



Published in final edited form as:

*J Biopharm Stat.* 2012 ; 22(2): 401–415. doi:10.1080/10543406.2010.544437.

## Average Cost-Effectiveness Ratio with Censored Data

Heejung Bang, Ph.D.\* and

Division of Biostatistics and Epidemiology, Department of Public Health, Weill Medical College of Cornell University, New York, New York, USA

Hongwei Zhao, Sc.D.

Department of Epidemiology and Biostatistics, School of Rural Public Health, Texas A&M Health Science Center, College Station, Texas, USA

### SUMMARY

In the cost-effectiveness analysis, interest could lie foremost in the incremental cost-effectiveness ratio (ICER) which is the ratio of the incremental cost to the incremental benefit of two competing interventions. The average cost-effectiveness ratio (ACER) is the ratio of the cost to benefit of an intervention without reference to a comparator. A vast literature is available for statistical inference of the ICERs, but limited methods have been developed for the ACERs, particularly in the presence of censoring. Censoring is a common feature in prospective studies and valid analyses should properly adjust for censoring in cost as well as in effectiveness. In this paper, we propose statistical methods for constructing a confidence interval for the ACER from censored data. Different methods - Fieller, Taylor, Bootstrap - are proposed and through simulation studies and data analysis, we address the performance characteristics of these methods.

### Keywords

Bootstrap; Confidence Interval; Fieller; Taylor

### 1 Introduction

Economic burden in health care has been a significant concern to virtually all members of our society. In the presence of competing methods (e.g., interventions and regimens) doing the same task, evaluating costs, relative to their perceived benefits, is critical. The strong need to control health care costs makes us search for intervention(s) that produce the greatest value, based partly on comparative economic evaluations.

Cost-effectiveness analysis (CEA) is a form of economic analysis that compares the relative expenditures and outcomes of two or more competing strategies. It has been frequently used in economic, biomedical, social sciences, public health and other fields over the past several decades. Various statistical methodologies have been developed for CEA and the incremental cost-effectiveness ratio (ICER), defined as the ratio of the difference in costs to the difference in effectiveness between two competing strategies, has been most widely accepted by researchers and policy makers [1], [2], [3] [4]. In addition to the ICER, the ACER that estimates average cost spent per effect has served as another important measure in the CEA [5], [6]. [Note: We can formulate the  $ACER = \mu^M / \mu^T$  and the

$ICER = (\mu_1^M - \mu_2^M) / (\mu_1^T - \mu_2^T)$ , where  $\mu^M$  and  $\mu^T$  denote mean cost and mean effectiveness,

\*Correspondence to: Heejung Bang, Department of Public Health, Weill Medical College of Cornell University, 402 East 67th Street, New York, NY 10065, U.S.A. Tel.: +1-646-962-8015; Fax: +1-646-962-0281. heb2013@med.cornell.edu.

respectively, and the subscript indexes comparison groups. The difference of the ACERs (as in two-sample problem),  $\mu_1^M/\mu_1^T - \mu_2^M/\mu_2^T$ , may be contrasted with the ICER.]

Apparently, the ICER and ACER estimate different parameters and the purposes of these measures are different. Some researchers reviewed and compared the roles of these two approaches [5], [6], [7], [8], [9]. Particularly, Laska et al. [10], Gardiner et al. [11], [12] and Wagner [13] studied the mathematical properties of the ACER and elucidated the relationships between the ACER and the ICER. Using ACERs, one may devise a decision rule based on a fixed budget, to maximize total effectiveness, similar to the idea of the Neyman-Pearson lemma [14].

Although, the ICER could be more relevant to health economics and policy decision, the ACER has several advantages/properties that should be noted: 1) it is a parameter that characterizes clinical and economical properties of a treatment independent of its comparators (thus, application to one group or more than 2 groups is straightforward); 2) it conveys an intuitive meaning and interpretation (say, cost spent per year) that even lay persons can understand - it is very likely that researchers, policy makers and patients/payers may want to see the ACERs (e.g., for short vs. long-term costs) even when the ICER is adopted for decision making; 3) it is less vulnerable to numerical instability, compared to the ICER - For example, the ICERs can be misleading and/or numerically unstable when two groups demonstrate similar effectiveness (e.g., resulting in a small denominator in the ICER); and 4) a subjective threshold (such as  $\lambda$ , which is typically needed for ICERs) is generally not needed [15], [16].

Currently, there are various methods available for statistical inference of the ICERs. In contrast, less attention has been paid to the ACERs. Also, to our knowledge, there is no statistical method published so far for the analysis of the ACERs with censored data. Censoring occurs commonly in prospective studies that entail patient follow-up (e.g., clinical trials). When it occurs, it should be properly accounted in the analyses of effectiveness, cost, and cost-effectiveness [17], [18], [19]. Also, when the effectiveness between two treatment groups are differential (e.g., one group survives longer than the other), it may be misleading to compare two mean costs (or other functions of the costs or their distributions) because one treatment group may incur larger cost primarily because patients in that group tend to survive longer (i.e., cost is confounded with survival time). In that case, one way to adjust survival experience could be to compute the ACERs and to compare the ACERs rather than costs between/among treatment groups directly. Even when the effectiveness between two treatment groups are not differential (so that estimation of the ICER may not be well justified or stable), estimating the ACERs and the associated confidence intervals (CI) would be still interesting, say, in order to understand how much money is expected to be spent per year of survival. Although the estimates of the ACER and ICER are similar mathematically as both being ratio statistics, there are some differences - the ACER is the difference of the ratios and the ICER is the ratio of the differences. Therefore, currently available methodological guidance and comparative reviews suited for the ICERs may not be directly applicable to the ACERs ([20], [21], [22], [23], [24], [25], among others).

In this article, statistical methods are proposed for the estimation and inference of the ACER in the presence of censoring (called 'one-sample problem'). Then the methods are extended to the comparison of the two ACERs ('two-sample problem'). The methods are outlined in Section 2. Section 3 presents simulation studies and Section 4 presents the analysis of the data collected in a cardiovascular clinical trial. Concluding remarks are provided in Section 5.

## 2 Statistical Methods

### 2.1 Notation and assumptions

Let us consider one group of samples (e.g., one arm in the trial) first and assume death is the endpoint of interest without loss of generality. For the  $i$ th person in the study, let  $T_i$  denote his/her overall survival time and  $C_i$  the censoring time, where censoring is assumed to be random and independent of survival time and total cost. This kind of ‘non-informative censoring’ is assumed for most of survival data analysis methods. For example, this assumption is satisfied when censoring is mainly caused by administrative reasons such as finite duration of a clinical trial. Due to censoring, only one value between  $T_i$  and  $C_i$  is observed, whichever occurs first. Thus, the observed data for survival analysis are the follow-up time,  $X_i = \min(T_i, C_i)$ , along with the corresponding censoring indicator,  $\Delta_i = I(T_i \leq C_i)$ , where  $\Delta_i = 1$  means the  $i$ th person’s death is observed and  $\Delta_i = 0$  means the person’s death is not observed so that his/her survival time is censored. Denote the cost history,  $M(t)$ , as the cost accumulates from time 0 (the point when the patient entered the study) to time  $t$ . Because of censoring, it is impossible to estimate the lifetime cost without making some restrictive assumptions [26], [27]. Therefore, we only consider cost as well as survival time accumulated up to a fixed time point  $L$ , where a reasonable amount of complete (i.e., uncensored) data are available over the time period  $[0, L]$ . Hence, we will consider  $T_i^L = \min(T_i, L)$  instead of  $T_i$ . But for simpler notation, we will suppress the superscript  $L$  and use  $T_i$  to denote  $T_i^L$  in the remaining manuscript. In practice,  $L$  should be chosen to make the largest survival time be uncensored [28]. However, we may choose  $L$  to define several largest survival times as complete observations in order to prevent numerical instability near the tail of the distribution [29]. [Remark: Depending on the goal and data availability, CEA may be conducted with different time frames (e.g., lifetime horizon, within trial, or upto 3 years). In these tasks, extrapolation or time restriction is often entailed.]

We are interested in estimating the expected value or mean of the medical cost  $\mu = E\{M_i(T_i)\}$  from a set of observed data  $[X_i, \Delta_i, \{M_i(t), t \leq X_i\}, i = 1, \dots, n]$ . The cost history is often measured in discrete time intervals (e.g., monthly or yearly) in practice. If the cost history is not recorded, we only observe the final (total) cost  $M_i \equiv M_i(X_i)$  for each individual. Those who died before being censored have  $M_i \equiv M_i(X_i) = M_i(T_i)$ .

### 2.2 Estimating the Mean Cost and its Variance

If all patients are followed up to time  $L$  or until their deaths (i.e., no censoring), then we would have complete survival time and total cost data for every patient so that the standard statistical method such as the sample mean could be used for estimating the mean cost and mean survival time. However, in most cases in prospective studies, some costs as well as survival times are censored, i.e., they are not completely observed for all patients. Bang and Tsiatis [30] proposed a simple weighted estimator for the mean cost using the idea of inverse-probability weighting as follows:

$$\hat{\mu}_{wt}^M = \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i M_i}{\hat{K}(T_i)}$$

where  $\hat{K}(T_i)$  is an estimator for  $K(t) = P(C > t)$ , the survival function of the censoring time  $C$  evaluated at time  $T_i$ . It can be estimated by the Kaplan-Meier estimator with the roles of  $T_i$  and  $C_i$  reversed (i.e., censoring indicator becomes  $1 - \Delta_i$  instead of  $\Delta_i$ ) [31]. In this simple weighted estimator, censoring is handled by weighting each uncensored individual with their probability of being uncensored so that each uncensored individual represents more than one

observation in order to account for censored observations. This inverse-probability weighting idea was originated by Horvitz and Thompson [32] in survey sampling and the simple weighted estimator has served as a building block upon which more advanced or sophisticated medical cost estimators have been developed. The simple weighted estimator was shown to be consistent and asymptotically normal and its variance can be consistently estimated by

$$\widehat{\text{Var}}(\widehat{\mu}_{wt}^M) = \frac{1}{n^2} \left[ \sum_{i=1}^n \frac{\Delta_i (M_i - \widehat{\mu}_{wt}^M)^2}{\widehat{K}(T_i)} + \sum_{j=1}^n \frac{(1 - \Delta_j)}{\widehat{K}^2(C_j)} \{G_j(M^2) - G_j^2(M)\} \right]$$

where

$$G_j(M) = \frac{1}{n\widehat{S}(X_j)} \sum_{i=1}^n \frac{\Delta_i M_i(X_i) I(X_i \geq X_j)}{\widehat{K}(X_i)}.$$

The simple weighted estimator only utilizes completely observed data, thus it is not efficient, especially when the proportion of censoring is high. A general way to improve the efficiency of an estimator could be by extracting and utilizing information from censored observations or from the cost history of available observations. For this purpose, a more efficient but practical estimator was proposed by Zhao and Tian [33] based on semi-parametric efficiency theories ([34], [30]). [A reason that we call ‘practical’ here is that there are several efficient estimators developed in this framework but most estimators are computationally and algorithmically complicated so that the use of these estimators is less attractive to practitioners and data analysts.] The Zhao and Tian estimator can be written as

$$\widehat{\mu}_{IMP}^M = \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i M_i}{\widehat{K}(T_i)} + \frac{1}{n} \sum_{j=1}^n \frac{(1 - \Delta_j)}{\widehat{K}(C_j)} \{M_j - \overline{M}(C_j)\}$$

where

$$\overline{M}(C_j) = \frac{\sum_{i=1}^n M_i(C_j) I(X_i \geq C_j)}{\sum_{i=1}^n I(X_i \geq C_j)}$$

is the average cost at  $C_j$  of individuals who are still under observation at time  $C_j$ . The formulation of the improved estimator clearly shows the contributions from uncensored data and those from censored data.

A variance estimator for the improved estimator is given by

$$\begin{aligned} \widehat{Var}(\widehat{\mu}_{IMP}^M) &= \widehat{Var}(\widehat{\mu}_{WT}^M) \\ &- \frac{2}{n^2} \sum_{j=1}^n \frac{(1-\Delta_j)}{Y(C_j)\widehat{K}(C_j)} \sum_{i=1}^n \frac{\Delta_i I(X_i \geq X_j)}{\widehat{K}(T_i)} \{M_i - G_j(M)\} \{M_i(X_j) - \overline{M}(X_j)\} \\ &+ \frac{1}{n^2} \sum_{j=1}^n \frac{(1-\Delta_j)}{Y(C_j)\widehat{K}(C_j)^2} \sum_{i=1}^n I(X_i \geq X_j) \{M_i(X_j) - \overline{M}(X_j)\}^2 \end{aligned}$$

This improved estimator was shown to be more efficient than the simple weighted estimator whenever  $M_i(u)$  is highly correlated with the overall cost  $M_i$ , which is often the case in practice [33]. See [35], [36] for analytic relationships among simple weighted, improved and/or other estimators.

### 2.3 Estimating the Covariance of the Mean Cost and the Mean Survival Time

In order to construct a CI for the ACER (as a ratio statistic,  $\widehat{\mu}^M/\widehat{\mu}^T$ ) using parametric methods outlined in the next subsection, we need to derive covariance between the mean cost estimator ( $\widehat{\mu}^M$ ) and the mean survival time estimator ( $\widehat{\mu}^T$ ). Here, the mean survival time can be estimated by the area under the Kaplan-Meier curve of the survival function over  $[0, L]$ ,

or equivalently by  $\widehat{\mu}_{WT}^T = \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i T_i}{\widehat{K}(T_i)}$ , an alternative formula via inverse-probability weighting [28], [37]. [Remark: ACER will denote parameter and estimator interchangeably as long as the context is clear for simpler presentation.]

Using the logics of Zhao and Tian [33] based on the martingale central limit theorem, the covariance of the simple weighted estimator of the mean cost and the mean survival time can be derived as

$$\widehat{Cov}(\widehat{\mu}_{WT}^M, \widehat{\mu}_{WT}^T) = \frac{1}{n^2} \sum_{i=1}^n \frac{\Delta_i M_i T_i}{\widehat{K}(T_i)} - \frac{1}{n^3} \sum_{i=1}^n \frac{\Delta_i M_i}{\widehat{K}(T_i)} \sum_{i=1}^n \frac{\Delta_i T_i}{\widehat{K}(T_i)} + \frac{1}{n^2} \sum_{j=1}^n \frac{(1-\Delta_j)}{\widehat{K}^2(C_j)} \{G_j(MT) - G_j(M)G_j(T)\}$$

where

$$G_j(MT) = \frac{1}{n\widehat{S}(X_j)} \sum_{i=1}^n \frac{\Delta_i M_i(X_i) T_i I(X_i \geq X_j)}{\widehat{K}(X_i)}$$

and

$$G_j(T) = \frac{1}{n\widehat{S}(X_j)} \sum_{i=1}^n \frac{\Delta_i T_i I(X_i \geq X_j)}{\widehat{K}(X_i)}.$$

The covariance of the improved estimator and the mean survival time can be derived as:

$$\widehat{Cov}(\widehat{\mu}_{IMP}^M, \widehat{\mu}_{WT}^T) = \widehat{Cov}(\widehat{\mu}_{WT}^M, \widehat{\mu}_{WT}^T) - \frac{2}{n^2} \sum_{j=1}^n \frac{(1-\Delta_j)}{Y(C_j)\widehat{K}(C_j)} \sum_{i=1}^n \frac{\Delta_i I(X_i \geq X_j)}{\widehat{K}(T_i)} \{T_i - G_j(T)\} \{M_i(X_j) - \overline{M}(X_j)\}.$$

## 2.4 Estimating the ACER and its Confidence Interval: One-Sample Problem

In this subsection, we continue to consider one-sample problem. This implies, in two (or multi) arm trial, we will compute the ACER for each arm separately.

We consider four different approaches for constructing CIs: Taylor's, Fieller's, percentiles-based Bootstrap, and normality-based Bootstrap methods. These methods along with others have been compared for the ICERs (as such, the same formulae provided below appeared elsewhere including [22], [38] among others) and a general consensus is that Fieller's method and Bootstrap percentile method are recommended for obtaining CIs for the ICER in common scenarios). [Remark: Yet, great caution may need to be exercised when we choose a CI method for the ICER because qualitatively different scenarios are possible such as  $a/b$ ,  $a/0$ ,  $0/0$ , where  $a$  and  $b$  are positive constants and 'identically 0' is highly unlikely in practice - See [39] and [40] on these issues. In contrast, the same caution is not generally needed for the ACER.] In this manuscript, we want to evaluate the performance of these methods for the ACER counterpart. Each method is summarized below. We assume that we have large samples so that asymptotic theory is satisfied. However, we will evaluate the methods with finite samples ( $N = 100$  and  $200$ ) using simulation in the next section.

**Taylor's method**—The variance of the ACER can be approximated via the standard Taylor's series expansion assuming a normal distribution based on the central limit theorem. The variance can be estimated by:

$$\widehat{Var}(\widehat{ACER}) = \widehat{Var}\left(\frac{\widehat{\mu}^M}{\widehat{\mu}^T}\right) \approx \left(\frac{\widehat{\mu}^M}{\widehat{\mu}^T}\right)^2 \left\{ \frac{\widehat{Var}(\widehat{\mu}^M)}{(\widehat{\mu}^M)^2} + \frac{\widehat{Var}(\widehat{\mu}^T)}{(\widehat{\mu}^T)^2} - \frac{2\widehat{Cov}(\widehat{\mu}^M, \widehat{\mu}^T)}{\widehat{\mu}^M \widehat{\mu}^T} \right\}$$

Then the  $(1 - \alpha)\%$  CI for the ACER parameter can be constructed as

$$\left( \widehat{ACER} - z_{\alpha/2} \sqrt{\widehat{Var}(\widehat{ACER})}, \widehat{ACER} + z_{\alpha/2} \sqrt{\widehat{Var}(\widehat{ACER})} \right)$$

where  $z_{\alpha/2}$  is the critical value with right tail area  $\alpha/2$  from the standard normal distribution.

**Fieller's method**—In order to circumvent the derivation of the exact variance of a ratio, Fieller [42] proposed a method that transforms the ratio into a linear function of the two random variables and then derives the variance of the linear function. Since the numerator  $x = \widehat{\mu}^M$  and the denominator  $y = \widehat{\mu}^T$  in the ACER are bivariate-normal asymptotically, the assumption needed for Fieller's theorem may well be satisfied. Hence the  $100(1 - \alpha)\%$  CI for the ACER can be given as

$$\frac{xy - z_{\alpha/2}^2 S_{xy} \pm \{f(x, y, S_{xx}, S_{xy}, S_{yy})\}^{1/2}}{y^2 - z_{\alpha/2}^2 S_{yy}}$$

where  $f(x, y, S_{xx}, S_{xy}, S_{yy}) = (xy - z_{\alpha/2}^2 S_{xy})^2 - (x^2 - z_{\alpha/2}^2 S_{xx})(y^2 - z_{\alpha/2}^2 S_{yy})$ , and  $S_{xx}$ ,  $S_{yy}$  and  $S_{xy}$  are the sample variances of  $x$ , the sample variance of  $y$ , and the sample covariance of  $x$  and  $y$ , respectively.

In addition, Bootstrap methods are natural choices for nonparametric estimation of the variance or CI. By generating  $B$  bootstrap replicates of effectiveness-cost pairs by ‘sampling with replacement’, we can obtain  $B$  ACER values, denoted by  $(\widehat{ACER}_1^*, \dots, \widehat{ACER}_B^*)$ , which could form an approximate distribution of  $\widehat{ACER}$ , where we want to preserve within-person correlation. Among various bootstrap methods, we intended to implement two most convenient ones [41].

**Bootstrap percentile method**—From  $(\widehat{ACER}_1^*, \dots, \widehat{ACER}_B^*)$ , a  $(1 - \alpha)\%$  CI for the ACER can be derived by using the  $(\alpha/2)$ th and  $(1 - \alpha/2)$ th percentiles, which are denoted by  $(\widehat{ACER}_{(\alpha/2)}^*, \widehat{ACER}_{(1-\alpha/2)}^*)$ . For example, If  $B = 1,000$  and the ACER values are sorted, then the 25th and 975th values can form a 95% CI.

**Bootstrap normal method**—The  $B$  ACER bootstrap resamples can also be used to estimate the variance of  $\widehat{ACER}$ , denoted by  $\widehat{Var}^*(\widehat{ACER})$ . Then the  $(1 - \alpha)\%$  CI is constructed as

$$\left( \widehat{ACER} - z_{\alpha/2} \sqrt{\widehat{Var}^*(\widehat{ACER})}, \widehat{ACER} + z_{\alpha/2} \sqrt{\widehat{Var}^*(\widehat{ACER})} \right).$$

## 2.5 Estimating the Difference of the ACERs and its Confidence Interval: Two-Sample Problem

After the effectiveness of two (or more) treatments are compared, we are often interested in comparing the associated costs as well. In some cases, it may be more legitimate to compare the ACERs rather than (or in addition to) the costs between two arms since the survival time tends to be positively associated with the cost (i.e., the total costs tend to be higher for longer survivors). Therefore, if controlling survival experience in the cost comparison is the goal, testing the equality of the ACERs between treatments (i.e.,  $H_0: ACER_1 = ACER_2$  where the subscript denotes the treatment indicator) would be more relevant than testing the equality of the mean costs. For this task, the two methods we suggested above for the one-sample problem (Taylor’s method and Bootstrap normal) can be straightforwardly extended to the two-sample problem as they estimate variance/standard error, whereas Bootstrap percentile method and Fieller’s method are not because they estimate CIs for the ratio statistic directly.

For the Bootstrap percentile method, we can sample ACER from each group (group 1 and group 2) separately and compute the difference in the ACERs (let us call  $\widehat{diff}$ ). Then we can generate  $B$  bootstrap samples of  $(\widehat{diff}_1^*, \dots, \widehat{diff}_B^*)$  and form a CI.

Fieller’s method cannot be used directly for obtaining a CI for the difference of the ACERs. However, in the illustrative example, we consider an ad-hoc way of obtaining a CI for the difference of ACERs, using standard error estimates obtained from the length of CI derived from the Fieller’s method.

## 3 Simulation Study

We conducted a simulation study to examine the performance of the presented methods for constructing CIs with finite samples. We adopted simulation settings that were similar to the ones used in previous papers [17], [30], [33]. Data for individual subjects were

independently and identically distributed. The total cost for each subject consisted of 3 cost components: the diagnostic cost that incurs at the beginning of the study; the annual cost that incur annually during the follow-up; and the terminal death cost that incurred shortly before or at death. As such, diagnostic and annual costs are relevant for all patients, while death cost is only entailed for those who died. We implemented the following two scenarios for cost data. For the first scenario, we generated:

$$\begin{aligned}\text{diagnostic cost} &\sim \exp(N(0, 1) * 1+5)+1000 \\ \text{annual cost} &\sim U[0, 1] * 1000+1000 \\ \text{terminal death cost} &\sim \exp(N(0, 1) * 1.5+6)+1000\end{aligned}$$

and for the second scenario, we assumed:

$$\begin{aligned}\text{diagnostic cost} &\sim \exp(N(0, 1) * 1.5+5)+1000 \\ \text{annual cost} &\sim U[0, 1] * 3000+1000 \\ \text{terminal death cost} &\sim \exp(N(0, 1) * 1.8+6)+1000\end{aligned}$$

where  $N$  and  $U$  denote the normal and uniform distributions and  $\exp$  denotes exponentiation. The first scenario intends to exemplify moderate skewness and variability, while the second scenario intends to exemplify higher skewness and variability.

For each scenario, we considered two types of distribution for generating survival times: a uniform distribution on  $U[0, 10]$  years and an exponential distribution with a mean of 6 years. Our interest is to estimate the mean cost accumulated up to death or 10 years whichever comes first. In the scenario 1, the true mean cost was \$10,990 for the uniform survival time and \$10,786 for the exponential survival time, where the skewness parameter was estimated to be 1.7 and 1.3, respectively. The corresponding true ACERs were \$2,198/yr and \$2,215/yr, respectively. In the scenario 2, the corresponding skewness parameters were 3.3 and 2.7 and the corresponding true ACERs were \$3,400/yr and \$3,426/yr - here, the true ACERs were numerically estimated using 1, 000, 000 uncensored data.

We considered two levels of censorship:  $C$  was generated from  $U[0, 20]$  and  $U[0, 12.5]$  years, independent of all other variables. The former was referred to as light censoring, resulting in about 25% censored data, and the latter was referred to as heavy censoring, resulting in about 40% censored data. Note that by definition, if the follow-up time (the minimum of the survival time and the censoring time) exceeds 10 years, it is equivalent to the uncensored, complete event. One thousand simulation runs were carried out and 1, 000 bootstrap samples were generated for the bootstrap methods. The same set of simulations was repeated for the sample size of 100 and 200.

We calculated a naive estimator (using observed costs without accounting for censoring in cost), the simple weighted estimator, and the improved estimator. Tables 1 and 2 present the (absolute) bias, the coverage probability of the 95% CIs, and the median length of the 95% CIs in each scenario. The simple weighted estimator and the improved estimator had small biases, whereas the naive estimator severely underestimates the true cost as expected. Enhanced efficiency in the improved estimator was demonstrated in the shorter length of the CIs as the theory predicts. All four methods of constructing CIs performed quite well and comparably without any consistent, noticeable inferiority of any method across different data distributions and sizes in the scenario 1. However, when the skewness and variability were increased for individual cost data (in the scenario 2), the different estimators tended to perform somewhat differently while coverage probabilities were deteriorated. Fieller's and Bootstrap-percentile methods seem to perform best overall. It may be noteworthy that



normality-based parametric methods could perform poorly under heavy censoring with small sample size (e.g., the coverage probability can be as low as 87%). Under heavy censoring, the improved estimator may be recommended because it utilizes some observations that are not used in the simple weighted estimator. The efficiency gain was more pronounced in the scenario 2, in which the observed cost at a given time is more strongly correlated with the total cost (as contributions from annual costs is increased), compared to the scenario 1. It is quite interesting to observe that Taylor's method does not perform terribly - this method is not generally recommended in the ICER context ([20], [22], [43] among others). The absence of the 'small denominator problem' (which is frequent in the ICER estimation) in the ACER estimation may have partly contributed to comparable performances of these methods. Also, asymptotic normality may be better achieved in the ACER setting (a ratio of means) than in the ICER (a ratio of differences in means) or cost setting. Taylor's method yielded slightly shorter CIs than Fieller's method, probably due to imposing the symmetry of the CIs. The coverage probability was improved as the sample size increased and/or the censoring proportion decreased within each scenario.

#### 4 Example

To illustrate the methods discussed in this paper, we analyzed the data collected from the Multicenter Automatic Defibrillator Implantation Trial (MADIT). MADIT was a randomized controlled trial that examined the effectiveness of an implantable cardiac defibrillator (ICD) in prevention of sudden death for patients who were at high risk for ventricular arrhythmia [44]. A total of 181 patients were enrolled from 36 centers with 89 patients assigned to the ICD arm, and 92 patients assigned to the control arm with the conventional treatment. The first enrolled patient was followed for 61 months and the last for less than 1 month, with an average follow-up of 27 months. After completion of the study, it has been shown that the use of ICD as prophylactic therapy yielded improved survival, compared to the conventional treatment ( $p = 0.009$ , [44]). Because of the high initial cost associated with the ICDs, cost analyses were warranted and cost data were collected for patients recruited from centers in the US. All the relevant medical costs incurred during the study were recorded, as described in Mushlin et al. [45].

The CEA for this trial was performed using the standard ICER and published previously [45]. In this paper, we analyzed the cost and survival data using the ACER. As in the original CEA, we also restricted the duration of the cost estimation to 4 years (denoted by  $L$  previously). The data were heavily censored; for example, 70% of subjects were censored in the ICD arm and 48% of subjects were censored in the conventional arm. As customarily done in the CEA, both costs and survival times were discounted at 3% annual rate in our analysis [2].

The simple weighted and improved estimators were implemented for the estimation of the mean costs, ACERs and their differences. Data analyses are reported in Table 3. The mean survival time is 2.65 years for the conventional arm and 3.45 years for the ICD arm, which yielded the difference of 0.80 year with 95% CI of (0.43 – 1.17). The mean cost for the ICD was estimated as \$110,109 using the simple weighted estimator and \$99,548 using the improved estimator. The corresponding mean costs for the conventional arm were \$70,044 and \$72,754, respectively. The difference of the mean costs was \$40,065 (95% CI: 17,383–62,747) and \$26,794 (6,901–46,687) by the simple weighted estimator and the improved estimator, respectively, which implies ICD significantly costs more than the conventional treatment. However, when we divided the mean cost by the mean survival time to compute the ACER, we obtained \$5,475/yr by the simple weighted estimator and \$1,395/yr by the improved estimator for the difference of the ACERs and all of the 95% CIs based on the four different methods included the null value 0, meaning that the difference of the ACERs

is not likely statistically significant; see Table 3 for the CIs derived from different methods. Therefore, we may conclude that the conventional and ICD treatments cost comparably when we adjust survival experiences in the MADIT. We think that the simple weighted estimator and the improved estimator were quite different because the censoring rate was high.

## 5 Discussion

In this article, we present methods for estimation and statistical inference for the ACER-based analyses when censoring is present. We evaluated and compared different methods using simulation and data analysis. The improved estimator offers an efficiency gain, while the simple weighted estimator is simpler to implement. In practice, the improved estimator may be preferred when censoring is heavy (so that only a small number of complete cost data contribute to the simple weighted estimator). Based on our results, we believe that Bootstrap percentile and Fieller's methods could be recommended - the same advice as in the ICER literature. Yet, estimation and inference of the ACERs with censored data seem to be feasible and numerically stable, thus we may promote researchers in the field of CEA to utilize these methods without major difficulties or barriers whenever useful or justifiable. More importantly, any estimators that do not account for censoring adequately in the cost and effectiveness analyses could be erroneous. Also any cost comparison between/among groups may need to adjust for the difference in effectiveness. One way to do this task is to use the ACER.

Note that we did not consider the case with covariates in this paper. Covariates may be incorporated via regression, stratification or other means ([46], [47], [48], [49] among others). The proposed method may also be applied to cost to time ratio or other measures [50].

Finally, we generally recommend that researchers consider to report the analyses of the effectiveness, cost, ICER and ACER all together, with proper consideration of censoring, in their CEA. Computer programs written in SAS are available upon request from the first author.

## Acknowledgments

We are grateful to Dr. Alvin I. Mushlin for making the cost data of MADIT available to us. We also thank Associate Editor and two referees for their constructive comments. This research was supported by R01 HL096575.

## References

1. Weinstein MC, Stasson WB. Foundations of cost-effectiveness analysis for health and medical practice. *New England Journal of Medicine*. 1977; 296:716–721. [PubMed: 402576]
2. Gold, MR.; Siegel, JE.; Russell, LB.; Weinstein, MC., editors. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.
3. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on cost-effectiveness in health and medicine. *Journal of the American Medical Association*. 1996; 276:1339–1341. [PubMed: 8861994]
4. Willan, AR.; Briggs, AH. *Statistical Analysis of Cost-effectiveness Data*. Chichester, England: John Wiley & Sons; 2006.
5. Laska EM, Meisner M, Siegel C. Statistical inference for cost-effectiveness ratios. *Health Economics*. 1997; 6:229–242. [PubMed: 9226141]
6. Briggs A, Fenn P. Trying to do better than average: a commentary on 'statistical inference for cost-effectiveness ratios'. *Health Economics*. 1997; 6:491–495. [PubMed: 9353649]

7. Hershey JC, Asch DA, Jepson C, Baron J, Ubel PA. Incremental and average cost-effectiveness ratios: Will physicians make a distinction? *Risk Analysis*. 2003; 23:81–89. [PubMed: 12635724]
8. Hoch JS. An illustration of a current debate in cost-effectiveness analysis: average cost-effectiveness ratios vs. incremental cost effectiveness ratios. *Abstract Book Association for Health Services Research Meeting*. 1999; 16:348–349.
9. Hoch JS, Dewa CS. A clinician's guide to correct cost-effectiveness analysis: think incremental not average. *Can J Psychiatry*. 2008; 53:267–274. [PubMed: 18478830]
10. Laska EM, Meisner M, Siegel C. The usefulness of average cost-effectiveness ratios. *Health Economics*. 1997; 6:497–504. [PubMed: 9353650]
11. Gardiner J, Hogan A, Holmes-Rovner M, Rovner D, Griffith L, Kupersmith J. Confidence intervals for cost-effectiveness ratios. *Medical Decision Making*. 1995; 15:254–263. [PubMed: 7564939]
12. Gardiner, JC.; Bradley, CJ.; Huebner, M. The cost-effectiveness ratio in the analysis of health care programs. In: Rao, CR.; Sen, PK., editors. *Handbook of Statistics, Bioenvironmental and Public Health Statistics 18*. New York: North-Holland; 2000. p. 841-869.
13. Wagner DH. Nonlinear functional versions of the Neyman-Pearson lemma. *SIAM Review*. 1969; 11:5265.
14. Neyman J, Pearson E. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London Series A*. 1933; 231:289337.
15. Birch S, Gafni A. Information created to evade reality (ICER): things we should not look to for answers. *Pharmacoeconomics*. 2006; 24:1121–1131. [PubMed: 17067196]
16. Gafni A, Birch S. Incremental cost-effectiveness ratios (ICERs): The silence of the lambda. *Social Science and Medicine*. 2006; 62:2091–2100. [PubMed: 16325975]
17. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics*. 1997; 53:419–434. [PubMed: 9192444]
18. Blackhouse G, Briggs AH, O'Brien BJ. A note on the estimation of confidence intervals for cost-effectiveness when costs and effects are censored. *Medical Decision Making*. 2002; 22:173–177. [PubMed: 11958499]
19. Raikou M, McGuire A. Estimating medical care costs under conditions of censoring. *Journal of Health Economics*. 2004; 23:443–470. [PubMed: 15120465]
20. Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Economics*. 1997; 6:243–252. [PubMed: 9226142]
21. Briggs AH, Mooney CZ, Wonderling DE. Constructing confidence intervals for cost-effectiveness ratios: an evaluation of parametric and non-parametric techniques using Monte Carlo simulation. *Statistics in Medicine*. 1999; 18:3245–3262. [PubMed: 10602149]
22. Fan MY, Zhou XH. A simulation study to compare methods for constructing confidence intervals for the incremental cost-effectiveness ratio. *Health Serv Outcomes Res Method*. 2007; 7:57–77.
23. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. *Health Economics*. 1997; 6:327–340. [PubMed: 9285227]
24. Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: An example from a randomized trial. *Statistics in Medicine*. 1996; 15:1447–1458. [PubMed: 8841654]
25. Gardiner JC, Huebner M, Etton J, Bradley CJ. On parameter confidence intervals for the cost-effectiveness ratio. *Biometrical Journal*. 2001; 43:283–296.
26. Huang Y. Cost analysis with censored data. *Medical Care*. 2009; 47:S115–S119. [PubMed: 19536024]
27. Zhao, H.; Wang, H. Cost and cost-effectiveness analysis with censored data. In: Faries, DE.; Leon, AC.; Haro, JM.; Obenchain, RL., editors. *Analysis of Observational Health-Care Data Using SAS*. Cary, NC: SAS Press Series; 2010. p. 363-382.
28. Klein, JP.; Moeschberger, ML. *Survival Analysis*. New York: Springer-Verlag; 1997.
29. Bang H. Medical cost analysis: Application to colorectal cancer data from the SEER Medicare database. *Contemporary Clinical Trials*. 2005; 26:586–597. [PubMed: 16084777]
30. Bang H, Tsiatis AA. Estimating medical costs with censored data. *Biometrika*. 2000; 87:329–343.

31. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*. 1958; 53:457–481.
32. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*. 1952; 47:663–685.
33. Zhao H, Tian L. On estimating medical cost and incremental cost-effectiveness ratios with censored data. *Biometrics*. 2001; 57:1002–1008. [PubMed: 11764238]
34. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*. 1994; 89:846–866.
35. Zhao H, Bang H, Wang H, Pfeifer PE. On the equivalence of some medical cost estimators with censored data. *Statistics in Medicine*. 2007; 26:4520–4530. [PubMed: 17380543]
36. O'Hagan A, Stevens JW. On estimators of medical costs with censored data. *Journal of Health Economics*. 2004; 23:615–625. [PubMed: 15120473]
37. Satten GA, Datta S. The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician*. 2001; 55:207–210.
38. Briggs AH, Gray AM. Handling uncertainty in economic evaluations of healthcare interventions. *BMJ*. 1999; 319:635–638. [PubMed: 10473486]
39. Obenchain RL. Resampling and multiplicity in cost-effectiveness inference. *Journal of Biopharmaceutical Statistics*. 1999; 9:563–582. [PubMed: 10576404]
40. Wang H, Zhao H. A study on confidence intervals for incremental cost-effectiveness ratios. *Biometrical Journal*. 2008; 50:505–514. [PubMed: 18663759]
41. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall; 1996.
42. Fieller EC. Some problems in interval estimation. *Journal of the Royal Statistical Society, Series B*. 1954; 16:175–185.
43. Dinh P, Zhou XH. Nonparametric statistical methods for cost-effectiveness analyses. *Biometrics*. 2006; 62:576–588. [PubMed: 16918923]
44. Moss AJ, Hall WJ, Cannom DS, Daubert JP, Higgins SL, Klein H, Levine JH, Saksena S, Waldo AL, Wilber D, Brown MW, Heo M. Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. *New England Journal of Medicine*. 1996; 335:1933–1940. [PubMed: 8960472]
45. Mushlin AI, Hall WJ, Zwanziger J, Gajary E, Andrews M, Marron R, Zou KH, Moss AJ. The cost-effectiveness of automatic implantable cardiac defibrillators: Results from MADIT. *Circulation*. 1998; 97:2129–2135. [PubMed: 9626173]
46. Lin DY. Linear regression analysis of censored medical costs. *Biostatistics*. 2000; 1:35–47. [PubMed: 12933524]
47. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Economics*. 2004; 13:461–475.
48. Gardiner J, Homes-Rovner M, Goddeeris J, Rovner D, Kupersmith J. Covariate-adjusted cost-effectiveness ratios. *Journal of Statistical Planning and Inference*. 1999; 75:291–304.
49. Gardiner, JC.; Liu, L.; Luo, Z. Estimating of medical costs from a transition model. In: Balakrishnan, N.; Silvapulle, M.; Pena, E., editors. *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor PK Sen*. Vol. 1. Institute of Mathematical Statistics; 2008. p. 350–363.
50. Hartmann M, Orlin J. Finding minimum cost to time ratio cycles with small integral transit times. *Networks*. 1993; 23:567–574.

Bias and coverage probability (median length) of 95% two-sided confidence intervals for the ACER - simulation study for scenario 1

Table 1

Sample size	Survival time	Light censoring		Heavy censoring	
		Uniform	Exponential	Uniform	Exponential
100	ACER-naive	-356	-383	-572	-632
	ACER (simple-weighted)	6	7	15	4
	Fieller-normal	92.3% (316)	94.6% (359)	90.8% (366)	92.0% (441)
	Taylor-normal	91.7% (313)	93.6% (353)	89.6% (361)	90.7% (433)
	Bootstrap-percentile	91.8% (314)	93.9% (354)	92.3% (389)	92.7% (497)
	Bootstrap-normal	92.2% (318)	93.8% (357)	93.3% (393)	94.0% (503)
	ACER (improved)	6	7	13	4
	Fieller-normal	92.8% (305)	94.2% (343)	92.4% (339)	94.4% (409)
	Taylor-normal	91.7% (302)	93.6% (337)	89.6% (334)	90.4% (402)
	Bootstrap-percentile	92.5% (300)	93.5% (337)	92.2% (353)	94.4% (448)
Bootstrap-normal	92.0% (305)	93.4% (340)	93.0% (357)	94.3% (457)	
200	ACER-naive	-361	-383	-589	-638
	ACER (simple-weighted)	2	6	7	11
	Fieller-normal	92.7% (236)	94.2% (265)	93.1% (276)	94.1% (329)
	Taylor-normal	92.2% (235)	94.0% (263)	92.6% (274)	93.3% (326)
	Bootstrap-percentile	92.5% (231)	93.0% (263)	93.5% (283)	93.4% (335)
	Bootstrap-normal	92.2% (233)	93.8% (264)	93.4% (285)	94.1% (339)
	ACER (improved)	2	6	5	9
	Fieller-normal	92.8% (226)	94.6% (254)	93.2% (253)	94.4% (293)
	Taylor-normal	92.2% (225)	94.0% (252)	92.6% (252)	93.3% (290)
	Bootstrap-percentile	91.9% (223)	93.3% (251)	92.8% (254)	93.5% (290)
Bootstrap-normal	91.8% (225)	93.7% (253)	92.5% (256)	93.7% (297)	

Naive denotes the estimator that does not account for censoring in cost. Bias is the absolute bias in the point estimate. For bootstrap methods, B=1000 replicates were generated. 1000 simulations were used.

Bias and coverage probability (median length) of 95% two-sided confidence intervals for the ACER - simulation study for scenario 2

**Table 2**

Sample size	Survival time	Light censoring		Heavy censoring	
		Uniform	Exponential	Uniform	Exponential
100	ACER-naive	-569	-617	-929	-1000
	ACER (simple-weighted)	-9	-13	-7	5
	Fieller-normal	92.7% (691)	93.1% (756)	91.6% (824)	87.6% (994)
	Taylor-normal	91.4% (685)	92.2% (747)	91.0% (816)	86.6% (978)
	Bootstrap-percentile	92.0% (692)	92.4% (756)	93.3% (861)	91.1% (1075)
	Bootstrap-normal	91.6% (691)	92.4% (757)	92.7% (860)	90.3% (1070)
	ACER (improved)	-7	-10	-8	2
	Fieller-normal	93.1% (636)	92.7% (677)	92.9% (730)	94.7% (889)
	Taylor-normal	91.4% (631)	92.2% (668)	91.0% (723)	86.6% (875)
	Bootstrap-percentile	91.2% (628)	92.3% (668)	92.0% (719)	94.3% (857)
Bootstrap-normal	91.2% (625)	91.4% (666)	91.6% (718)	93.8% (882)	
200	ACER-naive	-567	-615	-927	-1011
	ACER (simple-weighted)	3	-7	12	-6
	Fieller-normal	92.1% (515)	93.4% (567)	91.8% (644)	91.7% (780)
	Taylor-normal	91.2% (513)	92.7% (564)	90.8% (641)	91.0% (775)
	Bootstrap-percentile	90.9% (513)	92.5% (566)	92.4% (655)	92.4% (798)
	Bootstrap-normal	91.1% (511)	92.7% (569)	91.6% (654)	92.0% (792)
	ACER (improved)	4	-6	13	-2
	Fieller-normal	90.9% (478)	92.7% (513)	93.5% (548)	93.9% (620)
	Taylor-normal	91.2% (476)	92.7% (510)	90.8% (546)	91.0% (616)
	Bootstrap-percentile	90.0% (471)	90.8% (510)	91.5% (533)	92.0% (578)
Bootstrap-normal	89.5% (473)	91.4% (510)	91.3% (534)	92.2% (585)	

Naive denotes the estimator that does not account for censoring in cost. Bias is the absolute bias in the point estimate. For bootstrap methods, B=1000 replicates were generated. 1000 simulations were used.

**Table 3**

## Data analysis of MADIT

	ICD (95% CI)	Conventional (95% CI)	Difference (95% CI)
Effectiveness*	3.45 yrs (3.26, 3.65)	2.65 yrs (2.34, 2.96)	0.80 yr (0.43, 1.17)
Cost (simple-weighted)*	\$110109 (96526, 123692)	\$70044 (51879, 88209)	\$40065 (17383, 62747)
Cost (improved)*	\$99548 (88786, 110,310)	\$72754 (56023, 89485)	\$26794 (6901, 46687)
ACER (simple-weighted)	\$31883/yr	\$26408/yr	\$5475/yr
Fieller-normal	(27801, 36121)	(19290, 34198)	(-3061, 14011) <sup>+</sup>
Taylor-normal	(27730, 36036)	(19014, 33803)	(-2773, 13723)
Bootstrap-percentile	(27962, 36140)	(19992, 35088)	(-4030, 13788)
Bootstrap-normal	(27716, 36050)	(18801, 34015)	(-3289, 14191)
ACER (improved)	\$28825/yr	\$27430/yr	\$1395/yr
Fieller-normal	(25603, 32173)	(20799, 34746)	(-6313, 9140) <sup>+</sup>
Taylor-normal	(25546, 32104)	(20513, 34346)	(-6259, 9049)
Bootstrap-percentile	(25568, 32099)	(20753, 36503)	(-7693, 8856)
Bootstrap-normal	(25587, 32063)	(19723, 35137)	(-7043, 9833)

\* for cost and effectiveness and their differences, normal-based formulae were used for CI.

<sup>+</sup> for CIs for the difference of the ACERs using Fieller's method, we used the approximate relationship of standard error = (length of the CI)/(2\*1.96).

For bootstrap methods, B=1000 replicates were generated.