	MI	-0.008	0.120	0.116	94.2	-0.001	0.086	0.080	91.9
	BMI	-0.007	0.118	0.120	95.6	-0.003	0.086	0.083	94.0
Moderate (33%, 41%, 49%)	Comp obs.	0.065	0.116	0.125	91.4	0.053	0.083	0.087	91.6
	All obs.	-0.052	0.108	0.113	93.2	-0.056	0.077	0.079	91.4
	Plug-in	0.075	0.107	0.120	90.8	0.039	0.085	0.086	90.6
	MI	0.002	0.124	0.115	92.8	-0.009	0.084	0.084	94.6
	BMI	-0.003	0.124	0.126	96.0	-0.011	0.087	0.091	96.6
Heavy (60%, 70%, 78%)	Comp obs.	0.138	0.155	0.176	87.9	0.134	0.100	0.107	78.8
	All obs.	-0.060	0.109	0.114	93.4	-0.061	0.073	0.078	89.4
	Plug-in	0.121	0.134	0.158	92.2	0.077	0.107	0.122	92.6
	MI	-0.006	0.201	0.123	79.4	-0.022	0.158	0.090	77.2
	BMI	-0.017	0.200	0.166	89.8	-0.030	0.151	0.130	87.2

^aNaive estimator using datasets which exclude censored subjects.

^bNaive estimator using censored observations as observed observations.

^cThe proposed plug-in estimator.

^dKaplan-Meier-based imputation without bootstrap procedure.

^eKaplan-Meier-based imputation with bootstrap procedure.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 2

Simulation results for estimating strong broad sense agreement ($\rho_{bsa} = 0.827$) when $C \sim \text{Weibull+}$ Y : empirical biases (EmpBias), empirical standard deviation (EmpSD), estimated standard deviation (EstSD), and coverage probabilities of 95% confidence intervals (Cov95).

Equal censoring proportions across Y: C ~ Weibull+ Y									
Censoring	Methods	N=60				N=120			
		EmpBias	EmpSD	EstSD	Cov95	EmpBias	EmpSD	EstSD	Cov95
Low (27%,27%,27%)	Comp obs. ^a	0.047	0.048	0.049	87.2	0.045	0.036	0.035	77.8
	All obs. ^b	0.047	0.040	0.041	86.4	0.046	0.030	0.029	69.4
	Plug-in ^c	0.018	0.067	0.076	94.4	0.011	0.050	0.052	92.4
	MI ^d	-0.024	0.077	0.066	92.2	-0.018	0.055	0.047	90.4
	BMI ^e	-0.027	0.074	0.077	95.4	-0.015	0.052	0.054	92.6
Moderate (50%,50%,50%)	Comp obs.	0.099	0.046	0.043	76.2	0.098	0.031	0.030	37.4
	All obs.	0.099	0.031	0.029	33.4	0.098	0.021	0.021	8.2
	Plug-in	-0.020	0.120	0.130	87.6	-0.009	0.094	0.100	89.2
	MI	-0.067	0.136	0.077	73.4	-0.056	0.100	0.055	71.3
	BMI	-0.081	0.131	0.110	84.0	-0.067	0.098	0.087	81.9
Heavy (69%,69%,69%)	Comp obs.	0.137	0.042	0.034	90.1	0.139	0.026	0.024	31.0
	All obs.	0.137	0.019	0.019	4.4	0.139	0.013	0.013	0.0
	Plug-in	0.080	0.141	0.158	80.6	-0.074	0.136	0.142	74.6
	MI	-0.099	0.154	0.078	59.1	-0.111	0.129	0.061	44.1
	BMI	-0.114	0.143	0.120	74.0	-0.120	0.117	0.102	62.4
Unequal censoring proportions across Y: C ~ Weibull+0.5Y									
Censoring	Methods	N=60				N=120			
		EmpBias	EmpSD	EstSD	Cov95	EmpBias	EmpSD	EstSD	Cov95
Low (20%, 30%, 39%)	Comp obs.	0.021	0.056	0.058	94.2	0.021	0.041	0.039	92.0
	All obs.	-0.086	0.079	0.076	77.0	-0.085	0.051	0.052	54.2
	Plug-in	0.031	0.053	0.060	93.8	0.022	0.040	0.041	90.4
	MI	-0.010	0.066	0.061	93.8	-0.004	0.050	0.042	90.3
	BMI	-0.012	0.066	0.067	95.6	-0.005	0.050	0.046	88.2
Moderate (33%, 44%, 56%)	Comp obs.	0.034	0.059	0.063	96.0	0.035	0.040	0.042	87.8
	All obs.	-0.118	0.078	0.081	57.8	-0.122	0.056	0.057	27.0
	Plug-in	0.023	0.073	0.081	94.8	0.020	0.047	0.052	96.0
	MI	-0.026	0.092	0.068	86.4	-0.019	0.068	0.049	87.2
	BMI	-0.032	0.092	0.084	91.0	-0.027	0.072	0.063	91.5
Heavy (54%, 70%, 83%)	Comp obs.	0.069	0.077	0.083	97.6	0.072	0.050	0.048	80.2
	All obs.	-0.155	0.085	0.084	37.6	-0.163	0.056	0.059	6.4
	Plug-in	0.019	0.104	0.124	94.7	0.015	0.079	0.088	92.2
	MI	-0.047	0.144	0.074	75.3	-0.038	0.094	0.054	68.1
	BMI	-0.069	0.144	0.124	87.8	-0.046	0.090	0.090	91.5

^aNaive estimator using datasets which exclude censored subjects.

^bNaive estimator using censored observations as observed observations.

^cThe proposed plug-in estimator.

^dKaplan-Meier-based imputation without bootstrap procedure.

^eKaplan-Meier-based imputation with bootstrap procedure.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 3

Comparing computation times (in seconds) for the estimation of BSA index and jackknife standard error in one simulation using different estimation methods.

N	Peng et al. [1]'s method	Plug-in method	MI method	BMI method
60	12	0.9	117	118
120	247	2	2480	2486
240	4077	5	40593	40611

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 4

Surgical ICU patient study: estimates of ρ_{bsa} by different methods and associated standard errors (SE) and 95% confidence intervals (95% CI).

Method	$\hat{\rho}_{bsa}$	SE	95%CI
Comp obs. ^a	-0.023	0.103	(-0.221, 0.177)
All obs. ^b	0.053	0.153	(-0.243, 0.340)
Plug-in ^c	0.298	0.122	(0.045, 0.515)
MI ^d	0.276	0.109	(0.052, 0.474)
BMI ^e	0.269	0.120	(0.022, 0.485)

^aNaive estimator using datasets which exclude censored subjects.

^bNaive estimator using censored observations as observed observations.

^cThe proposed plug-in estimator.

^dKaplan-Meier-based imputation without bootstrap procedure.

^eKaplan-Meier-based imputation with bootstrap procedure.