



Published in final edited form as:

*J Biopharm Stat.* 2016 ; 26(3): 552–564. doi:10.1080/10543406.2015.1052482.

## MEDIAN-BASED INCREMENTAL COST-EFFECTIVENESS RATIOS WITH CENSORED DATA

Heejung Bang<sup>1</sup> and Hongwei Zhao<sup>2</sup>

<sup>1</sup>Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, Davis, California, USA

<sup>2</sup>Department of Epidemiology and Biostatistics, School of Public Health, Texas A&M Health Science Center, College Station, Texas, USA

### Abstract

Cost-effectiveness is an essential part of treatment evaluation, in addition to effectiveness. In the cost-effectiveness analysis, a measure called the incremental cost-effectiveness ratio (ICER) is widely utilized, and the mean cost and the mean (quality-adjusted) life years have served as norms to summarize cost and effectiveness for a study population. Recently, the median-based ICER was proposed for complementary or sensitivity analysis purposes. In this paper, we extend this method when some data are censored.

### Keywords

censoring; cost-effectiveness analysis; ICER; median

### 1. Introduction

Cost-effectiveness analysis (CEA) is an economic analysis that evaluates cost and effectiveness for competing treatments to understand the economic consequence of a new treatment option. The incremental cost-effectiveness ratio (ICER), defined as the extra cost incurred for gaining one unit of health benefit, has long been used as a standard measure in the CEA. The ICER is particularly useful when a new treatment is more effective but also more expensive than its competitor.

The mean has been most widely accepted as a summary measure for cost as well as effectiveness for various CEA measures including the ICER. Particularly, the ‘arithmetic’ mean cost is relevant since the total cost can be derived directly from the mean, and ideally CEA should address the total cost for treating *all* patients (Barber and Thompson, 2000; Ramsey, et al., 2005). Yet, due to high skewness in cost data, median or other quantiles deserve some attention. Recently, the median-based ICER was proposed and it was illustrated that the mean- and median-based ICERs could yield qualitatively different results, including different signs (Bang and Zhao, 2012). Their finding warrants some discussion in

---

Address correspondence to Heejung Bang, Division of Biostatistics, Department of Public Health Sciences, University of California, One Shields Ave., Davis, CA 95616, USA; hbang@ucdavis.edu.

the decision making process although the mean cost still serves as the parameter of primary interest. Also, it has been pointed out different parameters were being used in effectiveness analysis and CEA. For example, the CONSORT recommends hazard ratio or difference in *median* survival times as the effectiveness measure for censored survival data (CONSORT, 2010; Guyot, et al., 2011). It also has been noted that different statistical software packages can yield different mean estimates from the same data (e.g., due to different endpoints used), while the median was unchanged in the presence of censoring (Barker, 2009). As such, a natural extension would be to handle (right) censoring in the median-based ICER, which is the purpose of this paper.

This paper is organized as follows: In Section 2, we review existing methods, and introduce the median-based ICER that can handle censored data, its estimation and inference procedures. Simulation results for the median cost and median-based ICER along with the mean counterparts are presented in Section 3. Section 4 reports the analyses of two related cardiovascular clinical trials using the conventional and expanded methods. Discussions are provided in Section 5.

## 2. Methods

### 2.1 Notation and assumptions

We will consider one sample problem (e.g., a single group) first, followed by two sample problem that is needed for the ICER. Of note, for simpler exposition, we will use ICER for the parameter as well as its estimator.

For the  $i$ th person in the study, let  $T_i$  denote his/her survival time (so that the endpoint of interest is mortality without loss of generality) and  $C_i$  censoring time, where these two times are assumed to be independent. This assumption is usually satisfied when censoring is mainly caused by administrative reasons such as staggered entry and limited duration of a study, and is commonly imposed in standard survival analyses. Due to censoring, not all  $T_i$ s are observed, instead, the follow-up time  $X_i = \min(T_i, C_i)$  along with the censoring indicator  $\delta_i = I(T_i \leq C_i)$  is observed:  $\delta_i = 1$  means the  $i$ th person died before being censored, and  $\delta_i = 0$  this person was censored before death, with  $I(\cdot)$  an indicator function. We denote  $M_i$  as the *observed* total cost for the  $i$ th person, accumulated from time 0 to  $X_i$ . Thus, for the person who died before being censored,  $M_i$  is the true cost and  $X_i$  is the true survival time.

Here, because of censoring, it is impossible to estimate the cost over the entire health history without making distributional assumptions. Therefore, we only consider cost accumulated up to a fixed time point  $L$ , where one has a reasonable amount of data available for the time period  $[0, L]$ , and  $L$  could be equal to or shorter than study duration. Hence,  $T_i$  should be replaced by  $T_i^L = \min(T_i, L)$  but for ease of notation, we will suppress the superscript  $L$  and use  $T_i$  throughout the paper. Thus, observed data to be used for the proposed method consist of three variables  $\{X_i, \delta_i, M_i\}$  for  $i=1, \dots, n$ .

### 2.2 Estimating the mean cost and the ICER with censoring: Review

Suppose that we want to estimate the mean of the cost up to a maximum time of  $L$ . If every patient is followed up to time  $L$  or until one's death, we would have complete cost data with

no censoring and the standard statistical methods such as the sample mean or regression could be used for estimating the mean cost. However, in many situations that entail follow-up, cost as well as survival data are not completely observed for every patient due to censoring. To handle this issue, an inverse-probability-weighted (IPW) estimator for the mean cost has been proposed:

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

where  $\hat{K}(T_i)$  is the Kaplan-Meier estimator for  $K(t)=P(C>t)$ , survival function of the censoring time  $C$ , evaluated at time  $T_i$  (Bang and Tsiatis, 2000). The mean survival time can be estimated by Eq (1) with  $T$  in place of  $M$ :

$$\hat{\mu}^T = \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i T_i}{\hat{K}(T_i)}$$

which is equivalent to the area under the Kaplan-Meier survival curve (Wahed, 2011).

Next, we consider a study with two groups (e.g., two arms in RCT) and the associated ICER.

For arm  $k(k=0,1)$ , we denote the mean cost by  $\mu_k^M$  and the mean survival time by  $\mu_k^T$ . The standard ICER based on the mean cost and the mean survival time is estimated by (Willan, et al., 2002; Zhao and Tian, 2001):

$$\frac{\hat{\mu}_1^M - \hat{\mu}_0^M}{\hat{\mu}_1^T - \hat{\mu}_0^T}.$$

Alternative measures for effectiveness such as the median survival time have been used in the ICER estimation as well (Cordony, et al., 2008; Gardiner, et al., 2000; Vu, et al., 2008), and can be written as:

$$\frac{\hat{\mu}_1^M - \hat{\mu}_0^M}{\hat{m}_1^T - \hat{m}_0^T}.$$

A major rationale is that the median survival time  $\hat{m}_k^T$  is easily estimated from the Kaplan-Meier curve as long as the estimated survival probability reaches 0.5. Its advantages and wide acceptance as effectiveness measure (vs. mean survival time) have been well documented (Brookmeyer, 2005; CONSORT, 2010; Gardiner, et al., 1986; Guyot, et al., 2011). The methods described in this section can be implemented for 4 (=2×2) different settings defined by (mean or median cost)×(mean or median effectiveness) in a unified framework.

Remarks: In general,

- a. Skewness is more severe for cost data, compared to survival data, so that the impact of the mean vs. median in the numerator could be a greater concern.

- b. The mean above is restricted mean, not unrestricted mean, the reason for which was explained previously, and can also be seen in the Appendix where we compared estimability of the mean and median survival times and costs with or without censoring (Bang and Zhao, 2014; Huang, 2009).
- c. The confidence interval (CI) for the mean-based ICER can be obtained by various methods including Fieller's method or the bootstrap method, e.g. (Wang and Zhao, 2008; Zhao and Tian, 2001). If the median survival time is used in ICER, a bootstrap method can be employed to obtain a CI, similar to the approach stated in the following subsection.

### 2.3 Estimating the median cost and the ICER with censoring

We apply the IPW to an estimating equation for the median cost. The following equation can be used to find  $m$ :

$$\frac{1}{n} \sum_{i=1}^n \frac{\Delta_i \{I(M_i \leq m) - 0.5\}}{\hat{K}(T_i)} \approx 0$$

where  $m$  denotes the true median and  $\approx$  means approximation due to the discontinuous nature of the estimating function. An alternative expression for  $m$  is

$$\hat{m} = \inf \left\{ x : \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i I(M_i > x)}{\hat{K}(T_i)} \leq 0.5 \right\}.$$

The resulting median estimator can be shown to be consistent and asymptotically normal (Bang and Tsiatis, 2002; Ying, et al., 1995; Zhao, et al., 2012).

Now, we introduce the ICER based on the median costs in place of the mean costs. As long as the survival function reaches 0.5 before the time limit  $L$  so that the median survival time is estimable, we propose:

$$\frac{\hat{m}_1^M - \hat{m}_0^M}{\hat{m}_1^T - \hat{m}_0^T} \quad (2)$$

where  $\hat{m}_k^M$  and  $\hat{m}_k^T$  are the estimators, respectively, for the median cost and the median survival time in the  $k^{\text{th}}$  arm. If the survival function does not reach 0.5, we may use a projected median based on extrapolation, or restricted time mean:

$$\frac{\hat{m}_1^M - \hat{m}_0^M}{\hat{\mu}_1^T - \hat{\mu}_0^T}. \quad (3)$$

Eq (3) can be particularly useful when effectiveness is a binary outcome so that the expected value or mean is a probability. This formula has been used when data are uncensored (Fowler, et al., 2014).

Along with the point estimate, we suggest nonparametric bootstrap methods for making statistical inference for the proposed ICERs; a CI can be constructed and a CE plane with bootstrap replicates can guide interpretation (Black, 1990; Laupacis, et al., 1992; Obenchain, 1999). Specifically, we recommend the quadrant-based bootstrap method that can be used for the mean- and median-based ICERs together in all possible scenarios in the systematic and unified fashion described in (Bang and Zhao, 2012). An appropriate bootstrap method (e.g., percentile, re-ordered and wedge-based) must be selected depending on how many/ which quadrants contain bootstrap replicates in the CE plane (Obenchain, 1999; Wang and Zhao, 2006).

### 3. Simulation Study

We conducted a simulation study to examine the performance of the proposed method with limited sample sizes. We adopted simulation settings similar to the ones used in prior research (Bang and Zhao, 2012; Lin, et al., 1997). We conducted simulations for the median cost and median-based ICER. Of note, the mean cost and mean-based ICER (which are the current standard) were also computed and reported side by side for comparison in tables and figures. We imposed time-restriction, which is typical in practice, so that strictly speaking, mean should be read ‘restricted’ mean below. For succinct presentation, we provided some details in the footnotes of Tables, not in text.

#### 3.1 Cost estimation in one group

Data for individual patients were independently and identically generated. The total cost for each patient consists of three cost components: diagnostic cost that incurs at the beginning of the study; annual cost that incurs annually during follow-up; and terminal cost that incurs around death, so diagnostic and annual costs are relevant for all patients, while terminal cost is only relevant for those who died. We generated cost data in our “Cost scenario 1” as:

- diagnostic cost  $\sim \exp(N(0,1)*1+5)+1000$
- annual cost  $\sim U[0,1]*1000+1000$
- terminal cost  $\sim \exp(N(0,1)*1.5+6)+1000$

where  $N$  and  $U$  denote a normal and uniform distribution and  $\exp$  denotes exponentiation. Our interest in all simulations is to estimate the median of the patient-level total cost accumulated up to death or 10 years whichever comes first.

We generated survival times from two distributions: a uniform distribution on  $U[0, 10]$  years and an exponential distribution with a mean of 6 years. The true median cost is ~\$10,514 for the uniform and ~\$9,531 for the exponential survival time, where true costs were numerically estimated using 1,000,000 uncensored random samples.

Two levels of censoring were considered:  $C$  was generated from  $U[0, 20]$  and  $U[0, 12.5]$  years, independently of all other variables. The former was referred to as light censoring, resulting in ~25% censored data, and the latter as heavy censoring, resulting in ~40% censored data. Of note, 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 at the 10th year. Five hundred simulations were carried out for each setting. The same set of simulations was repeated for sample size (N) of 100, 200 and 500, and 500 bootstrap samples were generated for computing CI by the percentile method.

Simulation results for the median cost estimation are summarized in Table 1 in terms of the relative bias, and the median length and the coverage probability (CP) of the 95% CI estimates, computed from 500 simulation runs.

In this simulation study, we observe the bias estimate is small for all cases, which is mostly less than 1%. The length of 95% CIs is shorter when N is larger or censoring rate is lower, as anticipated. CPs appear to be fairly accurate (94.0 to 95.2%) when N is 500. Expected and stable performance is observed for N as low as 100, as previously observed (Zhao, et al., 2012). In all simulation settings, we observe CI is much (20–90%) wider for the median cost, compared to the mean cost.

We also tried the second set of simulations for symmetric cost components as follows:

- diagnostic cost  $\sim [\max(N(0,1)*1.5+6, 0)]*1000$
- annual cost  $\sim U[0,1]*1000$
- terminal cost  $\sim [\max(N(0,1)+2, 0)] * 1000$  (4)

Under these, we found that CPs are closer to 95% and observed biases are nearly 0% for both the mean and median cost, but the CIs for the median cost are about 30% wider than those for the mean cost.

### 3.2 ICER estimation in two groups

Next, we conducted simulations for the ICERs. For this part, we focused on a common and important scenario where CEA is most pertinent, so called, “NE quadrant” scenario where a new treatment is more effective but more expensive (Black, 1990; Hoch and Smith, 2006). For data generation, we used  $U[0, 10]$  for group 1 and  $U[0, 8]$  for group 2 for the uniform survival time setting. We also generated exponential survival time with a mean (median) of 6 ( $6*\log 2 \approx 4.16$ ) years for group 1 and of 4.5 ( $4.5*\log 2 \approx 3.12$ ) for group 2. In the light censoring scenarios, censoring rate is ~25% and ~20% in the two arms and in the heavy censoring scenarios, it is ~40% and ~32%. With the same 10 year restriction, cost data were generated from (‘Cost scenario 1’):

- diagnostic cost  $\sim \exp(N(0,1)*1+2)+1000$
- annual cost  $\sim U[0,1]*1000+1000$
- terminal cost  $\sim \exp(N(0,1)*1.1+1)+1000$

for group 1; and

- diagnostic cost  $\sim \exp(N(0,1)*0.9+2)+1000$
- annual cost  $\sim U[0,1]*800+500$

- terminal cost  $\sim \exp(N(0,1)*1.1)+1000$

for group 2. The true median-based ICER is  $\sim 4,019$  and  $\sim 3,508$  for uniform and exponential survival distribution settings, respectively.

The performance of the estimated ICERs and their 95% CIs is summarized in Table 2a. For the proposed median-based ICER, the (relative) bias in the point estimate shows the range of 0.5 to 4.0%. We learned that the ICER estimation, which is a ratio estimation, requires substantially larger  $N$  in order to achieve reasonable performance. We also observed that CIs tend to be conservative, as reflected in CPs larger than 95% in almost all cases we simulated, even as large as 99%. Yet, as  $N$  increases, conservativeness somewhat diminishes, but always fully (say, when we tried  $N=2000$ ). As above, the length of CI is shorter when  $N$  is larger or censoring rate is lower. Markedly reduced lengths of CIs and more accurate CPs are observed for the mean-based ICER, especially for the uniform survival distribution.

Additionally, we simulated ‘Cost scenario 2’, and results are presented in Table 2b. We generated the true survival time from normal distributions (with mean of 6 and 4.5 years for group 1 and 2, respectively, and a unit variance, truncated at time 0). We used the same distribution for exponential survival times and censoring as described in Section 3.1. Under these, censoring rate is increased (e.g., 30% for light and 50% for heavy). For group 1, each cost component was symmetrically generated following (4) as in Section 3.1, and for group 2, we used:

- diagnostic cost  $\sim [U[0,1]*1.4+5]*1000$
- annual cost  $\sim U[0,1]*1000$
- terminal cost  $\sim U[0,1]*1000$ .

Under these cost scenarios, normal survival times provide notably shorter lengths of CIs. Small bias in point estimate was maintained in all scenarios,  $<2\%$ . The observed CPs for the median-based ICERs are closer to nominal 95% in most cases so that the conservativeness of the CPs is ameliorated, compared to Cost scenario 1. Interestingly, normal vs. exponential survival times with the same cost components show profound difference in variability of the estimates measured by the length of CI, partly as a result of the big difference of variability of survival times generated from these two distributions.

## 4. Examples

### 4.1 Study overview of MADIT I and II

To illustrate the proposed methods, we analyzed the data from two cardiovascular clinical trials, the Multicenter Automatic Defibrillator Implantation Trial (MADIT) I and II. MADIT I was conducted in 1990–1996, to examine the effectiveness of an implantable cardioverter-defibrillator (ICD) in prevention of sudden death for patients at high risk of ventricular arrhythmia (Moss, et al., 1996). A total of 181 patients were enrolled from 36 centers, with 89 patients assigned to the treatment group to receive an ICD, and 92 to a conventional intervention. The first enrolled patient was followed for 61 months and the last for  $<1$  month, with average follow-up of 27 months. After study completion, it was shown that the



use of ICD led to improved survival, compared with the conventional intervention (hazard ratio of 0.46,  $p=0.009$ ).

After the completion of MADIT I, MADIT II was conducted in 1997–2002 to identify patients with coronary heart disease who would benefit from ICD. In MADIT II, patients with a prior myocardial infarction and a left ventricular ejection fraction of 0.30 or less were randomly assigned to ICD ( $N=742$ ) or conventional therapy ( $N=490$ ), with all-cause mortality as the primary endpoint (Moss, et al., 2002). The findings revealed that in these patients, the prophylactic use of an ICD, in addition to medications, significantly reduced the risk of death (hazard ratio of 0.69,  $p=0.016$ ).

The data were heavily censored, with 70% of patients in the ICD arm and 48% in the conventional therapy arm in MADIT I and the corresponding rates were 76% and 72% in MADIT II. Due to important economic consequences, CEA were performed for these trials (Mushlin, et al., 1998; Zwanziger, et al., 2006). As done in the original CEA, we restricted the duration of the cost estimation to  $L=4$  years for MADIT I and 3.5 years for MADIT II, and both costs and survival times were discounted at 3% annual rate.

## 4.2 Data analysis

We computed the (restricted) mean and median costs, and the (restricted) mean survival time as effectiveness measure, as median survival is not estimable because of the low event rate in limited follow-up. The corresponding ICERs along with 95% CIs were estimated from 1,000 bootstrap samples. We denote the ICER using the mean cost by  $ICER_{mean}$  and that using the median cost by  $ICER_{median}$ . Data analyses are reported in Table 3.

In MADIT I, the mean survival time was 2.65 years for the conventional arm and 3.45 years for the ICD arm, which yielded the difference of 0.80 year (95% CI: 0.44–1.18). In this trial, the mean and median costs for the conventional arm were \$70K and \$54K, respectively, while the mean and median costs for the ICD arm were much larger, \$110K and \$112K. Interestingly, the median cost is slightly larger than the mean cost in the ICD arm. Also, the cost distributions for the two arms are substantially different; see the upper panels in Figure 1. The cost difference was computed as \$40K (16K–63K) for the mean and \$58K (20K–92K) for the median. The ICD increased survival significantly but incurred larger costs, so that CEA could be well justified in this situation. The  $ICER_{mean}$  was estimated as \$50K/yr (18K–103K/yr) and the  $ICER_{median}$  as \$73K/yr (25K–161K/yr). CIs were computed by the bootstrap percentile method as the bootstrap replicates lie in the NE and SE quadrants that show natural ordering of the ICER estimates (Bang and Zhao, 2012). The corresponding CE planes are shown in Figure 2 with the observed ICERs with the 95% CIs and bootstrap replicates – we recommend the CE plane is interpreted together with the 5 decision regions (Bang and Zhao, 2014; Laupacis, et al., 1992; Obenchain, 1999).

In contrast, MADIT II had the mean survival time of 2.73 years in the conventional arm and 2.89 years in the ICD arm, which yield the difference of 0.16 year (0.04–0.31). Here, the mean and median costs for the conventional arm were \$47K and \$33K, whereas those for the ICD arm were \$91K and \$67K, respectively. Again, costs as well as survival differences are statistically significant, i.e., excluding 0 in CIs. Cost distributions are presented in Figure



1, lower panels. The  $ICER_{mean}$  was estimated as \$267K/yr (124K–1,422K/yr) and the  $ICER_{median}$  as \$206K/yr (89K–987K/yr). Here, CIs were estimated using the *re-ordered* percentile bootstrap method because replicates lie in the NE and NW quadrants (Bang and Zhao, 2012; Wang and Zhao, 2008). A small number of bootstrap replicates lie in the NW quadrant due to a small survival benefit; see Figure 2, lower panels. Also, large ICERs and wide CIs must be due to a small denominator in the ICER, albeit statistically significant, and high variability owing to small/moderate  $N$ s of uncensored data, which are common in CEA with relatively short study durations.

Depending on the ICER threshold chosen for the willingness to pay, the mean vs. median-based analyses may lead to different conclusions and decisions. Currently, thresholds such as 50K, 100K and 265K for the ICER have been used for the determination of cost-effectiveness (Ubel, et al., 2003). For example, if one had adopted 50K for MADIT I or 265K for MADIT II, the conclusion based on the mean vs. median costs could be different.

## 5. Discussion

In this paper, we proposed the median-based ICER in the presence of censoring, studied its numerical performance in simulation studies and conducted analyses with two cardiovascular disease trials. This work can be regarded as an extension of the median-based ICER in the absence of censoring that has been recently proposed. It is well documented that censoring mechanisms for survival data and cost data are different and should be properly accounted for in the CEA.

The usefulness of ICER has been well documented in the literature, where the economic concept of “increment” or “difference” and combining cost and effect into one measure are essential. Other alternatives that solve some problems associated with the ICER such as subjectivity of threshold and possible numerical instability are also available, including incremental net benefit, cost-effectiveness acceptability curve, and average cost-effectiveness ratio (Bang and Zhao, 2014; Fenwick, et al., 2001; Laska, et al., 1997; Willan and Lin, 2001).

We assert that mean and median-based methods be considered together for any highly skewed data, including medical cost although the mean-based ICER has been nearly unanimously advocated by health economists in current practice (Gold, et al., 1996; Thompson and Barber, 2000; Weinstein, et al., 1996). In some studies, the mean vs. median-based analyses give fundamentally different answers, which would make the interpretation of the results and the ultimate conclusion difficult. When that happens, presenting only one analysis between these two could be misleading although we agree on the importance of arithmetic mean in healthcare policy decision (Bang and Zhao, 2012; Thompson and Barber, 2000). The median-based ICER may be particularly relevant or useful as the CONSORT as well as statistical community recommend median survival time, instead of mean survival time, whenever data are censored and the median survival time is estimable. A long history of discrepancy (e.g., median survival time in the effectiveness analysis and mean survival time in the CEA) has been noted (Bang and Zhao, 2014; Guyot, et al., 2011).

Through simulations and data analyses, we found that the median cost and median-based ICER tend to be much more variable than the mean counterparts. Moreover, median-based ICER seems to be more vulnerable to numerical instability. We speculate it is caused by the non-continuous nature of the survival function for costs and survival times, among other reasons. In general, ICER estimation would need much larger sample size than cost estimation and cost estimation needs larger sample size than effectiveness estimation. Indeed, some undesirable behaviors have been noted for median-based or non-smooth parameters (Boos and Stefanski, 2013; Efron and Tibshirani, 1993; Price and Bonnet, 2002; Sima and Gonen, 2013). Although ICERs can be estimated *numerically* as plausible values in our MADIT data analyses with relatively small/moderate sample size, high censoring rate (thus, resulting in even smaller sample size of uncensored data), and small denominator, we should pay attention to possible numerical instability or variability that could be too substantial to be informative, as reflected in very wide CIs. Sensitivity analyses would be critical in this case.

In this paper, we adopted bootstrap methods for statistical inference. But more research on other inference methods is warranted. Also, as in the example we presented, if the median survival time is not estimable (e.g., due to low event rate with short term follow-up), we may need to use the restricted mean. But if the median survival time or other meaningful quantile (e.g., upper quartile) is estimable, these parameters might be advantageous, compared to the mean survival time that is destined to be underestimated whenever the largest survival time is censored (Barker, 2009; Brookmeyer, 2005). A reality is that the statistical parameter of interest could be dictated by estimability and stability of estimation, especially when we use limited but *real* data. When a small number of events were observed, projection for survival time and the associated costs beyond the study duration may be considered, which could be useful as ancillary or exploratory analysis (Zwanziger, et al., 2006). Currently, a predominating approach in the field of CEA is based on modeling, simulation and projection for lifetime, where cumbersome data-related anomalies would be absent. Although this approach could be indispensable for policy research, a number of strong assumptions are unavoidable. Finally, we used the standard IPW technique for obtaining consistent estimators. More efficient estimators might be proposed in future.

To conclude, despite some limitations (which may be more rigorously studied and addressed in future), the mean- and median-based ICERs look feasible and may be used in parallel in practice whenever appropriate – possibly the median-based version as a sensitivity analysis. We believe that the mean and median-based CEA complement each other, rather than one is correct and the other is misleading (Thompson and Barber, 2000); together they may provide a more complete analysis of available data and lead to better informed decisions, including teaching us the limitations of available evidence.

## Acknowledgments

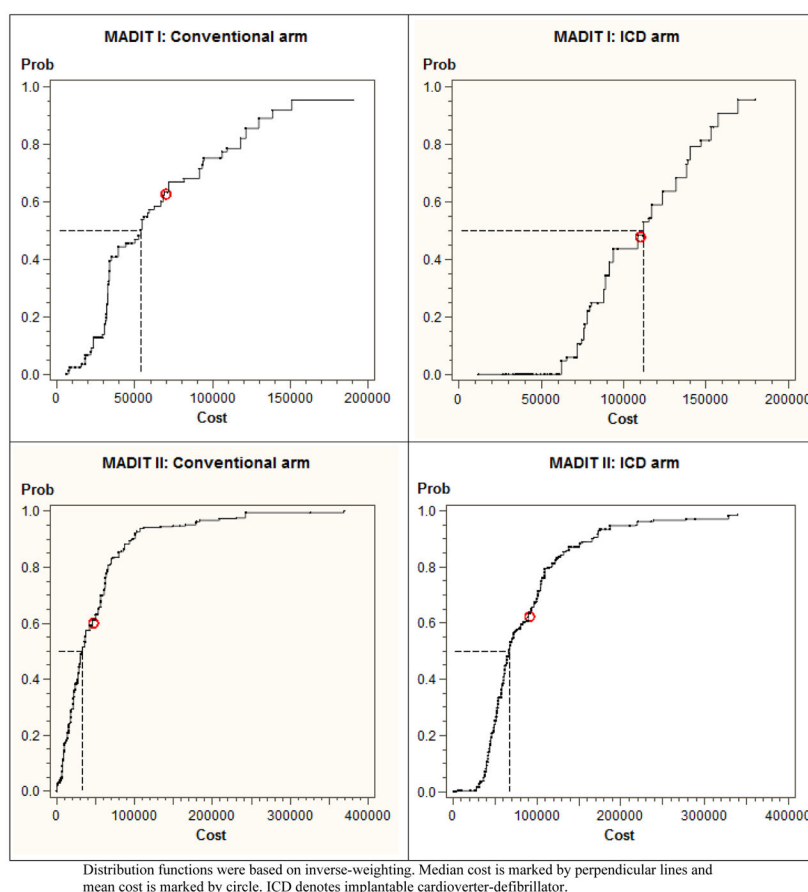
The authors are very grateful to Drs. Alvin I. Mushlin, Arthur J. Moss and Jack Zwanziger for making the cost data of MADIT I and II available to us. We thank Ms. Ya-Lin Chiu for programming advice. This research was supported by R01 HL096575. H. Bang was additionally supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), through grant #UL1 TR000002.

**Funding Source:** This research was supported by R01 HL096575 from the National Heart, Lung and Blood Institute. H. Bang was additionally supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), through grant #UL1 TR000002.

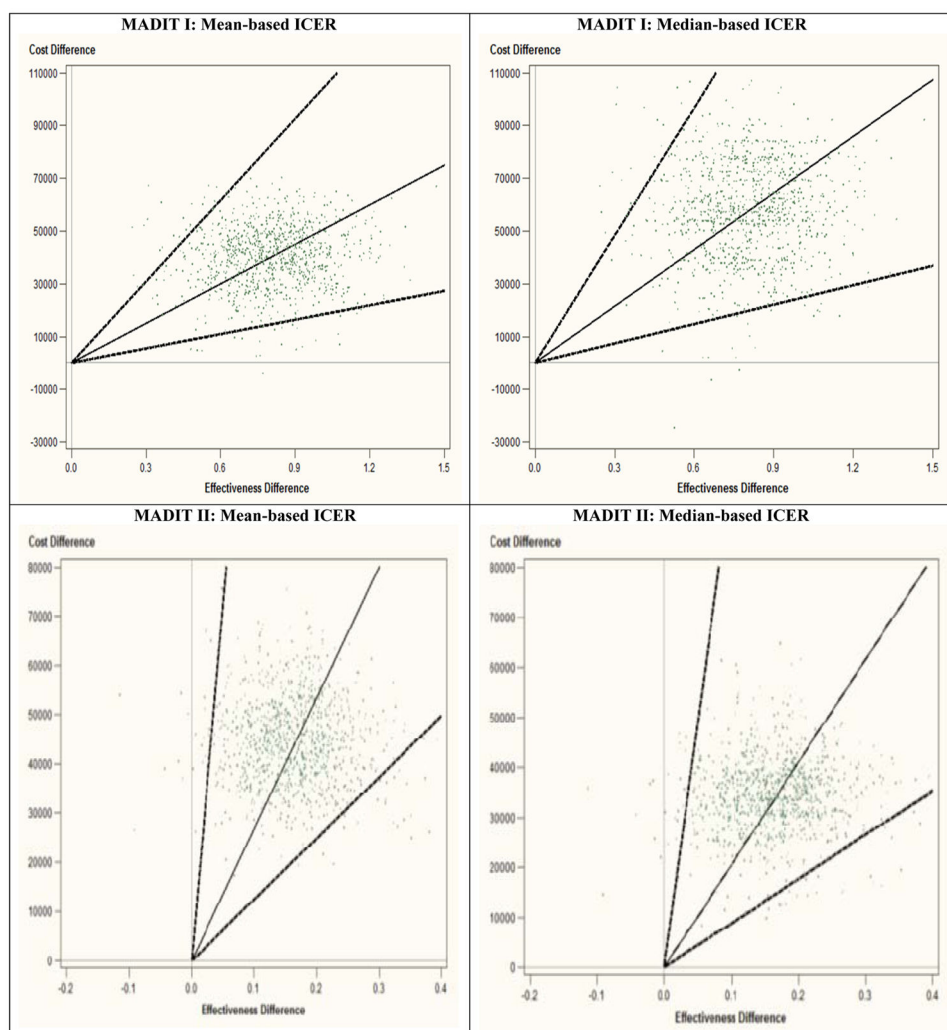
## References

- Bang H, Tsiatis AA. Estimating medical costs with censored data. *Biometrika*. 2000; 87:329–343.
- Bang H, Tsiatis AA. Median regression with censored cost data. *Biometrics*. 2002; 58:643–649. [PubMed: 12229999]
- Bang H, Zhao H. Average cost-effectiveness ratio with censored data. *Journal of Biopharmaceutical Statistics*. 2012; 22:401–415. [PubMed: 22251182]
- Bang H, Zhao H. Median-based incremental cost-effectiveness ratio (ICER). *Journal of Statistical Theory and Practice*. 2012; 6:428–442. [PubMed: 23441022]
- Bang H, Zhao H. Cost-effectiveness analysis: a proposal of new reporting standards in statistical analysis. *Journal of Biopharmaceutical Statistics*. 2014; 24:443–460. [PubMed: 24605979]
- Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. *Statistics in Medicine*. 2000; 19:3219–3236. [PubMed: 11113956]
- Barker C. The mean, median, and confidence intervals of the Kaplan–Meier survival estimate - computations and applications. *The American Statistician*. 2009; 63:78–80.
- Black WC. The CE plane: a graphic representation of cost-effectiveness. *Medical Decision Making*. 1990; 10:212–214. [PubMed: 2115096]
- Boos, DD.; Stefanski, LA. *Essential Statistical Inference*. Springer; New York: 2013.
- Brookmeyer R. Median survival time. *Encyclopedia of Biostatistics*. 2005
- CONSORT. 2010. <http://www.consort-statement.org/checklists/view/657-harms/1014-outcomes-and-estimation>
- Cordony A, et al. Cost-effectiveness of pemetrexed plus cisplatin: malignant pleural mesothelioma treatment in UK clinical practice. *Value in Health*. 2008; 11:4–12. [PubMed: 18237355]
- Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. Chapman and Hall; Boca Raton: 1993.
- Fenwick E, Claxton K, Sculpher MJ. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Economics*. 2001; 10:779–787. [PubMed: 11747057]
- Fowler R, et al. Cost-effectiveness of dalteparin vs unfractionated heparin for the prevention of venous thromboembolism in critically ill patients. *JAMA*. 2014; 312:2135–2145. [PubMed: 25362228]
- Gardiner, J.; Susarla, V.; Van Ryzin, J. *Adaptive Statistical Procedures and Related Topics*. Institute of Mathematical Statistics; Hayward, CA: 1986. Estimation of the median survival time under random censorship; p. 350–364.
- Gardiner JC, Bradley CJ, Huebner M. The cost-effectiveness ratio in the analysis of health care programs. *Handbook of Statistics, Bioenvironmental and Public Health Statistics in Medicine*. 2000; 18:841–869.
- Gold, MR., et al. *Cost-effectiveness in Health and Medicine*. Oxford University Press; New York: 1996.
- Guyot P, et al. Survival time outcomes in randomized, controlled trials and meta-analyses: the parallel universes of efficacy and cost-effectiveness. *Value in Health*. 2011; 14:640–646. [PubMed: 21839400]
- Hoch JS, Smith MW. A guide to economic evaluation: methods for cost-effectiveness analysis of person-level data. *Journal of Traumatic Stress*. 2006; 19:787–797. [PubMed: 17195977]
- Huang Y. Cost analysis with censored data. *Medical Care*. 2009; 47:S115–119. [PubMed: 19536024]
- Laska EM, Meisner M, Siegel C. The usefulness of average cost-effectiveness ratios. *Health Economics*. 1997; 6:497–504. [PubMed: 9353650]
- Laupacis A, et al. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *CMAJ*. 1992; 146
- Lin DY, et al. Estimating medical costs from incomplete follow-up data. *Biometrics*. 1997; 53:419–434. [PubMed: 9192444]

- Moss A, et al. Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. *NEJM*. 2002; 346:877–883. [PubMed: 11907286]
- Moss AJ, et al. Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. *NEJM*. 1996; 335:1933–1940. [PubMed: 8960472]
- Mushlin AI, et al. The cost-effectiveness of automatic implantable cardiac defibrillators: Results from MADIT. *Circulation*. 1998; 97:2129–2135. [PubMed: 9626173]
- Obenchain RL. Resampling and multiplicity in cost-effectiveness inference. *Journal of Biopharmaceutical Statistics*. 1999; 9:563–582. [PubMed: 10576404]
- Price RM, Bonnet DG. Distribution-free confidence intervals for difference and ratio of medians. *Journal of Statistical Computation and Simulation*. 2002; 72:119–124.
- Ramsey S, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value in Health*. 2005; 8:521–533. [PubMed: 16176491]
- Sima CS, Gonen M. Optimal cutpoint estimation with censored data. *Journal of Statistical Theory and Practice*. 2013; 7:345–359.
- Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? *BMJ*. 2000; 329:1197–1200. [PubMed: 10784550]
- Ubel PA, et al. What is the price of life and why doesn't it increase at the rate of inflation? *Archives of Internal Medicine*. 2003; 163:1637–1641. [PubMed: 12885677]
- Vu T, et al. Survival outcome and cost-effectiveness with docetaxel and paclitaxel in patients with metastatic breast cancer: a population-based evaluation. *Annals of Oncology*. 2008; 19:461–464. [PubMed: 18006892]
- Wahed A. On the equivalence of inverse-probability-of-censoring-weighted and Kaplan-Meier estimators. *Journal of the Applied Statistical Science*. 2011; 18
- Wang H, Zhao H. Estimating incremental cost-effectiveness ratios and their confidence intervals with differentially censored data. *Biometrics*. 2006; 62:570–575. [PubMed: 16918922]
- Wang H, Zhao H. A study on confidence intervals for incremental cost-effectiveness ratios. *Biometrical Journal*. 2008; 50:505–514. [PubMed: 18663759]
- Weinstein MC, et al. for the Panel on Cost-Effectiveness in Health and Medicine. Recommendations for reporting cost-effectiveness analyses. *JAMA*. 1996; 276:1253–1258. [PubMed: 8849754]
- Willan A, Lin D. Incremental net benefit in randomized clinical trials. *Statistics in Medicine*. 2001; 20:1563–1574. [PubMed: 11391688]
- Willan A, et al. Using inverse-weighting in cost-effectiveness analysis with censored data. *Statistical Methods in Medical Research*. 2002; 11:539–551. [PubMed: 12516988]
- Ying Z, Jung S-H, Wei LJ. Survival analysis with median regression models. *JASA*. 1995; 90:178–184.
- Zhao H, Tian L. On estimating medical cost and incremental cost-effectiveness ratios with censored data. *Biometrics*. 2001; 57:1002–1008. [PubMed: 11764238]
- Zhao H, et al. Nonparametric inference for median costs with censored data. *Biometrics*. 2012; 68:717–725. [PubMed: 22364557]
- Zwanziger J, et al. The cost effectiveness of implantable cardioverter-defibrillators: results from the MADIT-II. *JACC*. 2006; 47:2310–2318. [PubMed: 16750701]



**Figure 1. Cost distributions for conventional and ICD arms in MADIT I and II**  
 Distribution functions were based on inverse-weighting. Median cost is marked by perpendicular lines and mean cost is marked by circle. ICD denotes implantable cardioverter-defibrillator.



**Figure 2.**  
CE planes for Mean and Median for MADIT I and II

**Table 1**

Simulation results for Median and Mean costs

Survival time	N	Censoring	Median cost			Mean cost		
			Bias (%)	Length of CI	CP	Bias (%)	Length of CI	CP
Uniform	100	Light	-0.2	3258	94.0	0.2	2287	93.6
		Heavy	1.2	3699	94.0	0.2	2902	94.2
	200	Light	-0.3	2193	92.4	0.1	1652	95.0
		Heavy	0.1	2446	93.0	-0.6	1980	93.0
	500	Light	-0.8	1420	94.0	0.0	1073	94.2
		Heavy	-0.6	1535	95.2	-0.1	1218	94.0
Exponential	100	Light	-0.6	3766	93.6	-0.7	2237	94.6
		Heavy	0.3	4390	93.6	0.6	2384	94.2
	200	Light	0.8	2786	95.0	0.3	1676	94.5
		Heavy	-0.6	2985	95.2	-0.4	1939	94.2
	500	Light	-0.6	1768	94.8	-0.1	1070	94.2
		Heavy	-0.7	1866	94.8	0.0	1234	94.8

True median cost is \$10514 and \$9531 for uniform and exponential survival time, respectively.

True mean cost is \$10997 and \$10370 for uniform and exponential survival time, respectively.

Bias is relative bias in %, (median of observed median costs – true median cost)/true median cost.

Length of CI was summarized by the median length of the estimated CIs.

95% bootstrap CI was used in the computation of CP.

500 simulations and 500 bootstrap samples were used.

N denotes sample size; CI denotes 95% confidence interval; CP denotes coverage probability.



**Table 2**

Simulation results for Median- and Mean-based ICERs

<b>a. Cost scenario 1</b>									
Survival time	N per group	Censoring	Median-based ICER			Mean-based ICER			CP
			Bias (%)	Length of CI	CP	Bias (%)	Length of CI	CP	
Uniform	500	Light	2.6	6548	98.2	-0.1	2340	94.4	
		Heavy	0.5	6792	99.0	3.0	2621	95.0	
	1000	Light	-0.6	3282	94.4	1.5	1536	95.4	
		Heavy	1.1	3687	96.8	2.6	1691	96.0	
Exponential	500	Light	3.4	5377	98.5	-2.6	4622	96.6	
		Heavy	-3.6	5819	99.0	-3.5	5024	94.0	
	1000	Light	-4.0	2862	96.2	-1.3	2848	95.0	
		Heavy	-1.3	3190	97.0	-0.6	3133	95.2	

<b>b. Cost scenario 2</b>									
Survival time	N per group	Censoring	Median-based ICER			Mean-based ICER			CP
			Bias (%)	Length of CI	CP	Bias (%)	Length of CI	CP	
Normal	500	Light	-0.6	661	95.0	-0.3	461	96.5	
		Heavy	1.3	742	96.0	-0.3	530	94.4	
	1000	Light	-1.7	456	95.5	-0.9	323	95.0	
		Heavy	-0.6	535	95.2	-0.4	374	96.6	
Exponential	500	Light	-0.4	4647	96.8	1.8	2715	96.0	
		Heavy	-4.2	5981	97.0	4.0	3401	97.0	
	1000	Light	0.7	2443	96.0	1.6	1693	96.4	
		Heavy	1.7	2999	97.5	0.2	1984	95.0	

For a, true median-based ICER is 4019 and 3508 for uniform and exponential survival time, respectively.

And true mean-based ICER is 4073 and 4521 for uniform and exponential survival time, respectively.

For b, true median-based ICER is 1749 and 2230 for normal and exponential survival time, respectively.

And true mean-based ICER is 1710 and 2243 for normal and exponential survival time, respectively.

See the footnote in Table 1 for detailed simulation configurations.

N denotes sample size; CI denotes 95% confidence interval; CP denotes coverage probability; ICER denotes incremental cost-effectiveness ratio.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3**

Cost-effectiveness analysis for MADIT I and II

<b>a. MADIT I</b>				
	<b>ICD</b>	<b>Conventional</b>	<b>Difference</b>	<b>95% CI</b>
Cost (mean)	\$110,109	\$70,044	\$40,065	15,690–62,744
Cost (median)	\$111,940	\$53,794	\$58,146	20,380–92,289
Effectiveness (mean)	3.45 yrs	2.65 yrs	0.80 yr	0.44–1.18
ICER (mean)	\$50,007/yr			18,222–102,961
ICER (median)	\$72,575/yr			25,259–161,405

<b>b. MADIT II</b>				
	<b>ICD</b>	<b>Conventional</b>	<b>Difference</b>	<b>95% CI</b>
Cost (mean)	\$91,337	\$47,354	\$43,983	27,990–63,457
Cost (median)	\$67,016	\$33,118	\$33,898	18,107–49,993
Effectiveness (mean)	2.89 yrs	2.73 yrs	0.16 yr	0.04–0.31
ICER (mean)	\$266,739/yr			124,344–1,422,340
ICER (median)	\$205,575/yr			88,653–986,580

Effectiveness is the expected survival time within 4 years for MADIT I and 3.5 years for MADIT II.

Costs and survival times were discounted at 3% rate per annum.

CI of the ICER were estimated by bootstrap percentile method in MADIT I and by bootstrap reordered method in MADIT II, where 1000 bootstrap replicates were generated.

CI denotes 95% confidence interval; ICER denotes incremental cost-effectiveness ratio; ICD denotes implantable cardioverter-defibrillator; MADIT denotes Multicenter Automatic Defibrillator Implantation Trial

## Appendix

### Estimability of Mean and Median of survival time and cost

Parameter	Without censoring	With censoring without time restriction	With censoring with time restriction
Survival time			
Mean	Estimable	Estimable only if the largest survival time is uncensored. Not estimable otherwise. Remark: Due to tail problem in survival estimation, the mean estimation may not be reliable.	Estimable
Median		Estimable if survival function reaches 0.5. Not estimable otherwise.	Estimable but big step-down at the maximum time point may cause many quantiles to be tied.
Cost			
Mean	Estimable	Estimable only if the largest survival time is uncensored. Not estimable otherwise. Remark: Due to tail problem in survival estimation, it will likely be unreliable.	Estimable
Median		Same as mean cost	Estimable

All are without parametric assumptions. See Bang and Zhao (2014) for more details.