

UC Davis

UC Davis Previously Published Works

Title

A Tutorial on Net Benefit Regression for Real-World Cost-Effectiveness Analysis Using Censored Data from Randomized or Observational Studies

Permalink

<https://escholarship.org/uc/item/9kv6p84k>

Journal

Medical Decision Making, 44(3)

ISSN

0272-989X

Authors

Chen, Shuai
Bang, Heejung
Hoch, Jeffrey S

Publication Date

2024-04-01

DOI

10.1177/0272989x241230071

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

A Tutorial on Net Benefit Regression for Real-World Cost-Effectiveness Analysis Using Censored Data from Randomized or Observational Studies

Medical Decision Making
2024, Vol. 44(3) 239–251
© The Author(s) 2024



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0272989X241230071
journals.sagepub.com/home/mdm



Shuai Chen , Heejung Bang, and Jeffrey S. Hoch

Given the increasing popularity of person-level cost-effectiveness analysis using “real-world” data, there is a clear need to understand and use methods for observational data. When the cost-effectiveness data are subject to censoring, ignoring censoring is especially error prone for heavily censored data. We summarize best practice and provide a hands-on example of applying the net benefit regression framework for cost-effectiveness analysis, which works for both observational and randomized studies with possibly censored data. Many existing methods are special cases within this framework. We provide step-by-step guidance, user-friendly R programs, and examples to illustrate 1) fitting net benefit regressions for possibly censored cost-effectiveness data; 2) implementing doubly robust methods combining net benefit regressions and propensity scores, which may increase the chances to obtain consistent estimates in observational studies; 3) constructing cost-effectiveness acceptability curves; and 4) interpreting the results. The methods in this tutorial are easy to use and lead to more reliable and robust results using typical administrative data, thus providing an attractive option for real-world cost-effectiveness analysis using possibly censored observational data sets.

Highlights

- We illustrate the steps involved in carrying out cost-effectiveness analysis using net benefit regressions with possibly censored demo data by providing step-by-step guidance and code applied to a data set.
- We demonstrate the importance of these new methods by illustrating how naïve methods for handling censoring can lead to biased cost-effectiveness results.

Keywords

censoring, cost-effectiveness analysis, net benefit regression, non-randomized study, observational data, propensity scores

Date received: March 6, 2023; accepted: January 10, 2024

Background

Clinical studies often encounter challenges associated with right censoring. For instance, at the date of data capture, some patients may have dropped out, been lost

Corresponding Author

Shuai Chen, Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, One Shields Avenue, Med Sci 1C, Davis, CA 95616-5270, USA; (shschen@ucdavis.edu)

to follow-up, or not experienced the event of interest. In such cases, the last date of follow-up is referred to as the censoring time. Patients who are still alive at the time of data capture have costs and survival information available only until their last follow-up times, but their true total costs and full survival time until death are unknown. A recent study of statistical methods used in trial-based economic evaluations found that 15% of methodological studies reviewed described approaches for person-level cost-effectiveness analysis (CEA) with censored data.¹ The study concluded with a recommendation against naïve methods such as, “(1) Simply ignoring the fact that data are censored, and hence assuming that all patients in the study experience the event of interest, and (2) Treating the data as being complete, and hence simply omitting censored cases from the analyses.” All but 1 methodological study in the review was published more than a decade ago.¹ This is surprising given the increasing popularity of person-level CEA using “real-world” data; there is a clear need to understand and use methods for censored data in both randomized and observational studies.

While economic decision analytic models (e.g., Markov models) are commonly used to inform funding decision negotiations in many countries and agencies, the analysis of person-level data (e.g., from a data set with cost and effectiveness information from each patient) is also an important research activity, with many such studies published each year. Cost-effectiveness data sets can be assembled from randomized controlled trials, pragmatic studies, or administrative data sources (e.g., surveillance, epidemiology, and end results linked to Medicare administrative and claims data [SEER-Medicare]). These types of data sets, with person-level information on cost and health outcome, are valuable because they allow researchers to estimate the additional gain from new treatments in actual patients and place the estimate in context with the observed incremental costs.

Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, CA, USA (SC, HB); Division of Health Policy and Management, Department of Public Health Sciences, University of California, Davis, Sacramento, CA, USA (JSH); Center for Healthcare Policy and Research, University of California, Davis, Sacramento, CA, USA (SC, HB, JSH). The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this study was provided in part by a grant from National Institutes of Health (NIH). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. This research was partly supported by National Center for Advancing Translational Sciences, National Institutes of Health (UL1 TR001860). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

However, these sources often contain observational, censored data requiring special care to adjust for the non-randomized, incomplete nature of the correlated cost and effect data. Without accounting for the unique characteristics of the data, the CEA can provide a flawed assessment of the incremental costs associated with the incremental health outcomes, yielding biased estimates and inaccurate uncertainty assessments. However, there exist methods from statistical CEA that can address the challenges inherent in real-world data. Next, we summarize best practice and provide a hands-on example.

Main Summary Statistics of Interest in Statistical Cost-effectiveness

The incremental cost-effectiveness ratio (ICER) is a ratio of the incremental cost to the incremental effect, and as a ratio, it has statistical challenges. The incremental net benefit (INB) is not hampered by these shortcomings and provides an important alternative. Both are functions of the cost (c_i) and clinical effect outcome (e_i) for the new treatment ($Trt_i = 1$) and usual care ($Trt_i = 0$) ($i = 1, \dots, n$ for the i th patient). The $ICER = \Delta C / \Delta E$ and the $INB = \lambda \Delta E - \Delta C$, where ΔC and ΔE are the differences in the mean cost and mean effectiveness, respectively, and λ is a cost-effectiveness threshold value with various interpretations,^{2,3} such as the decision maker's unknown willingness-to-pay value for an additional unit of effectiveness. The INB is also the difference between the expected net benefits of new treatment and usual care, where an individual's net benefit is defined as $nb_i = \lambda e_i - c_i$.

There are various options for characterizing uncertainty (i.e., variability) in statistical CEA,⁴ including Fieller's theorem, bootstrap, and the cost-effectiveness acceptability curve (CEAC). The CEAC shows the probability that the new treatment or intervention is cost-effective as a function of the unknown cost-effectiveness threshold value given the data.^{5,6}

Net Benefit Regression with Complete (Uncensored) Data

A well-established method to analyze a cost-effectiveness data set in statistical CEA involves net benefit regression,⁷⁻⁹ which refers to a regression equation with net benefit ($nb_i = \lambda e_i - c_i$) as the dependent variable and the treatment indicator variable (Trt_i) and possibly other covariates (and potentially their interactions with treatment) as independent variables. The ability to study cost-effectiveness controlling for other covariates is a major

strength of placing CEA in a regression framework. The use of covariate adjustment is essential for observational studies, since treatment assignment or receipt is likely related to covariates, leading to potential confounding issues. In clinical trials, covariates can be used to adjust for imperfect randomization and to improve efficiency.¹⁰ In addition, detecting policy-relevant interactions (e.g., treatment by covariate) can help identify important subgroups with heterogeneous cost-effectiveness through hypothesis-generating analyses.

The above concepts can be formalized with equations by including the treatment indicator Trt_i , $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ip})'$ as a vector including p covariates (e.g., patient demographics), and possibly $Trt_i \times \mathbf{Z}_i$ as the treatment-covariate interactions. For a given cost-effectiveness threshold value of λ , the model is

$$nb_i = \beta_0 + \beta_{Trt}Trt_i + \beta'_Z \mathbf{Z}_i + \beta'_{TrtZ} Trt_i \times \mathbf{Z}_i + \varepsilon_i$$

where ε_i is the error term without a specified distribution. Large and statistically significant elements in β_{TrtZ} for an interaction term point toward important patient subgroups, which is helpful for tailoring decision making. Without censoring, the coefficients can be estimated by ordinary least squares (OLS) (for a given λ), and the standard error can be estimated by bootstrapping or the Huber-White robust estimator.¹¹ The CEAC can be constructed from net benefit regression,⁷ reflecting the probability that $\beta_{Trt} > 0$ for being cost-effective (in a Bayesian sense). The influence of skewness on statistical inference for the net benefit regression method was evaluated in simulation studies,¹² which concluded, "Apart from the confidence intervals for treatment effect being a little conservative (i.e., a little too wide), there appears to be no real cause for concern, even when cost data are log-normal and the total sample size is as small as 100."

Cost-effectiveness with Censored Data

As noted earlier, studies that capture or record the data before complete cost and effectiveness data are available face challenges associated with right censoring, with their true total costs and full survival time until death possibly unknown. It is tempting to simplify the analysis by using simple methods to handle censoring, such as using complete-case data only (i.e., discarding patients without complete data) or using all data ignoring censoring status (i.e., using the observed data at the last follow-up). Although there exist rare special situations in which the ICER and INB could be unbiased with such naïve methods, in general these simple methods are bias prone.

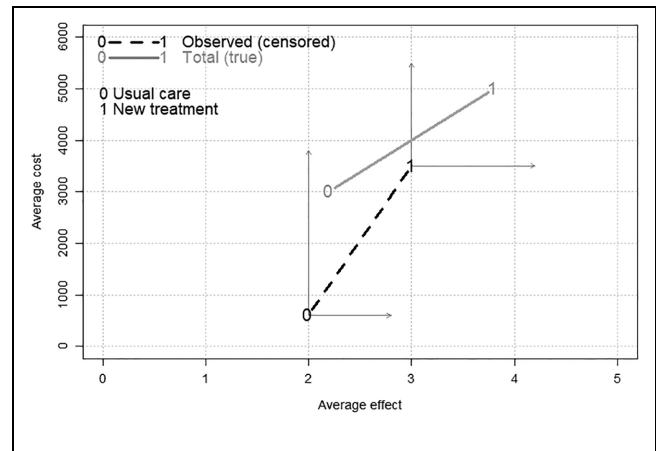


Figure 1 Sample averages for cost and effect for treatment options $Trt = 0$ (usual care as control) and $Trt = 1$ (new treatment), indicated with a “0” or “1,” respectively. By using observed data ignoring censoring, the naïve observed incremental cost-effectiveness ratio (ICER) is the slope of the dashed line. In contrast, the true data (with nothing left out) produce an ICER shown as the slope of the solid line. The vertical and horizontal arrows emanating from the observed 0 and 1 connected with the dashed line show that the true average total cost and effect data may be greater or equal (due to further including cost and effect after censoring), indicated by the 0 and 1 points connected by the solid line, whose slope shows the true ICER. For a purpose of illustration, this figure used hypothetical data, assuming that the control and new treatment groups have true average effects of 2.2 and 3.8 and true average costs of 3,000 and 5,000, respectively. Due to censoring, high costs before death are not fully observed. We assume the control and new treatment groups have the average observed effects of 2 and 3 and average observed costs of 600 and 3,500, respectively.

The relevant impact of censoring is related to how it affects estimates of ΔC and ΔE . Figure 1 shows the sample averages for observed cost and effect (e.g., survival time) for treatment options $Trt = 0$ and $Trt = 1$ (indicated with “0” or “1,” respectively). By using observed data ignoring censoring, the naïve observed ICER is the slope of the dashed line. Censoring means the “true” estimates of expected cost and effect will likely be to the northeast of the observed estimates because the full cost (and effect) will almost always be greater (or equal) to the observed cost (and effect) at the censoring time. The impact of censoring on the ICER is illustrated by the difference in slopes between the dashed and the solid connecting lines (with the true ICER equaling the slope of the solid line in Figure 1).

As the different ICERs in Figure 1 suggest, ignoring censoring is often not a good option. For example, if the

rate of cost accumulation increases around death, and costs at this time are censored, then costs that are unobserved are both different and important. Because health care costs tend to rise dramatically prior to death, it is incorrect to assume the rate of cost accumulation is similar for the period that is observed (before censoring) and the period that is not observed (censored).¹³ Supplementary Figure S1A demonstrates low coverage probabilities for 95% confidence intervals due to bias from using naïve methods (e.g., complete-case data only [CC] and using all data but ignoring censoring status [AL]), based on simulations performed by Chen and Hoch.⁸

Potentially Informative Censoring

Statistical methods to address censoring depend on whether the data are subject to informative or noninformative censoring. Most common survival analysis methods (e.g., the Kaplan-Meier method and Cox proportional hazards model) assume censoring is noninformative (i.e., the survival time is independent of censoring time). However, costs and effectiveness such as quality-adjusted life-years (QALYs) are likely to be subject to informative censoring.¹⁴ For example, a healthier patient will accumulate costs more slowly (with higher quality of life), with less costs (and higher QALYs) both at the censoring time and at the potential death time, leading to informative censoring for cost and QALYs.¹⁵ Consequently, standard survival analysis techniques are not valid to analyze costs and QALYs directly on the cost (or QALY) scale.^{14,16} According to Lagakos,¹⁷ “when the amount of censoring is small, very little bias is likely to result from the use of methods based on noninformative censoring. However, when informative censoring is extensive, substantial biases can occur, and so informative censoring models should be considered.”

Cost-effectiveness Analysis with Censored Observational Data

While using administrative data sets can boost sample size and produce real-world evidence, they are not randomized. For example, Ali et al.¹⁸ assessed the cost-effectiveness of breast-conserving surgery plus hormonal therapy with or without radiotherapy, using SEER-Medicare data, helping to inform the allocation of cancer care resources optimally by generating real-world evidence about the incremental costs of new treatments in relation to their better effectiveness. Their results, based on a real-world data set, suggested that the combination of radiotherapy and hormonal therapy could be

cost-effective from a US Centers for Medicare and Medicaid Services perspective.

Although propensity scores and standard survival analysis techniques are often used to evaluate survival using censored observational data, censoring in costs (and QALYs) are more likely to be informative and hence standard methods for noninformative censoring are not recommended. Performing a proper CEA using statistical methods for valid estimation and inference is not very commonplace for such nonrandomized censored data. Various methods for censored CEA^{14,19–33} have been proposed, providing a good foundation, yet there is room for improvement (e.g., estimates are not adjusted for covariates, not efficient, not doubly robust, and not straightforward to construct a CEAC). This article illustrates net benefit regression for censored cost-effectiveness data from observational studies,⁸ which overcomes the aforementioned disadvantages as well as unifying many methods as special cases (e.g., the methods work for uncensored data and unadjusted analysis).⁸ In the following sections, we introduce the methods and provide step-by-step instructions with R programs to illustrate 1) running net benefit regressions for censored cost-effectiveness data, 2) performing doubly robust methods that combine net benefit regressions and propensity scores, 3) constructing CEACs, and 4) interpreting the results.

Methods for Censored Net Benefit Regression: The Simple Weighted Estimator

Censored net benefit regression⁸ produces consistent estimates when handling censoring using inverse probability of censoring weighting.^{34,35} The simple weighted (SW) estimator uses total costs (and effectiveness) from patients with complete follow-up only, with the weights to represent potential patients that might have been observed. However, the SW estimator does not use cost/effectiveness “history” and wastes information from censored patients (e.g., the observed costs until censoring are ignored in the analysis).

Methods for Censored Net Benefit Regression: The Partitioned Estimator

When cost/effectiveness history (e.g., the monthly or yearly costs and QALYs) is available, the SW estimator can be improved through partitioning into smaller time “buckets.” For example, if the data are grouped into yearly data “buckets,” for patients censored after the k th year, their costs and effectiveness within the first

Table 1 Advantages Regarding Statistical Consistency (Asymptotically Unbiasedness) of the Doubly Robust Method Combining Net Benefit Regression and Propensity Scores in Randomized and Observational Studies

Study Type	Censored Data	NBR Correctly Specified	Propensity Correctly Specified	Asymptotically Unbiased Estimate Benefit from
Randomized	No	Yes	N/A	No advantage
		No	N/A	Robust for misspecification (NBR)
	Yes	Yes	N/A	Adjusting for censoring
		No	N/A	Adjusting for censoring and robust for misspecification (NBR)
Observational	No	Yes	Yes	No advantage
		No	Yes	Robust for misspecification (propensity)
		No	Yes	Robust for misspecification (NBR)
	Yes	Yes	Yes	No advantage
		Yes	Yes	Adjusting for censoring
		No	Yes	Adjusting for censoring and robust for misspecification (propensity)
		No	Yes	Adjusting for censoring and robust for misspecification (NBR)
		No	No	No advantage

NBR, net benefit regression; N/A, not applicable (since treatment is randomly assigned and propensity score is completely known in randomized studies, the consistency is guaranteed; note that propensity scores and doubly robust methods could still be used to improve efficiency in randomized studies).

k years are complete. Therefore, we can fit censored net benefit regression for each year separately, using the yearly cost and effectiveness data. The final partitioned (PT) regression coefficients estimates are the summation of coefficients estimates across all years from the yearly net benefit regressions.⁸

As such, net benefit regression can be used when carrying out CEA with censored data. If only the total costs and effectiveness are available, the SW estimator is useful. If the cost and effectiveness history for different periods is also available, the PT estimator is generally more efficient than the SW estimator is. Both the SW and PT methods are consistent in the statistical sense, with coverage probabilities that are close to the nominal level of 95% in simulation (Supplementary Figure S1B).⁸

Methods for Observational Data: Propensity Scores and Doubly Robust Methods

Define the causal average INB as $\delta = E(nb^{(1)}) - E(nb^{(0)})$, where $nb^{(1)}$ and $nb^{(0)}$ are an individual's potential outcomes of net benefits (one actual and one counterfactual) if the corresponding treatment Trt were 1 and 0 (e.g., new treatment v. usual care), respectively. Since an individual can only receive 1 treatment option, only $nb^{(1)}$ or $nb^{(0)}$ is observed for each person. In administrative data, treatment assignment is not random, leading potentially to selection bias, where patient characteristics are associated with the likelihood of receiving treatment and with the outcomes.

Both propensity scores and doubly robust methods are modern causal methods, which are useful for estimating causal average treatment effects (δ in our case).³⁶ Propensity score methods^{37–39} provide ways to balance measured covariates across treatment and comparison groups. The augmented inverse propensity weighted estimator^{40,41} combines a regression-based estimator and a propensity score-weighting method (inverse probability of treatment weighting) and is a “doubly robust” method in that it requires either the propensity or the outcome model to be correctly specified, but not necessary both.

Censored Net Benefit Regression with Observational Data: A Doubly Robust Estimator

When the net benefit regression (without interaction) is correctly specified, the causal average INB (δ) equals β_{Trt} in the net benefit regression. However, estimates are prone to bias if the net benefit regression model is misspecified. Likewise, if the propensity score model is not modeled correctly, bias may also be introduced. To diminish the possible bias in estimating “causal average INB” due to misspecified models, the doubly robust method⁸ combines both censored net benefit regressions and propensity scores and remains consistent when either the propensity score model or the net benefit regression model is correctly specified. This property may increase the chances to obtain consistent estimates in observational studies, and it can always obtain consistent estimates in randomized trials (since the propensity score is known). Table 1

Table 2 First 3 Rows of the Simulated Data^a

ID	Survival	Dead	Trt	Age65 y	LBBB	Female	cost.1	...	cost.15	tot.cost	QALY.1	...	QALY.15	tot.QALY
1	4.10	0	0	0	0	1	6.272	...	0	9.585	0.63	...	0	2.43
2	3.52	0	1	0	1	1	15.375	...	0	15.773	0.52	...	0	2.21
3	3.78	1	1	0	0	0	13.669	...	0	37.328	0.55	...	0	2.40

^aCosts are in \$1,000s, and survival and quality-adjusted life years (QALYs) are in years.

summarizes the benefits of using the doubly robust method in randomized and observational studies,^{8,24} supporting the contention that the doubly robust method is a useful tool for CEA, especially for administrative data with nonrandomized treatment assignment and possibly censored outcomes.

Data

To illustrate the methods, we simulated a cost-effectiveness data set mimicking administrative data including 2,000 hypothetical patients with cardiovascular disease, motivated by a real cardiovascular study. The data-generation process is similar to the simulation scenario in Chen and Hoch⁸ with these main differences: 1) more covariates are involved (increase from 1 to 3 covariates) and 2) in addition to using life-years as effectiveness, QALYs are also generated based on a heart failure event (i.e., a patient's quality of life tends to be lower after heart failure). More details of the data generation are provided in the Supplementary Materials. For this simulated data set, we know the true (uncensored and counterfactual potential) means based on statistical theory (Supplementary Table S1) and thus the true INBs, which allows us to evaluate the performance of the different methods.

The first few rows of the data are shown in Table 2. The *survival* variable is the follow-up time (in years) since the enrollment into the study. The *dead* variable is mortality indicator with "1" for the occurrence of death (i.e., complete data) and "0" indicating censoring (i.e., time until death is unknown but was known to be alive at the time indicated by *survival*). The *Trt* variable indicates the new treatment group (*Trt* = 1) versus the usual care comparison group (*Trt* = 0). Patients who are younger, with left bundle branch block (LBBB) conduction disturbance are more likely to receive the new treatment than the comparison. The *Age65*, *LBBB*, and *Female* variables are binary baseline covariates, with "1" for age ≥ 65 y, LBBB, and female, respectively, and "0" otherwise. U-shaped costs (high initial and terminal costs) were generated and grouped into yearly observed cost summary

variables called *cost.1*, *cost.2*, and all the way up to *cost.15* (in \$1,000s). In our example, the new treatment has higher initial costs for performing the treatment but lower annual costs subsequently. We used life-years (LYs) and QALYs as outcome measures of effectiveness and grouped QALYs into 15 annual summary variables as well. The patient health-related quality-of-life (QOL) was simulated in each of the yearly intervals, ranging from 1 for good health to 0 at death, and each yearly QALY was calculated as the integration of QOL within that year (see Supplementary Figure S3 for more illustration). When the variables *cost.1* to *cost.15* (and *QALY.1* to *QALY.15*) are summed, they provide the total observed costs (and QALYs) over 15 y, named *tot.cost* (and *tot.QALY*), respectively.

In this data set, costs and QALY data were collected up to 15 y; however, the longest follow-up time was slightly less than 15 y. Therefore, a smaller time limit should be chosen (e.g., 10 y), and time-restricted mean survival time/QALY/costs^{42,43} are used (more details in the following section). A consequence of applying such a restriction is that a patient alive at 10 y is considered as having complete data. This leads to a censoring rate of 49% within the 10-y horizon in this data set.

Steps Involved in Performing Net Benefit Regression Methods for Censored Observational Data

The theory behind censored net benefit regression and the doubly robust method have been described elsewhere.^{8,9} Here, we focus on the steps involved in performing net benefit regression methods for an administrative cost-effectiveness data set. R programs are provided for the main steps, and more examples and technical details are available in the Supplementary Materials.

Step 1: Preparation

Load required package. The first line loads the *NetBenReg* package to perform net benefit regression. Before

using this package, the `NetBenReg` package needs to be installed in R (see Supplementary Materials for details to install this package in R).

```
library(NetBenReg)
```

Load data. The next line loads the data set `CEdata`.

```
data(CEdata)
```

Choose time limit. Due to censoring, the right tail of the distribution of survival time cannot be estimated reliably, leading to unstable estimation of the mean survival/QALY/costs.⁴⁴ For instance, if all patients are followed for up to 10 y, the cost after 10 y is not observed, and thus, estimating mean costs after that time is not reliable. Consequently, researchers often turn to the time-restricted mean survival time/QALY/costs.^{42,43} For example, we can measure survival/QALY/costs saved within a time horizon of L years, where L is chosen such that a “reasonable” number of subjects are still being observed at that time (e.g., choose L as the upper quartile of the follow-up times). This means that we are interested in outcomes accumulated until death or up to L years, whichever occurs first. Although our data set provides the costs observed up to 15 y, the longest follow-up time is 14.99 y, meaning that no patients were observed at

the end of 15 y. Therefore, a smaller time limit L (e.g., 10 y with sufficient data using an intuitive cutoff) could be chosen, and patients alive at 10 y have complete data. This time limit is required for both SW and PT methods. Patients censored within L years are still considered as censored within this time horizon, which can be handled by the SW and PT methods appropriately. In the Supplementary Materials, we provide examples showing error or warning messages in R when we choose too large an L .

Choose cost-effectiveness threshold values. To perform net benefit regression, we need to choose a few cost-effectiveness threshold values denoted by λ . A sequence of λ values from 0 to multiple times of the ICER is a good start. We can first fit covariate-adjusted regressions using the SW method to estimate a covariate-adjusted ICER (\$2,680/QALY, see step 2). Based on this, we can create a cost-effectiveness threshold sequence of \$ 0, \$500, \$1,000, . . . , \$6,000 for 1 additional QALY:

```
lambda=seq(0,6,0.5)
```

This line generates `lambda` as a sequence from 0 to 6 (in \$1,000s), with 0.5 as the increment of the sequence (this creates a range from \$ 0 to \$6,000 jumping by \$500 increments).

Step 2: Fitting Censored Net Benefit Regressions

We can fit a covariate-adjusted net benefit regression using the SW method with QALY as effectiveness:

```
fit1<-NetBenReg(Followup=CEdata$survival, #follow-up time
  delta=CEdata$dead, #1=complete,0=censor
  group=CEdata$Trt, #1=treatment,0=comparison
  Cost=CEdata[,8:22], #a matrix of cost history
  Eff=CEdata[,24:38], #a matrix of effectiveness history
  Part.times=1:15, #end timepoints of intervals
  Method='SW', #use Simple Weighted method
  Z=CEdata[,5:7], #adjust covariates in regressions
  Eff.only=TRUE, #also fit effect-only regression
  Cost.only=TRUE, #also fit cost-only regression
  lambda=lambda, #cost-effectiveness threshold
  L=10) #time limit
Print(fit1)
```


The first line clarifies that the variable *survival* in the data set `CEdata` (described in “Data” section) denotes the follow-up time for each patient. The next line `delta=CEdata$dead` indicates if an individual was complete (1) or censored (0) after the survival time stopped being recorded. The line `group=CEdata$Trt` clarifies if the individual received new treatment (1) or usual care (0). Below that are the lines `Cost=CEdata[,8:22]` and `Eff=CEdata[,24:38]` clarifying the 15 columns for yearly cost and effect data. The line `Part.times=1:15` specifies the end time point as 1, 2, . . . , 15, meaning that the intervals are [0, 1], (1, 2], . . . , (14, 15] years for each grouped cost and effectiveness variable. The option `Method='SW'` requests the SW method. We can also fit the same model using the PT method with `Method='PT'`. Covariates, introduced through `Z=CEdata[,5:7]`, consist of a data frame of the variables *Age65*, *LBBB*, and *Female* stored in columns 5, 6, and 7 of `CEdata`. As a special case, the option `Z` can be removed if no covariates are desired for adjustment (e.g., in randomized studies). The options `Eff.only=TRUE` and `Cost.only=TRUE` additionally fit the regressions with effect and cost as dependent variables, respectively. The final 2 lines `lambda=lambda` and `L=10` refer to the cost-effectiveness threshold and the time horizon of *L* years described above.

To explore heterogeneous cost-effectiveness and identify subgroups, we can further include treatment-covariate interactions. The option `interaction=` specifies the names of covariates having interaction with treatment. For example, the following code adds a treatment-LBBB interaction (bold in following code):

```
fit2<-NetBenReg(Followup=CEdata$survival,
  delta=CEdata$dead, group=CEdata$Trt,
  Cost=CEdata[,8:22], Eff=CEdata[,24:
  38], Part.times=1:15, Method='PT', Z=
  CEdata[,5:7], interaction=c("LBBB"),
  Eff.only=TRUE, lambda=lambda, L=10)
```

Caution. Two naïve methods are also allowed (`Method='CC'` for complete-case only; `Method='AL'` for all data ignoring censoring). However, they are not recommended for censored data due to bias. Without censoring, the 4 methods (SW, PT, CC, AL) are equivalent to OLS.

Interpretation. Table 3 summarizes the results. To evaluate different methods, we used several methods to obtain “true” ICERs and INBs: 1) the “Theory” method is based on the true mean values analytically derived by

statistical theory, leading to the overall ICER of \$2,750/QALY (= \$2,713/0.99 QALY) in the whole population; 2) the “OLS using uncensored data” method is based on the simulated uncensored outcomes (including cost and effect after censoring time); 3) the “Using true outcomes” method is based on the simulated uncensored potential outcomes (counterfactuals), leading to an overall ICER of \$2,501/QALY (= \$2,585/1.03 QALY). Although the estimates from the methods (2) and (3) will approach the mean values from the method (1) when the sample size approaches infinity, the 3 methods could be slightly different with smaller sample sizes, as in our example. Methods (2) and (3) might be better “true” values, since our data set is a censored version of the full uncensored data set that was used to calculate (2) and (3).

For covariate-adjusted regression without interaction, the coefficient for the treatment is the adjusted INB. For example, when the cost-effectiveness threshold is \$6,000/QALY, the estimated covariate-adjusted INB is \$3,410 by the PT method. When $\lambda = 0$, the dependent variable is $-1 \times \text{cost}$, equivalent to the cost-only regression with switched signs for the estimated coefficients; for the effect-only model, the dependent variable = effect (indicated as $\lambda = \text{NA}$ in Table 3). Then it is easy to obtain a covariate-adjusted ICER: for example, the ICER is \$2,680/QALY using the SW method (= \$2,550/0.95 QALY as the ratio of the treatment coefficient estimates in cost-only and effect-only regressions). Comparing the estimates from covariate-adjusted regressions using the 4 methods to the true values (“OLS using uncensored data” method) in Table 3, it is apparent that estimates of both naïve methods are far away from the true values, with the CC method less biased than the AL method in this example. Both SW and PT methods produce similar estimates to the true values, but the PT method is generally more efficient, with 8% to 20% reductions in standard errors by using cost and QALY history.

The coefficient for the treatment-LBBB interaction is significantly positive for effect and all cost-effectiveness threshold λ values, indicating that effect as well as cost-effectiveness are heterogeneous across LBBB status. LBBB patients achieve significantly higher INBs from the new treatment as compared with their non-LBBB counterparts, adjusting for other covariates. The coefficients for the treatment (main effect) and treatment-LBBB interaction can be used to calculate adjusted INBs in LBBB and non-LBBB subgroups, respectively. For example, when the cost-effectiveness threshold is \$6,000/QALY, the regression adjusted INB is $-\$1,610$ for the non-LBBB subgroup and \$8,750 (= \$10,360–\$1,610) for the LBBB subgroup. This result provides evidence that

Table 3 Coefficient Estimates (Standard Errors) (in \$1,000s) and *P* Values for Treatment and Interaction Terms in Net Benefit Regressions with Different Values of Cost-Effectiveness Threshold λ (in \$1,000/QALY) for the Hypothetical Data Using Quality-Adjusted Life-Years (QALYs) as Effectiveness, Limited to a 10-year Time Horizon^a

Method	Model Term	Interpretation	$\lambda = 0$			$\lambda = 3$			$\lambda = 6$			$\lambda = \text{NA (Effect Only)}$		
			Est (SE) [P Value]	Est (SE) [P Value]	Est (SE) [P Value]	Est (SE) [P Value]	Est (SE) [P Value]	Est (SE) [P Value]	Est (SE) [P Value]	Est (SE) [P Value]	Est (SE) [P Value]	Est (SE) [P Value]	Est (SE) [P Value]	
Covariate-adjusted net benefit regression (without interaction)														
(Naïve) CC	Treatment v. control	Covariate-adjusted INB	-3.34 (0.43) [<0.001]	-0.75 (0.60) [0.21]	1.85 (0.93) [0.048]	0.86 (0.14) [<0.001]								
(Naïve) AL			-3.80 (0.30) [<0.001]	-2.03 (0.38) [<0.001]	-0.26 (0.59) [0.65]	0.59 (0.09) [<0.001]								
SW			-2.55 (0.39) [<0.001]	0.30 (0.62) [0.63]	3.14 (1.01) [0.002]	0.95 (0.15) [<0.001]								
PT			-2.49 (0.36) [<0.001]	0.46 (0.53) [0.39]	3.41 (0.82) [<0.001]	0.98 (0.12) [<0.001]								
OLS using uncensored data			-2.42 (0.29) [<0.001]	0.67 (0.44) [0.126]	3.77 (0.70) [<0.001]	1.03 (0.10) [<0.001]								
Covariate-adjusted net benefit regression with treatment-LBBB interaction														
PT	Treatment v. Control	Non-LBBB subgroup's covariate-adjusted INB	-3.24 (0.50) [<0.001]	-2.43 (0.71) [<0.001]	-1.61 (1.11) [0.145]	0.27 (0.16) [0.094]								
	Treatment-LBBB interaction	Difference between LBBB's and non-LBBB's covariate-adjusted INBs	1.52 (0.72) [0.036]	5.94 (1.04) [<0.001]	10.36 (1.61) [<0.001]	1.47 (0.23) [<0.001]								
OLS using uncensored data	Treatment v. Control	Non-LBBB subgroup's covariate-adjusted INB	-3.17 (0.41) [<0.001]	-2.69 (0.58) [<0.001]	-2.21 (0.91) [0.015]	0.16 (0.14) [0.24]								
	Treatment-LBBB interaction	Difference between LBBB's and non-LBBB's covariate-adjusted INBs	1.53 (0.58) [0.008]	6.88 (0.85) [<0.001]	12.22 (1.34) [<0.001]	1.78 (0.19) [<0.001]								
Doubly robust method (covariate-adjusted regression with interaction + propensity scores)														
DR-PT	Treatment v. Control	Causal average INB	-2.43 (0.37) [<0.001]	0.66 (0.53) [0.21]	3.75 (0.83) [<0.001]	1.03 (0.12) [<0.001]								
Using true outcomes			-2.58	0.52	3.62	1.03								
Theory			-2.71	0.25	3.21	0.99								

INB, incremental net benefit; LBBB, left bundle branch block; QALY, quality-adjusted life-years. ^aEst is the coefficient estimate (in \$1,000s) for treatment (dummy variable) and treatment-LBBB interaction from net benefit regressions or causal average INB from doubly robust methods; coefficients for other covariates are omitted. SE is the estimated standard errors. Covariate-adjusted INB is estimated by the coefficient for treatment (dummy variable) in covariate-adjusted regressions without interactions. Non-LBBB subgroup's covariate-adjusted INB is estimated by the coefficient for treatment in covariate-adjusted regressions with the treatment-LBBB interaction, and difference between LBBB's and non-LBBB's covariate-adjusted INBs is estimated by the coefficient for treatment-LBBB interaction in the same regressions, and LBBB subgroup's covariate-adjusted INB can be calculated by their summation. When $\lambda = 0$, the dependent variable = $-1 \times \text{cost}$, equivalent to the cost-only regression by switching the signs of estimated coefficients. For $\lambda = \text{NA}$, the dependent variable = effect (does not belong to net benefit regressions). PT is the partitioned net benefit regression using cost and QALY history. SW is the simple weighted net benefit regression. (Naïve) CC is the regression using complete-case (uncensored) cases only. (Naïve) AL is the regression using all data ignoring censoring status. For the uncensored data, the 4 methods (SW, PT, CC, AL) are equivalent to ordinary least squares (OLS). DR-PT is the doubly robust method with the partitioned method. Results from OLS using uncensored data and using true outcomes are based on the generated uncensored (and potential counterfactual outcomes) in the data simulation process, and results from theory are based on the true mean values analytically derived by theory. These 3 methods are used as true values to evaluate performance of other methods.

the cost-effectiveness of the new treatment can vary by LBBB status.

Step 3: Using a Doubly Robust Method to Estimate Causal Average INB with Observational Data

Propensity scores. The propensity score is the probability of receiving the new treatment, which could be estimated by fitting a logistic regression (or other models) with treatment indicator as the dependent variable and the potential confounders as explanatory variables. Generally, if a variable is thought to be related to the outcome but not the treatment, including it in the propensity score (for which it is an irrelevant variable) should reduce bias.⁴⁵ More practical guidelines about choice of variables in propensity score models can be found elsewhere.⁴⁵⁻⁴⁸

Implementing the doubly robust method. We can perform the doubly robust method by further specifying the option `Doubly.Robust=TRUE` to request the doubly robust method and the option `PS.Z=CEdata[,5:7]` to include the 3 covariates into the propensity score model using logistic regression.

```
fit3<-NetBenReg(Followup=CEdata$survival,
  delta=CEdata$dead, group=CEdata$Trt,
  Cost=CEdata[,8:22], Eff=CEdata[,24:
  38], Part.times=1:15, Method='PT', Z=
  CEdata[,5:7], interaction=c("LBBB"),
  PS.Z=CEdata[,5:7], Doubly.Robust=TRUE,
  Eff.only=TRUE, lambda=lambda, L=10)
```

Caution. We can examine whether there exists extreme propensity scores close to 0 or 1 (program in the Supplementary Materials), which is important because extreme propensity scores may lead to huge weights and hence unstable results. Crump et al.⁴⁹ suggested a rule of thumb of thresholds of 0.1 and 0.9 for extreme propensity scores. By default, the function `NetBenReg` replaces the propensity scores smaller than 0.1 by 0.1 (and greater than 0.9 by 0.9). It also produces warning messages when there are estimated propensity scores outside [0.1, 0.9]. Supplementary Figure S2 shows that the range of estimated propensity scores are sufficiently overlapping across treatment groups without extreme values, indicating that patients in the 2 groups are comparable after propensity score adjustment.

Interpretation. The doubly robust method estimates causal average INB, which is causal average treatment

effects on net benefit for the entire population (those who did and did not receive the new treatment). Table 3 summarizes the estimates from the doubly robust method, which are close to the true values.

Step 4: Constructing CEAC

For net benefit regression without interaction or the doubly robust method, the CEAC can be constructed based on the P value (p) for $\hat{\beta}_{Trt}$ (or $\hat{\delta}$ for the doubly robust method, where $\delta = E(nb^{(1)}) - E(nb^{(0)})$ is the causal average INB). With a given cost-effectiveness threshold λ value, the vertical axis for the CEAC is $p/2$ if $\hat{\beta}_{Trt} < 0$ or $1 - p/2$ if $\hat{\beta}_{Trt} > 0$ (i.e., 1-sided P value of $\hat{\beta}_{Trt}$).⁷ For net benefit regressions with interactions, the cost-effectiveness is heterogeneous, and hence, subgroup-specific CEACs can be constructed.⁸ Note that the most common interpretation of a CEAC is in a Bayesian framework. Thus, the 1-sided P value can be interpreted as the “probability that the intervention is cost-effective” due to its equivalence to Bayesian analysis incorporating a noninformative prior.⁵⁰ The impact of informative priors on CEACs has been discussed elsewhere.⁵¹

Interpretation. Figure 2 demonstrates 4 CEACs based on the fitted models (made using the program in the Supplementary Materials), showing the probabilities that the new treatment is cost-effective compared with the comparison group at different cost-effectiveness threshold values (a sequence of \$ 0, \$500, . . . , \$6,000 chosen in step 1). Among LBBB patients, the probability that the new treatment is cost-effective is much higher than the probability among non-LBBB patients, indicating heterogeneous cost-effectiveness across LBBB status. This was foreshadowed by the significant treatment-LBBB interaction observed earlier.

Conclusions

Funding negotiations are often informed by research evidence introduced through a health technology assessment (HTA) process. As the field of HTA begins to embrace real-world evidence to address well-known limitations in randomized trials, there will be demand for CEA conducted using person-level administrative data. Even in the United States, increases in available cost and health outcome data combined with Medicare’s new capabilities to consider drug costs as well as their effectiveness suggest that CEA may be a very strategic part of future comparative effectiveness research intent on informing health care funding decisions. In addition, if approvals for new

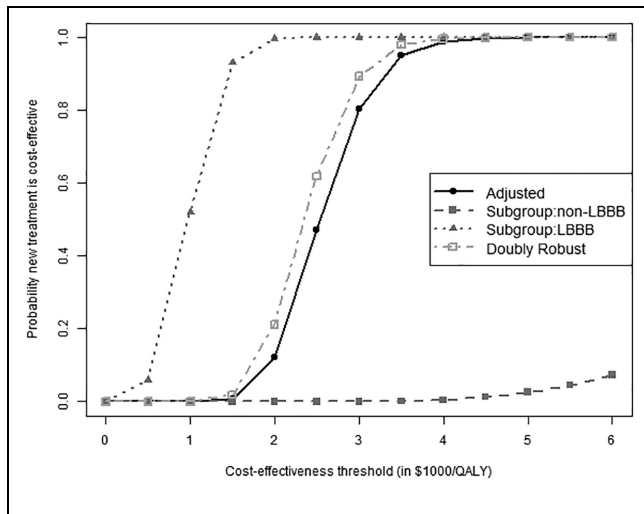



Figure 2 The cost-effectiveness acceptance curves (CEACs) for the hypothetical data limited to a 10-y time horizon, using quality-adjusted life-years (QALYs) as effectiveness with different values of cost-effectiveness threshold. The black solid curve is from covariate-adjusted partitioned regression. The gray dotted and dashed curves are for left bundle branch block (LBBB) and non-LBBB patients, respectively, based on the partitioned regression including treatment by LBBB interaction and adjusting for other covariates (age, gender, in addition to LBBB). The gray dash-dotted curve is from the doubly robust method (combining covariate-adjusted regression with an LBBB-treatment interaction and propensity scores).

drugs continue to outpace the evidence,⁵² the demand for real-world evidence about value (i.e., cost and effectiveness) promises to continue to grow.

Therefore, applying state-of-the-art methods to analyze censored observational data in a net benefit regression framework is essential. When cost-effectiveness data are censored, naïve methods to handle censoring should be avoided, especially for heavily censored data. The doubly robust method combines net benefit regressions and propensity scores in an easy-to-use manner, leading to more reliable results for observational data with censored costs or health outcomes. In addition, the methods can be applied to randomized clinical trials as well. This provides a strong option for CEA using possibly censored data from observational and randomized studies. With the methods illustrated in this article, potential challenges due to nonrandomized and censored data can be addressed in a sound manner.

ORCID iD

Shuai Chen  <https://orcid.org/0000-0002-4016-8217>

Data Availability

The user-friendly R package, simulated data sets, and code to generate them, worked example, and technical documentation are available at <https://github.com/shuaichencode/NetBenReg>.

Supplemental Material

Supplementary material for this article is available online at <https://doi.org/10.1177/0272989X241230071>.

References

1. El Alili M, van Dongen JM, Esser JL, Heymans MW, van Tulder MW, Bosmans JE. A scoping review of statistical methods for trial-based economic evaluations: the current state of play. *Health Econ.* 2022;31(12):2680–99. DOI: 10.1002/hec.4603
2. Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. *J Health Econ.* 1997;16(1):1–31. DOI: 10.1016/s0167-6296(96)00506-1
3. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess.* 2015;19(14):1–503, v-vi. DOI: 10.3310/hta19140
4. Briggs AH, O'Brien BJ. The death of cost-minimization analysis? *Health Econ.* 2001;10(2):179–84. DOI: 10.1002/hec.584
5. van Hout BA, Al MJ, Gordon GS, Rutten FFH. Costs, effects and C/E-ratios alongside a clinical trial. *Health Econ.* 1994;3:309–19.
6. Groot Koerkamp B, Hunink MG, Stijnen T, Hammitt JK, Kuntz KM, Weinstein MC. Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis. *Med Decis Making.* 2007;27(2):101–11. DOI: 10.1177/0272989X06297394
7. Hoch JS, Rockx MA, Krahn AD. Using the net benefit regression framework to construct cost-effectiveness acceptability curves: an example using data from a trial of external loop recorders versus Holter monitoring for ambulatory monitoring of “community acquired” syncope. *BMC Health Serv Res.* 2006;6(1):1–8. DOI: 10.1186/1472-6963-6-68
8. Chen S, Hoch JS. Net-benefit regression with censored cost-effectiveness data from randomized or observational studies. *Stat Med.* 2022;41(20):3958–74. DOI: 10.1002/sim.9486
9. 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 Econ.* 2002;11(5):415–30. DOI: 10.1002/hec.678
10. US Food and Drug Administration. Adjusting for covariates in randomized clinical trials for drugs and biological products final guidance for industry. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized->

- clinical-trials-drugs-and-biological-products. [Accessed 22 November, 2023].
11. Kleiber C, Zeileis A. *Applied Econometrics with R*. New York: Springer; 2008.
 12. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Econ*. 2004;13(5):461–75.
 13. Wijesundera HC, Wang X, Tomlinson G, Ko DT, Krahn MD. Techniques for estimating health care costs with censored data: an overview for the health services researcher. *Clinicoecon Outcomes Res*. 2012;4:145–55. DOI: 10.2147/CEOR.S31552
 14. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics*. 1997;53(2):419–34. DOI: 10.2307/2533947
 15. Lin DY. Regression analysis of incomplete medical cost data. *Stat Med*. 2003;22(7):1181–200. DOI: 10.1002/sim.1377
 16. Zhao H, Tsiatis AA. A consistent estimator for the distribution of quality adjusted survival time. *Biometrika*. 1997;84:339–48.
 17. Lagakos SW. General right censoring and its impact on the analysis of survival data. *Biometrics*. 1979;35(1):139–56.
 18. Ali AA, Tawk R, Xiao H, et al. Comparative cost-effectiveness of radiotherapy among older women with hormone receptor positive early-stage breast cancer. *Expert Rev Pharmacoecon Outcomes Res*. 2022;22(5):735–41. DOI: 10.1080/14737167.2022.2044309
 19. Lin DY. Linear regression of censored medical costs. *Biostatistics*. 2000;1(1):35–47. DOI: 10.1093/biostatistics/1.1.35
 20. Bang H, Tsiatis AA. Estimating medical costs with censored data. *Biometrika*. 2000;87(2):329–43.
 21. Zhao H, Tian L. On estimating medical cost and incremental cost-effectiveness ratios with censored data. *Biometrics*. 2001;57(4):1002–8. DOI: 10.1111/j.0006-341x.2001.01002.x
 22. Pullenayegum EM, Willan AR. Semi-parametric regression models for cost-effectiveness analysis: improving the efficiency of estimation from censored data. *Stat Med*. 2007;26(17):3274–99. DOI: 10.1002/sim.2814
 23. Goldfeld KS. Twice-weighted multiple interval estimation of a marginal structural model to analyze cost-effectiveness. *Stat Med*. 2014;33(7):1222–41. DOI: 10.1002/sim.6017
 24. Wang X, Beste LA, Maier MM, Zhou X-H. Double robust estimator of average causal treatment effect for censored medical cost data. *Stat Med*. 2016;35(18):3101–16. DOI: 10.1002/sim.6876
 25. Satten G, Datta S. The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *Am Stat*. 2001;55(3):207–10. DOI: 10.1198/000313001317098185
 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. *Med Decis Making*. 2002;22(2):173–7. DOI: 10.1177/0272989X0202200214
 27. Willan AR, Lin DY, Manca A. Regression methods for cost-effectiveness analysis with censored data. *Stat Med*. 2005;24(1):131–45. DOI: 10.1002/sim.1794
 28. Etzioni R, Urban N, Baker M. Estimating the costs attributable to a disease with application to ovarian cancer. *J Clin Epidemiol*. 1996;49(1):95–103. DOI: 10.1016/0895-4356(96)89259-6
 29. Fenn P, McGuire A, Phillips V, Backhouse M, Jones D. The analysis of censored treatment cost data in economic evaluation. *Med Care*. 1995;33(8):851–63. DOI: 10.1097/00005650-199508000-00009
 30. Li J, Vachani A, Epstein A, Mitra N. A doubly robust approach for cost-effectiveness estimation from observational data. *Stat Methods Med Res*. 2018;27(10):3126–38. DOI: 10.1177/0962280217693262
 31. Fontaine C, Daurès JP, Landais P. On the censored cost-effectiveness analysis using copula information. *BMC Med Res Methodol*. 2017;17(1):27. DOI: 10.1186/s12874-017-0305-9
 32. Handorf EA, Heitjan DF, Bekelman JE, Mitra N. Estimating cost-effectiveness from claims and registry data with measured and unmeasured confounders. *Stat Methods Med Res*. 2019;28(7):2227–42. DOI: 10.1177/0962280218759137
 33. Willan AR, Lin DY, Cook RJ, Chen EB. Using inverse-weighting in cost-effectiveness analysis with censored data. *Stat Methods Med Res*. 2002;11(6):539–51. DOI: 10.1191/0962280202sm308ra
 34. Horvitz D, Thompson D. A generalization of sampling without replacement from a finite universe. *J Am Stat Assoc*. 1952;47(260):663–85. DOI: 10.7717/peerj.1634
 35. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000;56(3):779–88. DOI: 10.1111/j.0006-341x.2000.00779.x
 36. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550–60. DOI: 10.1097/00001648-200009000-00011
 37. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41–55. DOI: 10.1093/biomet/70.1.41
 38. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *J Am Stat Assoc*. 2005;100(469):322–31. DOI: 10.1198/016214504000001880
 39. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med*. 2004;23(19):2937–60. DOI: 10.1002/sim.1903
 40. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J Am Stat Assoc*. 1995;90(429):106–21. DOI: 10.2307/2291134
 41. Kurz CF. Augmented inverse probability weighting and the double robustness property. *Med Decis Making*. 2022;42(2):156–67. DOI: 10.1177/0272989X211027181
 42. Han K, Jung I. Restricted mean survival time for survival analysis: a quick guide for clinical researchers. *Korean J Radiol*. 2022;23(5):495–99.

43. Lueza B, Mauguen A, Pignon JP, Rivero-Arias O, Bonastre J, MAR-LC Collaborative Group. Difference in restricted mean survival time for cost-effectiveness analysis using individual patient data meta-analysis: evidence from a case study. *PLoS One*. 2016;11(3):e0150032.
44. Huang Y. Calibration regression of censored lifetime medical cost. *J Am Stat Assoc*. 2002;97:318–27.
45. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol*. 2006;163(12):1149–56. DOI: 10.1093/aje/kwj149
46. Garrido MM, Kelley AS, Paris J, et al. Methods for constructing and assessing propensity scores. *Health Serv Res*. 2014;49(5):1701–20. DOI: 10.1111/1475-6773.12182
47. Austin PC. A tutorial and case study in propensity score analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality. *Multivariate Behav Res*. 2011;46(1):119–51. DOI: 10.1080/00273171.2011.540480
48. Brooks JM, Ohsfeldt RL. Squeezing the balloon: propensity scores and unmeasured covariate balance. *Health Serv Res*. 2013;48(4):1487–507. DOI: 10.1111/1475-6773.12020
49. Crump RK, Hotz VJ, Imbens GW, Mitnik OA. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*. 2009;96(1):187–99.
50. Briggs AH. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Econ*. 1999;8(3):257–61. DOI: 10.1002/(sici)1099-1050(199905)8:3<257::aid-hec427>3.0.co;2-e
51. Hoch JS, Blume JD. Measuring and illustrating statistical evidence in a cost-effectiveness analysis. *J Health Econ*. 2008;27(2):476–95. DOI: 10.1016/j.jhealeco.2007.07.002
52. Belluck P, Robbins R. Three F.D.A. advisers resign over agency's approval of Alzheimer's drug. *The New York Times*. 11 June, 2021. Available from: <https://www.nytimes.com/2021/06/10/health/aduhelm-fda-resign-alzheimers.html>