

# UC Davis

## UC Davis Previously Published Works

### Title

Cost-Effectiveness Analysis: A Proposal of New Reporting Standards in Statistical Analysis

### Permalink

<https://escholarship.org/uc/item/6nw8t0vz>

### Journal

Journal of Biopharmaceutical Statistics, 24(2)

### ISSN

1054-3406

### Authors

Bang, Heejung  
Zhao, Hongwei

### Publication Date

2014-03-04

### DOI

10.1080/10543406.2013.860157

Peer reviewed



Published in final edited form as:

*J Biopharm Stat.* 2014 ; 24(2): 443–460. doi:10.1080/10543406.2013.860157.

## Cost-Effectiveness Analysis: a proposal of new reporting standards in statistical analysis

**Heejung Bang** and

Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, CA, USA

**Hongwei Zhao**

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

Heejung Bang: hbang@ucdavis.edu; Hongwei Zhao: zhao@srph.tamhsc.edu

### Abstract

Cost-effectiveness analysis (CEA) is a method for evaluating the outcomes and costs of competing strategies designed to improve health, and has been applied to a variety of different scientific fields. Yet, there are inherent complexities in cost estimation and CEA from statistical perspectives (e.g., skewness, bi-dimensionality, and censoring). The incremental cost-effectiveness ratio that represents the additional cost per one unit of outcome gained by a new strategy has served as the most widely accepted methodology in the CEA. In this article, we call for expanded perspectives and reporting standards reflecting a more comprehensive analysis that can elucidate different aspects of available data. Specifically, we propose that mean and median-based incremental cost-effectiveness ratios and average cost-effectiveness ratios be reported together, along with relevant summary and inferential statistics as complementary measures for informed decision making.

### Keywords

average cost-effectiveness ratio (ACER); censoring; cost-effectiveness plane; incremental cost-effectiveness ratio (ICER); mean; median

## 1. Introduction

Economic burden in health care has been a significant concern of all parties involved - public, government, and industry - as our society is constantly faced with difficult decisions in allocating health care resources. The current cost consciousness in health care is a response to very large costs of some medical interventions, technologies, and regimens relative to their perceived health benefits. The strong need to contain health care costs leads us to consider which interventions produce the greatest value, based in part on economic implications [1, 2]. The importance of cost analysis is being more emphasized recently along with the contexts of Comparative Effectiveness Research, Health Care Reform and Affordable Care Act [3–6].

Cost-effectiveness analysis (CEA) is a form of economic analysis that compares the relative expenditures and outcomes of two or more strategies that perform the same task. It has been popularly used in various fields over the past several decades. Statistical methods for CEA have been developed and a measure, the incremental cost-effectiveness ratio (ICER), has been supported by various authorities and most widely adopted by researchers and policy makers [7–9]. However, there exist many controversies toward CEA/ICER despite the long

history of their usages and efforts that have been made to understand from various different angles. Many journals have written policies on the conduct and presentation of CEA in publication. This is partly due to the fact that different methods of CEA can yield results that are often counterintuitive or inconsistent, making the interpretation of the results challenging [10–15].

ICER is defined as the ratio of the difference in costs to the difference in effectiveness between two strategies. In CEA, the use of ‘arithmetic’ mean/average for cost is almost unanimously recommended because the total cost, which is important and relevant to society, can be directly estimated from the mean cost. As such, in the statistical analysis and reporting of a CEA, the mean-based ICER has been advocated thus far. However, it has been noted by many researchers that costs and cost-effectiveness ratios (CERs) are highly complicated in general with various methodological issues (see below). Cost analysis and CEA have potential to make great impacts in our lives so that such an important analysis may well deserve more than one methodology and perspective although the currently available recommendations or guidelines for statistical analyses had been formulated based on consensus statements [2, 8, 16–18]. Moreover, nearly unanimous endorsement for one method/parameter/perspective is rare in other scientific disciplines, including effectiveness research, although the necessity of some form of standardization is well understood.

In this paper, we intend to 1) propose a more comprehensive analysis for CEA using existing methods in a unified framework; 2) provide rationales for why additional measures are needed; and 3) offer a systematic guidance for statistical analyses that could be useful to researchers and help minimize common mistakes in practice. Specifically, we will summarize the challenging issues encountered in conducting CEA, and propose using ICERs based on the mean and the median and average cost-effectiveness ratio (ACER) together. We will also provide examples of how to conduct estimation and inference, and how to summarize and present the results. We believe that different analytic methods suggested in this paper are simple to implement (so that the added burden to researchers and practitioners for conducting additional analyses is not substantial) and complement one another as they fulfill different tasks and answer different questions; jointly, they would provide a more comprehensive view of evidence conveyed in available data that could lead to a more objective and enhanced decision making and a better understanding of the economic consequences.

## 2. Methods

### 2.1 Methodological issues and how to address

**Skewness: Both mean and median are important**—Skewness is a common feature of many monetary and biological data including medical cost. There are different ways to analyze skewed data in statistical and econometric literature. A simple but legitimate way to summarize skewed data is to compute the mean and median - or other parametric and nonparametric counterparts - together. The desirable properties of the mean are well documented and understood in our daily life and indeed, the concept of the mean is a foundation of modern statistical theories. For severely skewed data, the median is a simple and informative measure for central tendency and it may better represent a ‘typical’ value than the mean because the mean could be sensitive to outliers or erroneous data. For a symmetric distribution, the mean is equal to the median. Median or other quantiles-based methods have been extensively studied and used in econometrics and statistics. For instance, the median serves as a norm in housing price, mortgage, and income/salary statistics as in many national databases or surveys. Yet, it is not difficult to understand why the mean can be more meaningful in health economics and policy (See ‘Societal vs. Individual perspective’ below).

On the other hand, there are other practical circumstances. In treatment effectiveness research, there is no strong preference for the mean or the median. In the studies with the survival endpoint, the median is commonly used, regardless of the skewness of data, partly because the median is easily derived from the estimated survival function (e.g., as the time point at which the Kaplan-Meier curve reaches 0.5); in contrast, estimating the mean is not always straightforward when censoring is present [19] (See ‘Censoring’ and Table 1 below). As such, the median effectiveness is already being widely used as the standard summary measure in effectiveness studies (e.g., by FDA for drug approval). As a result, some inconsistency can be caused by different perspectives emphasized on cost in the numerator (societal perspective using the mean) and effectiveness in the denominator (individual or other perspective using the median) in CEA. In addition, when cost data are log-transformed to handle skewness/non-normality, a statistical testing based on the mean cost needs more attention. It has been reported that wrong hypothesis is frequently used in practice [20].

Thus, accepting both mean- and median-based analyses might be a reasonable way to resolve this potential inconsistency. Since the mean and median are conceptually and mathematically different and could capture different aspects of the distribution, the corresponding ICERs are not competing as the mean and median themselves are not competing. Even in the current practice where mean-based analysis is a preferred or official strategy for policy decision, median-based analysis could play an important role as a part of sensitivity or secondary analyses [21–24]. In that regard, we disagree with the authors who stated “Standard non-parametric methods (for example, Mann-Whitney U-test) and analyses of transformed costs are generally inappropriate because they are not focused on arithmetic means.” [25].

**Censoring: It should be adjusted and it could affect the parameter of primary interest**—Censoring arises commonly in prospective studies, where endpoints are not observed for all subjects. If censoring does not occur so that complete endpoint data are obtained for all subjects, then standard statistical methods (e.g., sample mean, median, t-test, Wilcoxon-Mann-Whitney test) could be valid analytic tools. However, if the effectiveness measure and/or cost are censored, specialized methods that properly account for censoring should be used. In the analysis of the effectiveness, censoring is adequately handled in the current practice – for example, using the Kaplan-Meier estimator, log-rank test and Cox model instead of sample mean/median, t-test and linear/logistic regression in the analysis of survival data.

If effectiveness data are censored, the associated cost data are likely to be censored as well, which implies that censoring should be accounted for in cost analysis and CEA, therefore, in both the numerator and denominator in CERs [26–29]. Although not so intuitive, it is important to understand that the underlying mechanism of censoring and how to address this issue in statistical methods differ for effectiveness and cost data. Even though it is often reasonable to assume censoring is noninformative for survival data, it is almost always informative for cost data; see references [30–32] and Table 1 for these issues. Various censored cost estimators have been published over the last decade and they are increasingly used [33–37]. It is worth noting that the quality-adjusted life year (QALY), which is commonly considered as an effectiveness measure in the ICER, shares a similar censoring mechanism to cost [38].

Here, we will revisit some fundamentals in survival data analysis that are relevant to CEA but less understood by many practitioners. In the absence of censoring, the distributions of effectiveness and lifetime cost - say, from diagnosis or study entry to death - and a function thereof (e.g., mean or median) can be easily estimated. In contrast, when censoring is present, estimating these parameters is not a simple task. Oftentimes, we should compromise

our goal depending on the data availability. Moreover, the parameter for statistical estimation could be altered depending on what the observed data permit – see Table 1 for the estimability of fundamental parameters in effectiveness and cost analyses and CEA.

Let us discuss effectiveness data first. In theory, the mean survival time can be estimated as long as the longest survival time is not censored, while the median survival time can be estimated as long as the survival probability (e.g., estimated from the Kaplan-Meier curve) at the largest observation time reaches 0.5 or below. However, we need to be aware of that generally the survival curve cannot be estimated very well in the tail area, due to the small sample size. These imply that the mean and median survival times are not always estimable, and even if they are, they might not be estimated accurately or reliably. One might artificially force that the largest survival time be uncensored and accept that the true mean is underestimated by this remedy – indeed, this remedy is chosen by many statistical packages (e.g., SAS) in order to produce the Kaplan-Meier curve. Another strategy is to estimate the *restricted* means, restricting the longest observations to some time limit, where a reasonable amount of data is still available; see Table 1.

This issue can be more complicated for cost. Nonparametrically, when the largest observation is censored, which happens frequently for studies with limited follow-up time, the (marginal) distribution of cost is nowhere identifiable for any time point beyond zero. Among ways to tackle this issue, time restriction is adopted most commonly in practice [30, 39]. Since cost estimation becomes harder and easily unstable near the end of follow-up, it is not uncommon to find a study in which 5 year data are used in the effectiveness analysis, whereas 3 or 4 year data are used in the cost analysis or CEA [40, 41]. Therefore, it is important to understand that in many situations, the best we can estimate is a time-restricted mean or some quantiles, while estimating the overall mean or median of lifetime cost could be an impossible goal using observed data without introducing additional assumptions.

**Bi-dimensionality: CI and CE plane should be constructed and interpreted correctly**—ICER is a ratio statistic with the difference in the numerator as well as in the denominator. Thus, it is essential to have a ‘2-dimensional’ thinking/plot, and cost-effectiveness (CE) plane offers a visual representation of the joint distribution of cost and effectiveness data as an exploratory and inferential tool [42]. Due to this bi-dimensionality, the confidence interval (CI) of the ICER is not guaranteed to be a ‘closed’ interval that we normally see (see Table 2 and Figure 1 as example). Various methods are available on how to construct a CI for the ICER - see [27, 43–47] among others. The bootstrap method showed some distinct advantages as long as computational burden is not too severe. It can be uniformly applied for mean or median-based methods and different CERs, and particularly useful to create a CI on the CE plane. Yet, despite a number of publications on the related topics, we still think that there is limited systematic guidance on how to construct the CI or conduct statistical inference for the ICERs that cover all possible scenarios in a unified manner. Some – if not many – researchers seem to naively use the standard bootstrap method (e.g., percentile-based method) that always generates a closed interval regardless of qualitatively different underlying scenarios, or they restrict attentions to simplistic scenarios (e.g., NE quadrant of the CE plane).

The CE plane can be formulated with the effectiveness difference as x-axis and the cost difference as y-axis, where the ICER represents a slope in it [42]. Different quadrants represent different preferences for a new vs. standard treatment [46, 48, 49] – readers may refer to ‘5 Decision Regions’ described in Appendix A that could aid interpretation [46]. Since the bootstrap samples can occupy anywhere between 1 to 4 quadrants (NE, NW, SE, SW) in the CE plane, the determination of which bootstrap method should be used depends on where bootstrap samples lie. For example, for ‘one quadrant only’ situations or some

‘two quadrants’ (i.e., ‘NE/SE’ or ‘NW/SW’) situations where the ordering of the data is natural, standard (e.g., percentile-based) methods can be used; but for other situations where the ordering of the data could be unnatural as it can be interrupted by discontinuity (ICER = + or – infinity) [44–47, 50], the use of standard methods could be erroneous so that different methods such as reordered or angle/wedge-based method should be used depending on the scenarios - systematic guidance is provided in a unified framework [51]. In Appendix B, we explained that construction of the cost-effectiveness acceptability curve (CEAC), another CE method, also needs bi-dimensional thinking and similar caution [52].

The International Society for Pharmacoeconomics and Outcomes Research Task Force recommended CEA be performed even when clinical effectiveness fails to be demonstrated [18]. We agree with this recommendation but this situation would need more care in the analysis as we may face a mathematical scenario of ‘division by 0’ for ICER so that standard methods may not work as we briefly explained above. In this situation, a general consensus in the literature is that we may compare costs directly rather than compute ICER. Yet, how to determine whether the denominator is equal or close to zero is not that straightforward, and can be subjective although statistical and/or clinical significance could help. A paired tool of ‘CE plane with properly selected statistical methods for ICERs could be particularly useful as they can be *systematically* applied to any mathematical scenario encountered in practice [47, 51].

**Average vs. Incremental: We should think both ways—**Controversy between ACER and ICER has a long history [53–56] and it is well summarized in a recent review paper [55]. Incremental and marginal gain is a critical concept in economic decision making and this concept provides a theoretical basis and strong support for the role of ICER in CEA [1]. However, it is also well understood that great care is needed in the use of ICER. First, ICER estimates can be highly variable or numerically unstable when the denominator, health benefit, is small, which is quite common in real life examples – when the denominator is near 0, ICER is virtually estimating + or – infinity. Second, for ICER-based methods, such as incremental net benefit or CEAC [1, 49, 57], the upper limit of what society is willing to pay for an additional unit of health benefit, conventionally denoted by  $\lambda$ , is needed for decision making, but this value is inherently subjective and depends on the effectiveness measure, disease, society/country, and year although some suggestions are available [15, 58–60].

In contrast, ACER has several conceptual, statistical and numerical advantages that could complement the role and capacity of ICER [54, 61–63]. Although interest could lie foremost in ICER, ACER still could provide additional useful information because many people (e.g., patients and clinicians) may want to know how much money is expected to be spent on average per benefit (e.g., per year until death), with or without consideration of comparator(s). Therefore, we argue that we need to think incremental as well as average in a more complete analysis of given data and in the relevant decision making processes.

**Societal vs. Individual perspective: Both are important—**The importance of the societal perspective in medical decision making under limited resources has been repeatedly emphasized [2]. The fact that the mean cost serves as the primary statistical parameter can be understood in this context as the total cost can be directly derived from the mean but not from alternative measures [20]. Society must pay all costs incurred including a small proportion of implausibly large ones and the total cost should be basis for health care policy decision [8, 17]. On the other hand, the individual perspectives may not be ignorable because treatment decisions considering benefit and cost are also made on individual levels, say, by patients and health care providers on the daily basis. The median might be an important parameter that may be more relevant to consumers than providers, more to



patients and practicing physicians than policy makers, and more on the individual level than the societal level. Moreover, we should be aware of potential dangers of outliers and errors (e.g., administrative or coding errors) in cost databases, and the importance of robust procedures in cost analysis. These views may further justify the use of the mean and the median together in CEA as a partial solution.

Along the line, a more empirical, objective CEA using observed data only (e.g., *actual* costs incurred for the numerator and survival time for the denominator) may be warranted in parallel with the CEA based on the set of ‘gold standard’ recommendations that reflect the societal perspective (e.g., comprehensive cost estimation covering opportunistic, indirect, and/or potential cost for the numerator and QALY for the denominator) [2, 8].

## 2.2 Proposal of a more comprehensive analysis: Beyond a single measure

In the current practice of CEA, most researchers compute and report the mean-based ICER (let us denote by  $ICER_{mean}$ ) only, where the associated CI and/or CE plane are often accompanied. We propose to expand this current practice to further include the median-based ICER (denoted by  $ICER_{median}$ ) [51] and ACER in the primary CEA. Mathematically, these 3 measures are defined as:

$$ICER_{mean} = \frac{mean(M_1) - mean(M_2)}{mean(E_1) - mean(E_2)}$$

$$ICER_{median} = \frac{median(M_1) - median(M_2)}{median(E_1) - median(E_2)}$$

$$ACER_i = \frac{mean(M_i)}{mean(E_i)}$$

where  $M_i$  and  $E_i$  denote cost and effectiveness measures, respectively, for the  $i$ -th group ( $i=1,2$ ) and  $mean(X)$  and  $median(X)$  denote the mean and median of variable  $X$ , respectively. [Remark: For simple presentation, we will use parameters and estimates interchangeably.] The difference of ACERs =  $ACER_1 - ACER_2$  between two treatments is often contrasted to  $ICER_{mean}$ .

ICER could be interpreted as the additional cost incurred per additional health outcome, where a difference or increment can be measured by either the mean or median, and ACER as the net cost incurred per a unit of health outcome. Naturally,  $ICER_{mean}$  and  $ICER_{median}$  could yield very different values. When this happens with life year as health outcome, we might interpret as: it costs \$y1 to save one year of life from the societal point of view, however, it is more likely to cost \$y2 to have one more year of life from the payer’s point of view.

Naturally, the mean and median estimate different population quantities, and similarly for ICER and ACER. Thus, they are not competing measures but rather they may be regarded as complementary measures that could address different questions. We propose reporting cost, effectiveness, ACER and ICERs (using mean and median) in one table – a sample table is provided in Table 2, to be discussed below. We generally recommend reporting the differences and CIs for pertinent measures in the same table, along with the CE plane, whenever possible and justifiable - possibly in supplemental document if space is an issue in publication.

## 3. Examples

In this section, we reanalyze data from two published trials using the proposed methods and provide a sample table for how the results can be summarized and reported. We also illustrate the impact of ignoring censoring in the CEA.

### 3.1 Schizophrenia trial

An international trial of pharmacotherapy of olanzapine vs. haloperidol for treating schizophrenia was conducted among patients who met Diagnostic and Statistical Manual III criteria for schizophrenia-related disorders [64]. A total of 1,996 patients entered the trial at 174 sites across 17 countries during 1993–1995. The standard CEA was conducted based on  $ICER_{mean}$  and published previously [65]. As the original investigators did, the same 812 patients (548 for olanzapine and 264 for haloperidol) who had both cost and effectiveness data from at least their first post-baseline visit from US centers were included. All costs were calculated from pricing algorithms based on 1996 standard price lists for drugs and units of medical services, and effectiveness was defined as responder days – refer to references [51, 64, 65] for more details on trial and cost-related information.

Based on the standard analysis using the means, olanzapine was both less costly (\$27,765 vs. \$38,066) and more effective (188 vs. 169 days) than haloperidol. However, when we computed the medians, the treatment effect was similar (200 vs. 182 days) but olanzapine was more costly than haloperidol (\$7,790 vs. \$6,486), i.e., the order is reversed (Table 2). In addition, the mean and median costs were highly different, which implies that the cost distribution is severely skewed. Note also that the 95% CI of the mean cost difference did not include 0, but for the median cost the 95% CI of the difference included 0, while the 95% CI of the difference in effectiveness measure included 0 either by the mean or median.

The resulting  $ICER_{mean}$  and  $ICER_{median}$  were computed as  $-\$563/\text{day}$  and  $\$75/\text{day}$ , respectively, i.e., the signs are different. We constructed a CE plane with 1000 bootstrap samples. Since the bootstrap ICER samples lie in 3 or 4 quadrants, we used the angle/wedge method for computing CIs [46, 51]. In this scenario, standard bootstrap methods that provide closed intervals should not be used. The inner angle formed by the two slopes of  $-174$  and  $2,110$  in the SE and SW quadrants contains 95% of the bootstrap samples of  $ICER_{mean}$  as shown in Figure 1. Numerically, we can write 95% CI as  $(-\infty, -174)$  in SE and  $(2110, \infty)$  in SW, which may look strange. However, when they are overlaid in the CE plane, conjoined open intervals may not be surprising any more. Figure 1 strongly supports the dominance of olanzapine as almost all bootstrap samples lie in favorable or highly favorable regions (Readers may interpret the results with reference to the 5 CE regions in Appendix A.).

We repeated the same analysis using  $ICER_{median}$ . This time, we found the outer angle formed by the two slopes of  $-152$  and minus infinity in the SE region encompasses 95% of the bootstrap samples. This ‘around the circle’ situation may indicate that this scenario mirrors the 0/0 situation mathematically, and non-significant differences in cost as well as in effectiveness seem to support this claim. Indeed, the CE plane exhibits the cost difference is consistently around 0, while the effectiveness difference is more likely to be positive. In this situation, we may choose to postpone the decision or draw CEACs to estimate the probability that one treatment is more cost-effective than the other (as y-axis) over different willingness-to-pay thresholds (in x-axis) as in Figure 2. In Appendix B, we explained how to construct CEAC when data are on more than one quadrant. The preference of olanzapine was apparent in Figure 2 for most meaningful values of willingness-to-pay, however, mean and median-based analyses resulted in highly different curves.

Next, we computed the ACER for each treatment. The ACER was estimated as  $\$148/\text{day}$  for olanzapine and  $\$225/\text{day}$  for haloperidol, and the difference was  $-\$77/\text{day}$  (95% CI:  $-133, -20$ ), which indicates higher cost per day for haloperidol, and again olanzapine shows cost advantage.



In this example, we may conclude that olanzapine seems to be more cost-effective based on the estimated ACER and ICERs and the associated CEACs. Yet, the fact of reversed signs in the cost difference as well as ICERs using the mean vs. median can make the interpretations of the cost analysis and CEA challenging and controversial, thus more discussions are warranted from clinical, economic and statistical perspectives. In such situations, either mean or median-based analysis alone could be limited or even misleading, so both analyses should be presented for readers and it may be reasonable to wait for further studies or evidence.

### 3.2 MADIT

To illustrate the importance of a proper treatment of censoring in the cost analysis and, potentially in the CEA, 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 [66]. A total of 181 patients were enrolled from 36 centers with 89 patients assigned to the ICD arm and 92 patients assigned to the conventional intervention arm. 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 significantly improved survival, compared to the conventional intervention [66]. Cost data were collected for patients recruited from centers in the US and all relevant medical costs incurred during the study were recorded, as described previously [40].

As in the original CEA, we also restricted the duration of the cost estimation to 4 years. The data were heavily censored; 70% of subjects were censored in the ICD arm and 48% of subjects were censored in the conventional arm. Both costs and survival times were discounted at 3% annual rate [2, 40]. For statistical illustration, we analyzed the mean cost in both arms without and with accounting for censoring, and results are reported in Table 3. We estimated the mean costs using the Zhao-Tian estimator that yields asymptotically unbiased estimates and handles heavily censored data well [27, 34] – the estimated mean was \$99,548 for ICD vs. \$72,754 for conventional arm. To compare, when we naively computed the mean from the full sample, underestimation was apparent as the costs after censoring were not included in calculations. When we computed the mean among uncensored/complete data only, the means are destined to be biased toward the subjects with shorter survival times and variability/uncertainty measures tend to increase due to reduced sample size [31]. As such, the standard error in addition to the mean/bias could be meaningfully changed.

## 4. Discussion

In this article, we presented the three analytic approaches ( $ICER_{mean}$ ,  $ICER_{median}$  and ACER) jointly in a unified framework, where  $ICER_{mean}$  has been regarded as a gold standard measure in the CEA field, while other measures did not receive enough attention. We demonstrated these approaches could yield highly different results that could lead to different conclusions. This is not surprising because the mean and median estimate different population quantities and so do different ratio statistics. We do not think different results obtained from the same dataset are undesirable. Different results may mean additional new information rather than confusion or inconsistency. Indeed, different results from multiple methods could be more useful than one result from one method, and an uninformative answer could be better than a misleading one [67]. Keep also in mind that the CERs were originated to aid in decision making and not in themselves make the decision [16, 56].

Nowadays, researchers are increasingly interested in examining costs of care, and administrative databases have made relevant data available. Therefore, correct analysis and interpretation are strongly called for, and key methodological issues commonly encountered in cost data should be properly addressed. The traditional paradigm that could reveal limited aspects of the data may need to be expanded, improved and/or evolved. Indeed, accepting more than one method and perspective would be an important step in evidence-based medicine and may make CEA more frequently used in the real world settings [13].

Reading this article, readers should note the following points. First, the proposed methods may be better suited for randomized controlled trials, rather than observational studies. For the latter, issues inherent in observational studies (e.g., confounders, selection bias) should be addressed as well. Although cost data from observational studies would capture more naturalistic settings, statistical methods are anticipated to be more complicated, similarly to what is known in effectiveness research [68]. Second, it is easily noticeable that we did not invent or add a new method in this proposal. Instead, we aimed to use existing fundamental methods and concepts jointly, which could provide a more complete picture of a given dataset, while still making it easy to implement. Yet, advanced or newer methods and concepts could be useful for some situations [28, 69–73]. Third, the proposed methods are compatible with empirical, patient-level data, minimizing the need for assumptions or modeling efforts. In different situations (e.g., when empirical data are not fully available or long term extrapolation is aimed), fundamentally different approaches are generally entailed [74]. Fourth, more discussion is needed on the interpretation and statistical properties of the median-based ICER as a relatively new measure.

In practice, more discussions would be definitely called for regarding how to best interpret results, synthesize the evidence obtained from the proposed analytic strategy, and guide the ultimate decision. While waiting for the consensus, our advice is that 1) if all of 3 analyses (i.e.,  $ICER_{mean}$ ,  $ICER_{median}$  and ACER) yield qualitatively similar results, say, having the same direction, we are more confident about adopting a new treatment; 2) if they conflict each other, it may imply different aspects of the data, and decision making may not be as simple. But we will have a clearer picture of treatment implications. For example, if there are a few patients who incurred huge costs, which drive the mean cost up for a treatment, whereas the majority of patients enjoyed lower costs, then this treatment may still be considered [75]. Finally, it is important to acknowledge all these analyses are from one dataset/study. Cumulative or total evidence in similar and different settings should be emphasized not only in effectiveness research but also in CE research although methods for meta-analysis or systematic review suited for cost analysis and CEA are highly limited currently.

## Acknowledgments

**Role of the Funding Source:** This research was supported by R01 HL096575 from the National Heart, Lung and Blood Institute.

We want to thank Dr. Robert Obenchain and Dr. Alvin Mushlin for providing de-identified data. We also appreciate Ms. Ya-Lin Chiu's advice in programming.

## References

1. Drummond, MF.; Sculpher, MJ.; Torrance, GW.; O'Brien, BJ.; Stoddart, GL. *Methods for the Economic Evaluation of Health Care Programmes*. 3. Oxford University Press; Oxford: 2005.
2. Gold, MR.; Sigel, JE.; Russell, LB.; Weinstein, MC. *Cost-effectiveness in Health and Medicine*. Oxford University Press; New York: 1996.

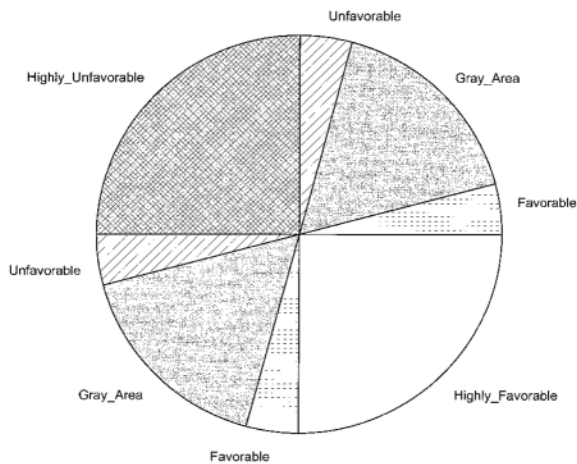
3. Garber AM, Harold CS. The role of costs in comparative effectiveness research. *Health Affair*. 2010; 29:1805–1811.
4. Owens DK, Qaseem A, Chou R, Shekelle P. for the Clinical Guidelines Committee of the American College of Physicians. High-value, cost-conscious health care: concepts for clinicians to evaluate the benefits, harms, and costs of medical interventions. *Annals of Internal Medicine*. 2011; 154:174–180. [PubMed: 21282697]
5. Gruber J. The cost implications of health care reform. *NEJM*. 2010; 362:2050–2051. [PubMed: 20463330]
6. Swain S, Hudis C. Health policy: Upholding the Affordable Care Act-implications for oncology. *Nat Rev Clin Oncol*. 2012 Epub ahead of print.
7. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practice. *NEJM*. 1977; 296:716–721. [PubMed: 402576]
8. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. for the Panel on Cost-Effectiveness in Health and Medicine. Recommendations for reporting cost-effectiveness analyses. *JAMA*. 1996; 276:1253–1258. [PubMed: 8849754]
9. Rawlins MD. NICE and the public health. *British Journal of Clinical Pharmacology*. 2004; 58:575–580. [PubMed: 15563355]
10. Walker D. How to do (or not to do). Cost and cost-effectiveness guidelines: which ones to use? *Health Policy and Planning*. 2001; 16:113–121. [PubMed: 11238438]
11. Eddy DM. Oregon’s methods: did cost-effectiveness analysis fail? *JAMA*. 1991; 265:2135–2141. [PubMed: 1920704]
12. Udvarhelyi S, Colditz GA, Rai A, Epstein AM. Cost-effectiveness and cost-benefit analyses in the medical literature: are the methods being used correctly? *Annals of Internal Medicine*. 1992; 116:238–244. [PubMed: 1530808]
13. Neumann PJ. Why don’t Americans use cost-effectiveness analysis? *The American Journal of Managed Care*. 2004; 10:308–312. [PubMed: 15152700]
14. Birch S, Gafni A. Information created to evade reality (ICER): things we should not look to for answers. *Pharmacoeconomics*. 2006; 62:1121–1131. [PubMed: 17067196]
15. Gafni A, Birch S. Incremental cost-effectiveness ratios (ICERs): the silence of the lambda. *Social Science and Medicine*. 2006; 62:2091–2100. [PubMed: 16325975]
16. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. for the Panel on Cost-Effectiveness in Health and Medicine. The role of cost-effectiveness analysis in health and medicine. *JAMA*. 1996; 276:1172–1177. [PubMed: 8827972]
17. Siegel JE, Weinstein MC, Russell LB, Gold MR. for the Panel on Cost-Effectiveness in Health and Medicine. Recommendations for reporting cost-effectiveness analyses. *JAMA*. 1996; 276:1339–1341. [PubMed: 8861994]
18. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, Cook J, Glick H, Liljas B, Petitti D, Reed S. 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]
19. Brookmeyer R. Median survival time: confidence intervals and tests. *Encyclopedia of Biostatistics*. 1998; 4:2538–2540.
20. Zhou XH, Melfi CA, Hui SL. Methods for comparison of cost data. *Annals of Internal Medicine*. 1997; 127:752–756. [PubMed: 9382393]
21. Jain R, Grabner M, Onukwuga E. Sensitivity analysis in cost-effectiveness studies: from guidelines to practice. *Pharmacoeconomics*. 2011; 29:297–314. [PubMed: 21395350]
22. Rossner M, Van Epps H, Hill E. Show me the data. *Journal of Cell Biology*. 2007; 179:1091–1092. [PubMed: 18086910]
23. Bishai D, Colchero A, Durack DT. The cost-effectiveness of antiretroviral treatment strategies in resource-limited settings. *AIDS*. 2007; 21:1333–1340. [PubMed: 17545710]
24. Nadler E, Broderick WC, Zarotsky V, Kim J. How do medical and pharmacy directors perceive the value of new cancer drugs? *Drug Benefit Trends*. 2009; 21:120–130.
25. 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]

26. 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]
27. Zhao H, Tian L. On estimating medical cost and incremental cost-effectiveness ratios with censored data. *Biometrics*. 2001; 57:1002–1008. [PubMed: 11764238]
28. Fenwick E, Marshall DA, Blackhouse G, Vidaillet H, Slee A, Shermanski L, Levy AR. Assessing the impact of censoring of costs and effects on health-care decision-making: an example using the Atrial Fibrillation Follow-up Investigation of Rhythm Management (AFFIRM) study. *Value in health*. 2008; 11:365–375. [PubMed: 17854433]
29. Wang H, Zhao H. Estimating incremental cost-effectiveness ratios and their confidence intervals with differentially censored data. *Biometrics*. 2006; 62:570–575. [PubMed: 16918922]
30. Huang Y. Cost analysis with censored data. *Medical Care*. 2009; 47:S115–119. [PubMed: 19536024]
31. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics*. 1997; 53:419–434. [PubMed: 9192444]
32. Zhao H, Cheng Y, Bang H. Some insight on censored cost estimators. *Statistics in medicine*. 2011; 30:2381–2388. [PubMed: 21748774]
33. Young TA. Estimating mean total costs in the presence of censoring: a comparative assessment of methods. *Pharmacoeconomics*. 2005; 23:1229–1242. [PubMed: 16336017]
34. 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]
35. Zhao H, Zuo C, Chen S, Bang H. Nonparametric inference for median costs with censored data. *Biometrics*. 2012 Epub ahead of print.
36. Nietert PJ, Wahlquist AE, Herbert TL. Characteristics of recent biostatistical methods adopted by researchers publishing in general/internal medicine journals. *Statistics in medicine*. 2012 Epub ahead of print.
37. Ye X, Henk HJ. An introduction to recently developed methods for analyzing censored cost data. *ISPOR Connections: uniting science and practice*. 2007; 13:11–13.
38. Zhao H, Tsiatis AA. A consistent estimator for the distribution of quality adjusted survival time. *Biometrika*. 1997; 84:339–348.
39. Bang H. Medical cost analysis: application to colorectal cancer data from the SEER Medicare database. *Contemporary clinical trials*. 2005; 26:586–597. [PubMed: 16084777]
40. 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]
41. Zwanziger J, Hall WJ, Dick AW, Zhao H, Mushlin AI, Hahn RM, Wang H, Andrews M, Mooney C, Wang H, Moss AJ. The cost effectiveness of implantable cardioverter-defibrillators: results from the MADIT-II. *JACC*. 2006; 47:2310–2318. [PubMed: 16750701]
42. Black WC. The CE plane: a graphic representation of cost-effectiveness. *Medical decision making*. 1990; 10:212–214. [PubMed: 2115096]
43. 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.
44. Heitjan DF, Moskowitz AJ, Whang W. Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Economics*. 1999; 8:191–201. [PubMed: 10348414]
45. Heitjan DF, Moskowitz AJ, Whang W. Problems with interval estimates of the incremental cost-effectiveness ratio. *Medical decision making*. 1999; 19:9–15. [PubMed: 9917015]
46. Obenchain RL. Resampling and multiplicity in cost-effectiveness inference. *Journal of Biopharmaceutical Statistics*. 1999; 9:563–582. [PubMed: 10576404]
47. Glick, H.; Doshi, J.; Sonnad, S.; Polsky, D. *Economic Evaluation in Clinical Trials*. Oxford University Press; New York: 2007.
48. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Economics*. 2002; 11:415–430. [PubMed: 12112491]

49. Willan, AR.; Briggs, AH. *The Statistical Analysis of Cost-effectiveness Data*. Wiley; Chichester, UK: 2006.
50. Wang H, Zhao H. A study on confidence intervals for incremental cost-effectiveness ratios. *Biometrical journal*. 2008; 50:505–514. [PubMed: 18663759]
51. Bang H, Zhao H. Median-based incremental cost-effectiveness ratio (ICER). *Journal of Statistical Theory and Practice*. 2012; 6:428–442. [PubMed: 23441022]
52. Fenwick E, Claxton K, Sculpher MJ. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Economics*. 2001; 10:779–787. [PubMed: 11747057]
53. Briggs AH, Fenn P. Trying to do better than average: a commentary on ‘statistical inference for cost-effectiveness ratio’. *Health Economics*. 1997; 6:491–495. [PubMed: 9353649]
54. Laska EM, Meisner M, Siegel C. The usefulness of average cost-effectiveness ratios. *Health Economics*. 1997; 6:497–504. [PubMed: 9353650]
55. Hoch JS, Dewa CS. A clinician’s guide to correct cost-effectiveness analysis: think incremental not average. *The Canadian Journal of Psychiatry*. 2008; 53:267–274.
56. 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.
57. Fenwick E, Marshall DA, Levy AR, Nichol G. Using and interpreting cost-effectiveness acceptability curves: an example using data from a trial of management strategies for atrial fibrillation. *BMC Health Services Research*. 2006; 6:52. [PubMed: 16623946]
58. Hutubessy, R.; Baltussen, R.; Torres-Edejer, TT.; Evans, DB. *WHO-CHOICE: choosing interventions that are cost-effective*. WHO Editions; Geneva, Switzerland: 2003.
59. Braithwaite RS, Meltzer DO, King JTT, Leslie D, Roberts MS. What does the value of modern medicine say about the \$50,000 per quality-adjusted life-year decision rule? *Medical Care*. 2008; 46:349–356. [PubMed: 18362813]
60. Ubel PA, Hirth RA, Chernew ME, Fendrick AM. 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]
61. Laska EM, Meisner M, Siegel C. Statistical inference for cost-effectiveness ratios. *Health Economics*. 1997; 6:229–242. [PubMed: 9226141]
62. Bang H, Zhao H. Average cost-effectiveness ratio with censored data. *Journal of Biopharmaceutical Statistics*. 2012; 22:401–415. [PubMed: 22251182]
63. Olchowski AE, Foster EM, Webster-Stratton CH. Implementing behavioral intervention components in a cost-effective manner: Analysis of the incredible years program. *Journal of Early and Intensive Behavior Intervention*. 2007; 3.4–4.1:284–304.
64. Tollefson GD, Beasley CM, Tran PV, Street JS, Krueger JA, Tamura RN, Graffeo KA, Thieme ME. Olanzapine versus haloperidol in the treatment of schizophrenia and schizoaffective and schizophreniform disorders: results of an international collaborative trial. *American Journal of Psychiatry*. 1997; 154:457–465. [PubMed: 9090331]
65. Obenchain RL, Johnstone BM. Mixed-model imputation of cost data for early discontinuers from a randomized clinical trial. *Drug information journal*. 1999; 33:191–209.
66. 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. *NEJM*. 1996; 335:1933–1940. [PubMed: 8960472]
67. Jiang G, Wu J, Williams GR. Fieller’s interval and the Bootstrap-Fieller interval for the incremental cost-effectiveness ratio. *Health Serv Outcomes Res Method*. 2000; 1:291–303.
68. Faries, DE.; Leon, AC.; Haro, JM.; Obenchain, RL. *Analysis of Observational Health-Care Data Using SAS*. SAS Press Series; Cary, NC: 2010.
69. Eckermann S, Willan A. Expected value of information and decision making in HTA. *Health Economics*. 2007; 16:195–209. [PubMed: 16981193]
70. Obenchain RL. ICE preference maps; nonlinear generalizations of net benefit and acceptability. *Health Serv Outcomes Res Method*. 2008; 8:31–56.

71. Ginnelly L, Claxton K, Sculpher MJ, Golder S. Using value of information analysis to inform publicly funded research priorities. *Appl Health Econ Health Policy*. 2005; 4:37–46. [PubMed: 16076237]
72. Severens J, Brunenberg D, Fenwick E, O’Brien B, Manuela J. Cost-effectiveness acceptability curves and a reluctance to lose. *Pharmacoeconomics*. 2005; 23:1207–1214. [PubMed: 16336015]
73. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics*. 1999; 18:341–364. [PubMed: 10537899]
74. Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. *BMJ*. 2011:342.
75. DeNavas-Walt, C.; Proctor, BD.; Smith, JC. US Census Bureau, Current Population Reports. 2011. Income, poverty, and health insurance coverage in the United States: 2010.

**Appendix A. The five regions in the CE plane for decision making**



This figure was reproduced from Obenchain (1999)

**Appendix B. How to construct CEAC when data are on more than 1 quadrant**

If a point lies in the SE quadrant, it means the treatment is preferred; similarly, if a point lies in the NW quadrant, it means that the treatment is not preferred. It is more complicated when the points lie in the NE or the SW quadrant. We use an example to illustrate the method with proper interpretations.

**In NE quadrant**

Point A, ICER=10, meaning that treatment is preferred most likely, since ICER is very low.

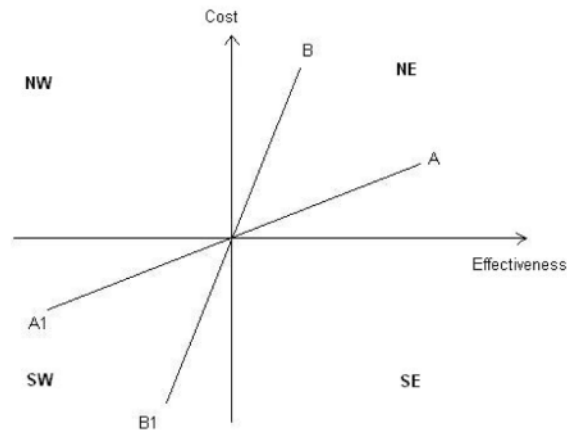
Point B, ICER=1000, meaning that treatment is probably not preferred, since ICER is too high.

**In SW quadrant**

Point A1, ICER=10, meaning that treatment is not preferred, since control is more cost-effective.

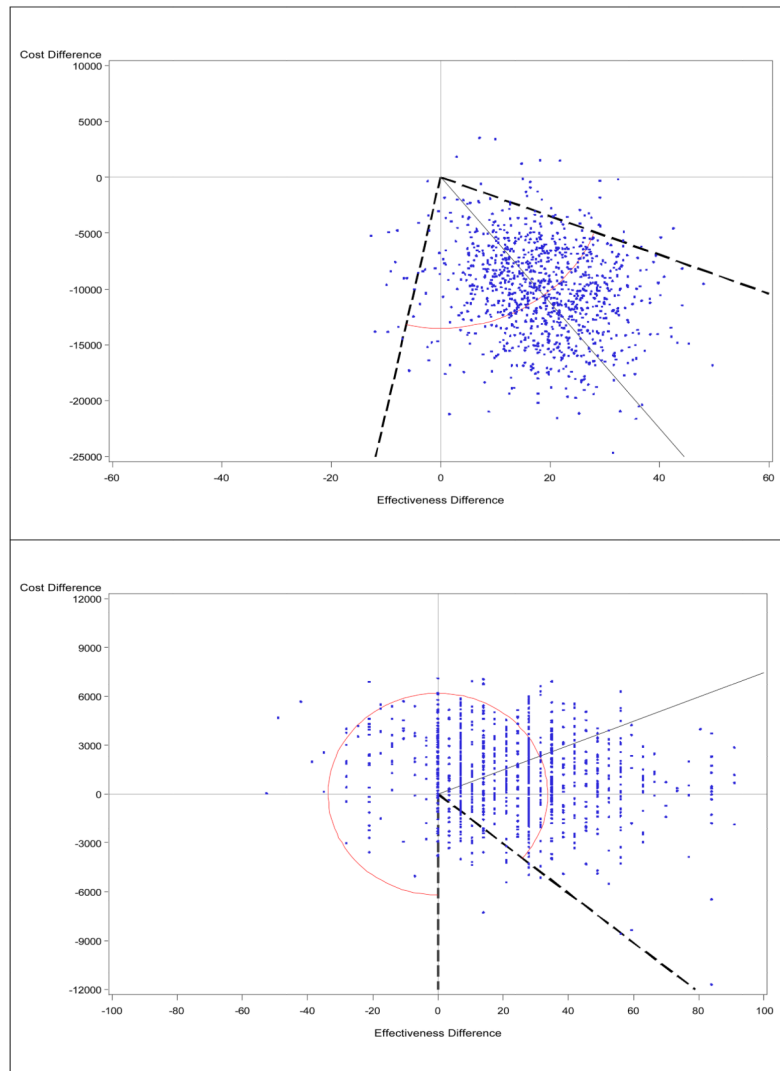


Point B1, ICER=1000, meaning that treatment is preferred, since control is too costly compared to its effectiveness.



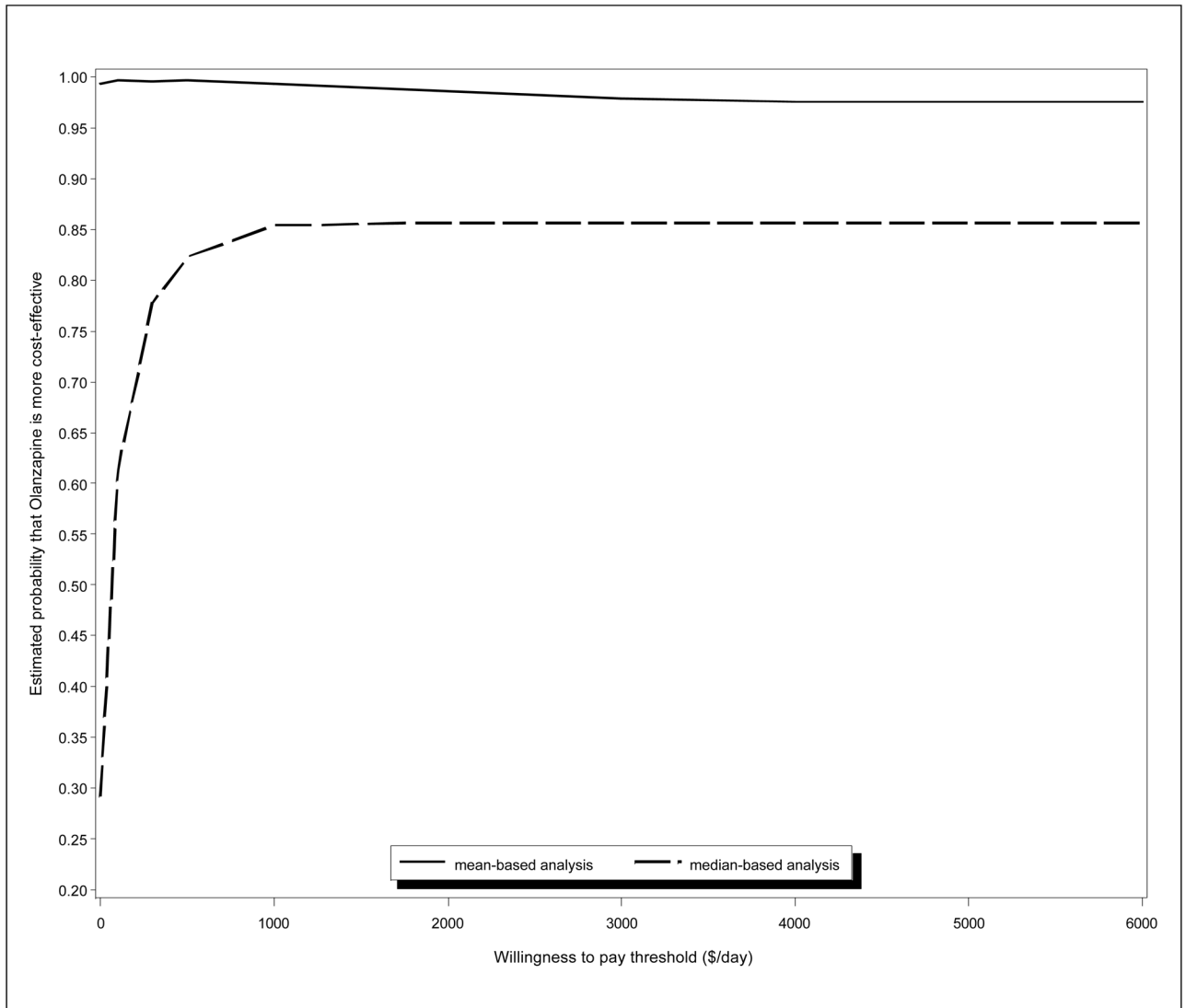
When we plot a CEAC using bootstrap method, we need to count the frequency that a resample appears below the line whose slope is  $\lambda$ . In the above example,

1.  $0 < \lambda < 10$ ,
  - A and B are not included
  - A1 and B1 are included (since control is not chosen)
2.  $10 < \lambda < 1000$ 
  - A and B1 are included
  - A1 and B are not included
3.  $\lambda > 1000$ 
  - A and B are included
  - A1 and B1 are not included



**Figure 1. Cost-effectiveness plane using mean (upper) vs. median (lower): Schizophrenia trial example**

Solid line indicates the point estimate and broken lines indicate a 95% confidence interval for the ICER (connected by arc). Note that the scale is different in Y-axis. This figure was reproduced from Bang and Zhao (2012).



**Figure 2.** Cost-effectiveness acceptability curves: Schizophrenia trial example

**Table 1**

Estimability of the mean and median of survival time and cost\*

Parameter	Without Censoring	With Censoring without time restriction**	With Censoring with time restriction
<b>Survival time</b>			
Mean	Estimable	Estimable only if the largest survival time is uncensored. Not estimable otherwise. The latter situation is much more common. Remark: Due to tail problem in survival estimation (see below), the mean estimation, although can be done, may not be reliable.	Estimable
Median	Estimable	Estimable if survival function (e.g., Kaplan-Meier curve) reaches 0.5 or below. Not estimable otherwise.	Estimable but big step-down at the maximum time point (i.e., time limit) may cause many quantiles to be tied (e.g., 75%tile= median and/or median=25%tile can happen)
Distribution	Estimable	Estimable up to the largest uncensored survival time. However, estimation in tail area tends to be unstable, meaning it has large variability.	Same as median survival time
<b>Cost</b>			
Mean	Estimable	Estimable only if the largest survival time is uncensored. Not estimable otherwise. The latter situation is much more common. Remark: Due to tail problem in survival estimation, it will likely be unreliable.	Estimable
Median	Estimable	Same as mean cost	Estimable
Distribution	Estimable	Same as mean cost	Estimable

\* All entries are for nonparametric estimation. With parametric assumptions, all can be estimable but this approach basically needs extrapolation and additional assumptions for unobserved data.

\*\* Without time restriction, survival and cost means survival time and lifetime cost. With time restriction, survival and cost means survival time and cost up to a certain, fixed time point (e.g., L=3 yrs), which is chosen so that  $P(\text{time of censoring} > L)$  is sufficiently larger than 0.

**Table 2**

Cost-effectiveness analysis: Schizophrenia trial example

	<b>Olanzapine (N=548)</b>	<b>Haloperidol (N=264)</b>	<b>Difference</b>	<b>95% confidence interval</b>
Cost (mean)	\$27765	\$38066	-\$10301	(-18159, -2256)
Cost (median)	\$7790	\$6486	\$1304	(-3903, 5574)
Effect (mean)	188 days	169 days	19 days	(-2.1, 37)
Effect (median)	200 days	182 days	18 days	(-21, 72)
ACER	\$148/day	\$225/day	-\$77/day	(-133, -20)
ICER (mean)	\$-563/day			( $-\infty$ , -174) in SE (2110, $\infty$ ) in SW*
ICER (median)	\$75/day			(-152, $\infty$ ) in SE/NE ( $-\infty$ , $\infty$ ) in NW/SW*

\* See Figure 1 for the CE plane for easier understanding of these intervals.

See Figure 2 for corresponding CEACs.

Data were uncensored in this example.

**Table 3**

Impact of ignoring censoring in cost estimation: MADIT example

Mean cost (standard error)		
	ICD (N=89)	Conventional (N=92)
Sample mean of all observed cost data	\$75493 (4072)	\$59207 (8038)
Sample mean of uncensored cost data	\$106496 (6417)	\$68189 (11190)
Mean accounting for censoring	\$99548 (5491)	\$72754 (8536)