Title

stata.com

stteffects intro — Introduction to treatment effects for observational survival-time data

Description

This entry provides an overview of the treatment-effects estimators that use observational survival-time data and are implemented in `stteffects`. It also provides an overview of the potential-outcomes framework and its application to survival-time data and to the interpretation of the treatment-effects parameters estimated.

The `stteffects` command estimates average treatment effects (ATEs), average treatment effects on the treated (ATETs), and potential-outcome means (POMs). Each of these effect parameters is discussed in this entry. `stteffects` implements a variety of estimators for the ATE, ATET, and POM. The treatment effects can be estimated using regression adjustment (RA), inverse-probability weights (IPW), inverse-probability-weighted regression adjustment (IPWRA), and weighted regression adjustment (WRA). This entry also provides some intuition for the estimators and discusses the tradeoffs between them.

Remarks and examples

stata.com

Remarks are presented under the following headings:

Introduction

The `stteffects` command estimates treatment effects using observational survival-time data.

For some intuition about the methods implemented in the `stteffects` command, consider the following question: Does smoking decrease the time to a second heart attack in the population of women aged 45–55 who have had one heart attack? Three aspects of this question stand out.

1. For ethical reasons, these data will be observational.
2. This question is about the time to an event, and such data are commonly known as survival-time data or time-to-event data. These data are nonnegative and, frequently, right-censored.
3. Many researchers and practitioners want an effect estimate in easy-to-understand units of time.

Aspect 1 is one of the most common reasons for using observational data, and aspect 2 focuses interest on survival-time data.

We are most concerned with aspect 3 because it helps us define and understand the effect of interest. In particular, we would like to know the average change in time to a second heart attack that would occur in the population if all women smoked instead of if no women smoked. This effect is an ATE.

We must solve a missing-data problem to estimate the ATE. The ATE is the population average of the contrast in outcomes when everyone gets the treatment and when no one gets the treatment. Formally, we write this as

$$\text{ATE} = E(t_1 - t_0)$$

where t_1 is the survival time when a subject gets the treatment and t_0 is the survival time when a subject does not get the treatment. For each treatment level, there is a potential outcome that would be observed if a subject received that treatment level: t_1 is the potential outcome that would occur if someone gets the treatment and t_0 is the potential outcome that would occur if someone does not get the treatment. The missing-data problem arises because each subject receives only one treatment level, and so we observe only one of the two potential outcomes.

Much of the survival-time literature uses a hazard ratio as the effect of interest. The ATE has three advantages over the hazard ratio as an effect measure.

1. The ATE measures the effect in the same time units as the outcome instead of in relative conditional probabilities.
2. The ATE is much easier to explain to nontechnical audiences.
3. The models used to estimate the ATE can be much more flexible. Hazard ratios are useful for population effects when they are constant, which occurs when the treatment enters linearly and the distribution of the outcome has a proportional-hazards form. Neither linearity in treatment nor proportional-hazards form is required for the ATE, and neither is imposed on the models fit by the estimators implemented in `stteffects`.

The estimators implemented in `stteffects` use the common missing-data techniques of regression modeling, weighting, and combinations thereof to account for data lost to censoring and to unobserved potential outcomes.

Here we note only a few contributions and entry points to the vast literature on estimating ATEs. The use of potential outcomes to define treatment effects has proved extraordinarily useful; see Holland (1986), Rubin (1974), and Heckman (1997). Cameron and Trivedi (2005, chap. 25), Wooldridge (2010, chap. 21), and Vittinghoff et al. (2012, chap. 9) provide excellent general introductions to estimating ATEs.

❑ Technical note

Left-truncation would be another type of missing data. The estimators implemented in `stteffects` do not adjust for left-truncation, so `stteffects` cannot be used with delayed-entry data.

`stteffects` cannot be used with time-varying covariates or multiple-record data because these add a repeated-measure structure that significantly complicates the estimation problem.

❑

A quick tour of the estimators

The stteffects command implements five estimators of treatment effects. We introduce each one by showing the basic syntax used to apply it to a common example dataset. See each command's entry for detailed information.

We have some fictional data on the time to a second heart attack among women aged 45–55 years. The treatment, smoking, is stored in the 0/1 indicator smoke. These data also contain each woman's age at the time of her first heart attack (age), and indices of her exercise level (exercise), diet quality (diet), and education attainment (education) prior to her first heart attack.

Like streg and other survival-time commands, stteffects uses the outcome variable and the failure indicator computed by stset. In this dataset, atime is the observed time in years to the second heart attack, and fail is the 0/1 indicator that a second heart attack was observed and recorded in atime. (When fail is 1, atime records the time to the second attack; when fail is 0, atime records a censored observation of the time to the second attack.)

We begin our examples by first reading in the data and then specifying the raw outcome and failure variables to stset.

```
. use https://www.stata-press.com/data/r18/sheart
(Time to second heart attack (fictional))
. stset atime, failure(fail)
Survival-time data settings
         Failure event: fail!=0 & fail<.
Observed time interval: (0, atime]
     Exit on or before: failure
──────────────────────────────────────────────────────────────────────────
      2,000  total observations
          0  exclusions
──────────────────────────────────────────────────────────────────────────
      2,000  observations remaining, representing
      1,208  failures in single-record/single-failure data
  3,795.226  total analysis time at risk and under observation
                                                At risk from t =         0
                                     Earliest observed entry t =         0
                                          Last observed exit t =  34.17743
```

The output indicates that 1,208 of the 2,000 observations record actual time to a second heart attack. The remaining observations were censored. Now that we have stset the data, we can use stteffects.

Regression adjustment

Regression modeling of the outcome variable is a venerable approach to solving the missing-data problem in treatment-effects estimation. Known as the regression-adjustment (RA) estimator, this method uses averages of predicted outcomes to estimate the ATE. If the outcome model is well specified, this approach is surprisingly robust.

Example 1: RA estimation

We now use `stteffects ra` to estimate the ATE by RA. We model the outcome as a function of `age`, `exercise`, `diet`, and `education`, and we specify that `smoke` is the treatment variable.

```
. stteffects ra (age exercise diet education) (smoke)
        Failure _d: fail
  Analysis time _t: atime
Iteration 0:  EE criterion =  1.006e-14
Iteration 1:  EE criterion =  2.305e-25
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : regression adjustment
Outcome model  : Weibull
Treatment model: none
Censoring model: none
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
smoke (Smoker vs Nonsmoker)	-1.956657	.3331787	-5.87	0.000	-2.609676	-1.303639
POmean						
smoke Nonsmoker	4.243974	.2620538	16.20	0.000	3.730358	4.75759

When all women in the population smoke, the average time to a second heart attack is estimated to be 1.96 years less than when no women smoke. The estimated average time to a second heart attack when no women smoke is 4.24 years.

The output reports that a Weibull model was used for the outcome. The other outcome models available are exponential, gamma, and log normal. See example 2 in [CAUSAL] **stteffects ra** for an application of the gamma parameterization to this model.

The ratio of the ATE to control-level POM measures the importance of the effect. In this example, when all women smoke, the time to a second heart attack falls by an estimated 46% relative to the case in which none of them smoke. See example 3 in [CAUSAL] **stteffects ra** for an example that uses `nlcom` to compute a point estimate and a confidence interval for this ratio.

◁

Unlike the IPW estimator discussed in the next section, RA does not model treatment assignment or the censoring process. Treatment assignment is handled by fitting separate models for each treatment level and averaging the predicted outcomes. As is standard in the survival-time literature, the censoring term in the log-likelihood function accounts for censoring; see Kalbfleisch and Prentice (2002, chap. 3), Cameron and Trivedi (2005, chap. 17), Cleves, Gould, and Marchenko (2016, chap. 13), and Wooldridge (2010, chap. 22).

See [CAUSAL] **stteffects ra** for further discussion of this command and the RA estimator.

Inverse-probability weighting

Sometimes researchers are more comfortable modeling treatment assignment than the outcome. Inverse-probability-weighted (IPW) estimators use weighted averages of the observed outcome to estimate the POMs and the ATE. The weights correct for the missing data. When there is no censoring, the missing potential outcome is the only missing data, and the weights are constructed from a model of treatment assignment. When the data may be censored, the weights must control for censoring and the missing potential outcome. In this case, IPW estimators construct the weights from two models, one for the censoring time and one for treatment assignment.

▷ Example 2: IPW estimation

Here we use `stteffects ipw` to estimate the effect of smoking on the time to a second heart attack. The model of assignment to the treatment `smoke` depends on `age`, `exercise`, `diet`, and `education`. The time-to-censoring model also depends on `age`, `exercise`, `diet`, and `education`.

```
. stteffects ipw (smoke age exercise diet education)
> (age exercise diet education)

        Failure _d: fail
  Analysis time _t: atime

Iteration 0:  EE criterion =  2.042e-18
Iteration 1:  EE criterion =  3.283e-31

Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
Censoring model: Weibull
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
smoke (Smoker vs Nonsmoker)	-2.187297	.6319837	-3.46	0.001	-3.425962	-.9486314
POmean						
smoke Nonsmoker	4.225331	.517501	8.16	0.000	3.211047	5.239614

When all women in the population smoke, the average time to a second heart attack is estimated to be 2.19 years less than when no women smoke. The estimated average time to a second heart attack when no women smoke is 4.23 years. When all women smoke, the average time to a second heart attack falls by an estimated 52% relative to the case when no women smoke.

The estimates have changed; however, the interpretation is the same as for the RA estimator because the IPW and RA estimators are estimating the same population effects. Under correct model specification, the estimates will differ in finite samples, but the size of these differences will decrease as the sample size gets larger. For the case at hand, the estimated ATE and control-level POM are roughly similar to those produced by the RA estimator using the Weibull model for the outcome.

Recall that IPW estimators are weighted averages of observed outcomes and that the weights control for the missing outcomes. Weights in survival-time data have two components: one for the missing potential outcome and one for data lost to censoring. We used a logit model for treatment assignment, so the component of the weights that controls for the missing potential outcome comes from the

estimated logit parameters. We used a Weibull model for the time to censoring, so the component of the weights that controls for data lost to censoring comes from the estimated Weibull parameters. ◁

Using weighting from an estimated treatment-assignment model to control for the missing potential outcome is standard in the treatment-effects literature; for example, see [CAUSAL] **teffects intro advanced**, Wooldridge (2010, chap. 21), Vittinghoff et al. (2012, chap. 9), Hirano, Imbens, and Ridder (2003), Cattaneo (2010), and Cattaneo, Drukker, and Holland (2013). Modeling the time to censoring is specific to the survival-time treatment-effects literature; see Bai, Tsiatis, and O'Brien (2013) and Robins and Rotnitzky (2006). See *Methods and formulas* in [CAUSAL] **stteffects ipwra** for more details.

See [CAUSAL] **stteffects ipw** for further discussion of this command and the IPW estimator.

Combinations of RA and IPW

More efficient estimators are obtained by combining IPW and RA, due to Wooldridge (2007) and Wooldridge (2010, chap. 21) and denoted by IPWRA. Unlike the estimators discussed in Wooldridge (2010, chap. 21), both the treatment and the outcome models must be correctly specified to estimate the ATE.

The IPWRA estimator uses estimated weights that control for missing data to obtain missingness-adjusted regression coefficients that are used to compute averages of predicted outcomes to estimate the POMs. The estimated ATE is a contrast of the estimated POMs. These weights always involve a model for treatment assignment. You choose whether to account for censoring by including a term in the log-likelihood function or whether to use weights that also account for the data lost to censoring.

▷ Example 3: Likelihood-adjusted-censoring IPWRA estimation

We model the outcome (time to a second heart attack) as a function of `age`, `exercise`, `diet`, and `education`. We model assignment to the treatment `smoke` as a function of the same covariates.

```
. stteffects ipwra (age exercise diet education)
> (smoke age exercise diet education)
       Failure _d: fail
  Analysis time _t: atime
Iteration 0:  EE criterion =  6.476e-15
Iteration 1:  EE criterion =  1.116e-26
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : IPW regression adjustment
Outcome model  : Weibull
Treatment model: logit
Censoring model: none
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
smoke (Smoker vs Nonsmoker)	-1.592494	.4872777	-3.27	0.001	-2.54754	-.637447
POmean						
smoke Nonsmoker	4.214523	.2600165	16.21	0.000	3.7049	4.724146

The estimated ATE of −1.59 and control-level POM of 4.21 are similar to the reported values of −1.96 and 4.24 in example 1.

We did not specify a model for the time to censoring, so censoring is handled by including a term in the log-likelihood function in the Weibull outcome model. We denote this likelihood-adjusted-censoring (LAC) version of the IPWRA estimator by LAC-IPWRA.

◁

▷ Example 4: Weighted-adjusted-censoring IPWRA estimation

Instead of including a term in the log-likelihood function, the weighted-adjusted-censoring IPWRA (WAC-IPWRA) estimator uses estimated weights to adjust for censoring. We model the time to a second heart attack as a function of `age`, `exercise`, `diet`, and `education`; we model assignment to the treatment `smoke` as a function of the same covariates; and we model the time to censoring as a function of `age`, `exercise`, and `diet`.

```
. stteffects ipwra (age exercise diet education)
> (smoke age exercise diet education) (age exercise diet)
        Failure _d: fail
  Analysis time _t: atime
Iteration 0:  EE criterion =  2.797e-13
Iteration 1:  EE criterion =  2.032e-25
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : IPW regression adjustment
Outcome model  : Weibull
Treatment model: logit
Censoring model: Weibull
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
smoke (Smoker vs Nonsmoker)	-2.037944	.6032549	-3.38	0.001	-3.220302	-.855586
POmean						
smoke Nonsmoker	4.14284	.4811052	8.61	0.000	3.199891	5.085789

The estimated ATE of −2.04 and control-level POM of 4.14 are similar to the reported values of −1.96 and 4.24 in example 1.

The weights for censoring are constructed from the estimated parameters because we specified a time-to-censoring model.

◁

Under correct specification, both versions of the IPWRA estimator estimate the same ATE and control-level POM as estimated by RA and IPW.

The addition of the time-to-censoring model makes the WAC-IPWRA somewhat less robust than the LAC-IPWRA estimator. Weighting methods to control for censoring also place more restrictive assumptions on the censoring process. For example, the censoring time must be random, otherwise it would be impossible to construct the weights. In *Assumptions and tradeoffs* below, we discuss the tradeoffs among the estimators and the assumptions that each requires. For the moment, we note that we believe the LAC-IPWRA estimator is more robust than the WAC-IPWRA estimator.

See [CAUSAL] **stteffects ipwra** for further discussion of this command and the IPWRA estimator.

Weighted regression adjustment

When estimating the parameters of an outcome model, the weighted regression-adjustment (WRA) estimator uses weights instead of a term in the log-likelihood function to adjust for censoring. These weights are constructed from a model for the censoring process. The estimated parameters are subsequently used to compute averages of predicted outcomes that estimate the POMs. A contrast of the estimated POMs estimates the ATE.

▷ Example 5: WRA estimation

We model the time to a second heart attack as a function of `age`, `exercise`, `diet`, and `education`; we specify that `smoke` is the treatment; and we model the time to censoring as a function of `age`, `exercise`, and `diet`.

```
. stteffects wra (age exercise diet education) (smoke) (age exercise diet)
        Failure _d: fail
  Analysis time _t: atime
Iteration 0:  EE criterion =  7.116e-15
Iteration 1:  EE criterion =  5.859e-27
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : weighted regression adjustment
Outcome model  : Weibull
Treatment model: none
Censoring model: Weibull
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
smoke (Smoker vs Nonsmoker)	-2.152014	.4986005	-4.32	0.000	-3.129253	-1.174775
POmean						
smoke Nonsmoker	4.079273	.4379517	9.31	0.000	3.220903	4.937642

The estimated ATE of −2.15 and control-level POM of 4.08 are similar to the reported values of −1.96 and 4.24 in example 1. Like the other estimators discussed, the WRA estimators estimate the same effect parameters as the RA estimator, so the interpretation is the same.

◁

In many survival-time applications, using weights to adjust for censoring is probably less robust than just including a term in the log-likelihood function for the outcome model. The model used to construct the weights is just as complicated as the outcome model, and including the term in the log-likelihood function places fewer restrictions on the censoring process, as discussed in *The correct adjustment for censoring assumption* below.

See [CAUSAL] **stteffects wra** for further discussion of this command and the WRA estimator.

Average treatment effect on the treated

Intuitively, the average treatment effect on the treated (ATET) is the effect in a well-defined, at-risk subpopulation. Sometimes the subpopulation that gets the treatment defines such an at-risk subpopulation. For example, we may want to know the average change in time to a second heart attack among female smokers aged 45–55 who have had a heart attack if they all became nonsmokers. This effect is the ATET.

Below, we use `stteffects ra` to estimate the ATET by RA.

```
. stteffects ra (age exercise diet education) (smoke), atet
        Failure _d: fail
  Analysis time _t: atime
Iteration 0:  EE criterion =  1.006e-14
Iteration 1:  EE criterion =  2.970e-26
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : regression adjustment
Outcome model  : Weibull
Treatment model: none
Censoring model: none
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATET						
smoke (Smoker vs Nonsmoker)	-1.527476	.2489203	-6.14	0.000	-2.015351	-1.039602
POmean						
smoke Nonsmoker	3.436937	.2217808	15.50	0.000	3.002255	3.87162

Now, all effects are calculated only for the subpopulation of women aged 45–55 years who smoke after their first heart attack. If no women in the subpopulation were to smoke, the average time to a second heart attack would be 3.44 years. When all women in the subpopulation smoke (the observed behavior), the average time to a second heart attack is estimated to be 1.53 years less than if no women in the subpopulation had smoked. In other words, if we could somehow turn all smokers in the subpopulation into nonsmokers, the average time to a second heart attack would be 3.44 years instead of 1.91 years ($3.44 - 1.53 = 1.91$).

These point estimates are a little different than those for the ATE and the control-level POM in the full population of women aged 45–55 years who have had one heart attack. The difference indicates that this particular health cost of smoking may be smaller among women who choose to smoke than in the full population.

Comparison of treatment-effects estimators

We can classify the estimators implemented in `stteffects` into five categories: 1) estimators based on a model for the outcome variable; 2) estimators based on models for the treatment assignment and the censoring time; 3) estimators based on models for the outcome variable and the treatment assignment; 4) estimators based on models for the outcome variable, the treatment assignment, and the censoring time; and 5) estimators based on models for the outcome variable and the censoring time.

Because there are several categories of estimators, the user must decide whether to model the outcome, the probability of treatment, the time to censoring, or some combination thereof.

Each category of estimator contains a variety of choices about the functional forms for the models.

We now provide some intuition behind each category of estimator and discuss the relationships.

1. When modeling only the outcome, separate outcome models for each treatment level account for treatment assignment, and censoring is adjusted for in the log-likelihood function. This approach is used in the RA estimators.
2. Some researchers would rather avoid modeling the outcome. Some estimators use weighted averages of the observed outcome to estimate the effect. When estimating treatment effects from observational survival-time data, the weights used must account for treatment assignment and censoring. Models for treatment assignment and time to censoring are used to construct the weights. This approach is used in the IPW estimators.
3. When seeking a more efficient estimator, it is natural to model both the outcome and the treatment and to adjust for censoring in the outcome model. This approach is used in the LAC-IPWRA estimators.
4. When seeking a more efficient estimator, another natural approach is to model both the outcome and the treatment and to adjust for censoring by weights that come from a time-to-censoring model. This approach is used in the WAC-IPWRA estimators.
5. We could modify approach 1 to model the outcome and the time to censoring so that censoring is handled by weighting and its own model instead of by likelihood adjustment. This approach is used in the WRA estimators.

While researcher preferences over what to model largely dictate the approach selected, we quickly note two points that could affect which approach works best. First, we can adjust for censoring by weighting only when censoring time is random. Second, weighting estimators become unstable if the weights get too large.

In the next section, we elaborate on the assumptions needed and the tradeoffs among the approaches to estimation.

Assumptions and tradeoffs

The estimators implemented in `stteffects` require three assumptions: conditional independence, sufficient overlap, and correct adjustment for censoring.

The conditional independence assumption

All estimators implemented in `stteffects` require the potential outcomes to be independent of the treatment assignment after conditioning on the covariates. Randomized experiments and the Heckman selection model are two motivating frameworks for the conditional independence assumption.

When the treatment is assigned randomly, the randomization ensures that the potential outcomes are independent of the treatment assignment. In observational data, the treatment is not randomly assigned. However, many important questions can only be answered using observational data because it would be unethical to randomly allocate hazardous treatments, for example, smoking. The conditional independence assumption in observational data says that treatment assignment is as good as random after conditioning on the covariates.

We can also understand conditional independence from a modeling framework. The Heckman selection model specifies that each of the potential outcomes and the treatment assignment process are functions of observable covariates and unobservable errors. The potential outcomes are conditionally independent of the treatment assignment when the unobservable errors in the treatment-assignment process are independent of the unobservable errors in each of the potential-outcome processes. See *The CI assumption* in [CAUSAL] **teffects intro advanced** for a detailed example.

Both frameworks lead to the same conclusion: we need to observe and to condition on a sufficient number of covariates.

Essentially, all the estimators in `stteffects` are equally susceptible to violations of the conditional independence assumption. No one estimator is any more robust to the conditional independence assumption than any other one.

Estimating the ATE among the subpopulation of those who get the treatment requires a significantly weaker version of the CI assumption; see *Assumptions for the ATET* below.

For more details about the conditional independence assumption, see *The CI assumption* in [CAUSAL] **teffects intro advanced**, and see Rosenbaum and Rubin (1983), Heckman (1997), Imbens and Wooldridge (2009), Cameron and Trivedi (2005, sec. 25.2), Wooldridge (2010, chap. 21), and Vittinghoff et al. (2012, chap. 9).

The sufficient overlap assumption

The sufficient overlap assumption requires that each individual have a sufficiently positive probability of being assigned to each treatment level. We believe that the RA estimator is more robust than the other estimators to near violations of the sufficient overlap condition, under correct model specification.

The overlap condition has no specification test, but using `teoverlap` and then summarizing the predicted treatment probabilities presents good diagnostics of overlap problems.

The correct adjustment for censoring assumption

The correct adjustment for censoring assumption has two parts. First, either the censoring time must be fixed or the process must be conditionally-on-covariates independent of the potential outcomes and the treatment-assignment process. This assumption is standard in survival analysis; see, for example, Kalbfleisch and Prentice (2002, chap. 3).

Second, the method used to adjust to censoring must be correct. For the RA and LAC-IPWRA estimators, which use likelihood-adjusted censoring, the second assumption is no more restrictive than assuming correct specification of the outcome model. For the IPW, WAC-IPWRA, and WRA estimators, which adjust by weighting, the second assumption requires that the censoring be random and that the censoring process be correctly modeled.

Under correct specification, all the estimators in `stteffects` perform well. However, we believe that estimators that use likelihood adjustment instead of weighting are more robust for three reasons.

1. The estimators that use weighting to adjust for censoring cannot handle fixed censoring processes. If the censoring process is not random, the weights are not well defined.
2. The estimators that use weighting to adjust for censoring do not allow the random censoring process to vary by treatment level.
3. The estimators that use weighting to adjust for censoring require an additional sufficient overlap condition: the probability of not being censored must be sufficiently greater than 0 or else the weights that adjust for censoring get too large.

While the estimators that use WAC instead of LAC require a few more assumptions, some researchers are more comfortable modeling the treatment and censoring than the outcome. In this case, the IPW or WAC-IPWRA estimator would be the estimator of choice.

See *Specification diagnostics and tests* below for information about testing these assumptions.

Assumptions for the ATET

We noted in *Average treatment effect on the treated* that the ATET is sometimes more interesting than the ATE. We can also estimate the ATET under less restrictive versions of the conditional independence assumption and the sufficient overlap assumption than those required for the ATE.

While ATE estimation requires that the potential outcomes for both the treated and the not treated be conditionally independent of treatment assignment, ATET estimation requires that only the not treated potential outcome be conditionally independent of treatment assignment.

This weaker version of conditional independence allows the gains from the treatment to be related to treatment assignment, after conditioning on the covariates. We can estimate the ATET, but not the ATE, if some unobserved factor increases (or decreases) the likelihood of assignment to the treatment, increases (or decreases) the time to event in the treatment group, and has no effect on the time to event when not in the treatment group.

For example, suppose that smoking is an acquired taste and that individuals who acquire the taste for smoking more easily are less adversely affected by smoking and otherwise similar to everyone else when not smoking. Taste for smoking is unobservable, and our data have no measure of this variable. In this case, we could estimate the ATET but not the ATE.

The weaker version of the sufficient overlap assumption only requires that each individual in the treated subpopulation have a positive probability of not getting treated. In contrast, ATE estimation requires that each individual in the population have a positive probability of getting each treatment level. In particular, we can estimate the ATET, but not the ATE, when some individuals in the population have zero chance of getting the treatment. For example, we could estimate the ATET, but not the ATE, if some women will never smoke for religious reasons.

Even when the conditions for ATE estimation hold, the ATE and ATET may differ. Finding that the ATET is significantly different from the ATE does not mean that the ATE is incorrectly estimated.

See Heckman (1997) and Wooldridge (2010, 911–912) for more information about the assumptions necessary to estimate the ATET.

Specification diagnostics and tests

After `stteffects ipw` and `stteffects ipwra`, some specification checks for the treatment-assignment model and the overlap condition are available.

The checks for the treatment-assignment model are known as balance checks. When the covariate distributions are invariant to the treatment level, the covariates are said to be balanced. The concept of balanced covariates comes from the experimental literature, in which random treatment assignment ensures that the covariates are balanced.

In observational data, the covariates are almost never balanced in the raw data. Weighting methods can be viewed as using a treatment-assignment model to balance the covariates. If the treatment-assignment model is well specified, the weights constructed from this model will balance the covariates. One of the nice features of balance checks is that they do not depend on the outcome or its distribution. This fact is especially useful for survival-time outcomes because censoring of the outcome has no effect on the balance checks, so the balance checks implemented in `tebalance` work without modification.

Conditional on the treatment-assignment model being well specified, we can use the estimated probabilities of treatment, known as the propensity scores, to look for signs that the overlap condition is violated. These checks depend only on the estimated treatment probabilities and are not affected by any censoring of the outcome, so the methods implemented in `teoverlap` work without modification.

We begin examining our model by using `tebalance summarize` after refitting the models used by the LAC-IPWRA estimator.

```
. quietly stteffects ipwra (age exercise diet education)
> (smoke age exercise diet education)
. tebalance summarize
Covariate balance summary
                          Raw     Weighted
------------------------------------------
Number of obs =         2,000      2,000.0
Treated obs   =           738        994.1
Control obs   =         1,262      1,005.9
------------------------------------------
```

	Standardized differences		Variance ratio	
	Raw	Weighted	Raw	Weighted
age	-.3122094	-.0184574	.8547308	.9370065
exercise	-.4975269	-.0458412	.4966778	.8342339
diet	-.2479756	.0021802	.7937645	1.095347
education	-.4801442	-.0216366	.6015139	.978078

The weighted standardized differences are much closer to 0 than the raw standardized differences, and the weighted variance ratios are much closer to 1 than the raw variance ratios. These results indicate that the model-based treatment weights balanced the covariates; see [CAUSAL] **tebalance** and [CAUSAL] **tebalance summarize** for details.

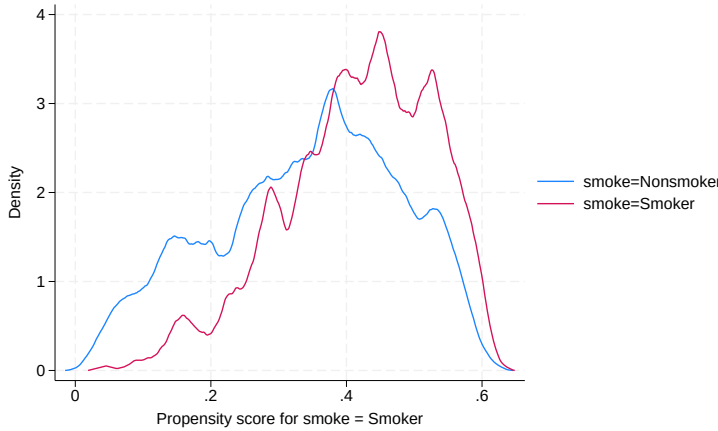
The diagnostics presented by `tebalance summarize` are not a formal test. However, we can use `tebalance overid` to conduct a formal test of the hypothesis that the weights constructed from the treatment-assignment model balanced the covariates.

```
. tebalance overid
Iteration 0:  Criterion =  .22681884
Iteration 1:  Criterion =  .22692316  (backed up)
Iteration 2:  Criterion =  .23090158
Iteration 3:  Criterion =   .2311461
Iteration 4:  Criterion =  .23256285
Iteration 5:  Criterion =  .23286304
Iteration 6:  Criterion =  .23335858
Iteration 7:  Criterion =   .2335567
Iteration 8:  Criterion =   .2335671
Iteration 9:  Criterion =  .23356711
Overidentification test for covariate balance
H0: Covariates are balanced
         chi2(5)      =  3.28142
         Prob > chi2  =   0.6567
```

There is no significant evidence against the null hypothesis. The interpretation is that we do not reject the null hypothesis that the treatment-assignment model is well specified; see [CAUSAL] **tebalance** and [CAUSAL] **tebalance overid** for details.

Given that we do not reject the treatment-assignment model, we can use this model to look for evidence that the overlap condition is violated. We begin by using `teoverlap`.

```
. teoverlap, ptlevel(Smoker)
```



The densities of the propensity scores for the smokers and nonsmokers appear to have the same support, indicating that there is no violation of the overlap condition. The only indicator of a possible problem is that the support of the density for nonsmokers gets very close to 0. This problem would affect ATE estimation but not ATET estimation, as discussed in *Assumptions and tradeoffs*. To further investigate, we compute and summarize the predicted propensity score by treatment level.

```
. predict ps1, ps tlevel(Smoker)
. summarize ps1 if smoke == 0
```

Variable	Obs	Mean	Std. dev.	Min	Max
ps1	1,262	.3410001	.1381673	.014819	.6161401

```
. summarize ps1 if smoke == 1
```

Variable	Obs	Mean	Std. dev.	Min	Max
ps1	738	.4168805	.1107557	.0454891	.6216282

To interpret these results, recall that ATE estimation requires that the minimum propensity score for each treatment level be sufficiently greater than 0 and that the maximum propensity score for each treatment level be sufficiently less than 1. Also recall that ATET estimation only requires that the maximum propensity score for each treatment level be sufficiently less than 1.

For ATE estimation, only the minimum predicted propensity score for nonsmokers presents a challenge, and 0.015 is probably not too small. For ATET estimation, neither maximum causes concern.

For information about choosing among the `stteffects` estimators and their functional forms for the different models, see *Model choice* under *Remarks and examples* in [CAUSAL] **teffects intro advanced**.

Multivalued treatments

`stteffects` can estimate treatment effects for multivalued treatments; here we provide some examples. See [CAUSAL] **teffects multivalued** for an introduction to interpreting effects from multivalued treatments.

Example 6: Multivalued ATE estimation

We have another fictional dataset that records the time to a second heart attack among women aged 45–55 years. In this dataset, `atime` is the observed time in years to the second heart attack, and `fail` is the 0/1 indicator that a second heart attack was observed and recorded in `atime`. (When `fail` is 1, `atime` records the time to the second attack; when `fail` is 0, `atime` records a censored observation of the time to the second attack.)

These data also contain the age at the time of the first heart attack (`age`), and indices of each woman's exercise level (`exercise`), diet quality (`diet`), and education attainment (`education`) prior to her first heart attack.

The treatment, smoking, is stored in the categorical variable `smoke`, which has the following value labels. The women who never smoked are labeled as `N`; the women who previously smoked but quit before their first heart attack are labeled as `B`; the women who previously smoked but quit after their first heart attack are labeled as `A`; and the women who continued to smoke after their first heart attack are labeled as `S`.

We begin by first reading in the data and then reviewing information previously stored using `stset`.

```
. use https://www.stata-press.com/data/r18/sheartm, clear
(Time to second heart attack (fictional))
. stset
-> stset atime, failure(fail)
Survival-time data settings
         Failure event: fail!=0 & fail<.
Observed time interval: (0, atime]
     Exit on or before: failure
──────────────────────────────────────────────────────────────────────────
     10,000  total observations
          0  exclusions
──────────────────────────────────────────────────────────────────────────
     10,000  observations remaining, representing
      9,741  failures in single-record/single-failure data
 27,999.155  total analysis time at risk and under observation
                                             At risk from t =         0
                                  Earliest observed entry t =         0
                                       Last observed exit t =  17.40826
```

We continue by tabulating the treatment variable `smoke`.

```
. tabulate smoke
    Smoking |
      level |      Freq.     Percent        Cum.
────────────┼──────────────────────────────────────
          N |      3,167       31.67       31.67
          B |      2,263       22.63       54.30
          A |      1,924       19.24       73.54
          S |      2,646       26.46      100.00
────────────┼──────────────────────────────────────
      Total |     10,000      100.00
```

We see that 31.67% of the women never smoked, 22.63% of the women previously smoked but quit before their first heart attack, 19.24% of the women previously smoked but quit after their first heart attack, and 26.46% of the women continued to smoke after their first heart attack.

We now use `stteffects ra` to estimate the ATE by RA. We model the outcome as a function of `age`, `exercise`, `diet`, and `education`, and we specify that `smoke` is the treatment variable.

```
. stteffects ra (age exercise diet education) (smoke)
        Failure _d: fail
  Analysis time _t: atime
Iteration 0:  EE criterion =  1.390e-20
Iteration 1:  EE criterion =  2.220e-29
Survival treatment-effects estimation           Number of obs     =     10,000
Estimator      : regression adjustment
Outcome model  : Weibull
Treatment model: none
Censoring model: none
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
smoke						
(B vs N)	-.4129793	.0317	-13.03	0.000	-.47511	-.3508485
(A vs N)	-1.281031	.032866	-38.98	0.000	-1.345447	-1.216614
(S vs N)	-2.167359	.0338994	-63.93	0.000	-2.233801	-2.100917
POmean						
smoke						
N	3.745919	.0289014	129.61	0.000	3.689273	3.802565

The average time to a second heart attack is 0.41 years sooner when all the women smoked at some point but quit before their first heart attack than when all the women never smoked. The average time to a second heart attack is 1.28 years sooner when all the women smoked at some point but quit after their first heart attack than when all the women never smoked. The average time to a second heart attack is 2.17 years sooner when all the women continued to smoke after their first heart attack than when all the women never smoked.

◁

▷ Example 7: Multivalued ATET estimation

In the at-risk subpopulation of women who continued to smoke, we want to estimate the effect of continuing to smoke (S) versus quitting after the first heart attack (A). Below we estimate the ATETs by RA, specifying A to be the control level and S to be the treatment level.

```
. stteffects ra (age exercise diet education) (smoke), atet control(A) tlevel(S)
        Failure _d: fail
  Analysis time _t: atime
Iteration 0:  EE criterion =  1.390e-20
Iteration 1:  EE criterion =  2.049e-29
Survival treatment-effects estimation           Number of obs     =     10,000
Estimator      : regression adjustment
Outcome model  : Weibull
Treatment model: none
Censoring model: none
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATET						
smoke						
(N vs A)	1.290123	.0377552	34.17	0.000	1.216125	1.364122
(B vs A)	.8748349	.0239595	36.51	0.000	.8278751	.9217946
(S vs A)	-.8869257	.0272301	-32.57	0.000	-.9402958	-.8335557
POmean						
smoke						
A	2.500108	.0217833	114.77	0.000	2.457413	2.542802

The parameter (S vs A) is the one of interest. The estimate implies that the average time to a second heart attack among women who continue to smoke is 0.89 years sooner when they all continue to smoke than when they all quit smoking after their first heart attack.

◁

Acknowledgments

We thank Xiaofei Bai of Servier, Matias Cattaneo of Princeton University, Barbara Sianesi of the Institute for Fiscal Studies, and Jeff Wooldridge of Michigan State University for helpful conversations about methods and the literature.

References

Bai, X., A. A. Tsiatis, and S. M. O'Brien. 2013. Doubly robust estimators of treatment-specific survival distributions in observational studies with stratified sampling. *Biometrics* 69: 830–839. https://doi.org/10.1111/biom.12076.

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cattaneo, M. D. 2010. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics* 155: 138–154. https://doi.org/10.1016/j.jeconom.2009.09.023.

Cattaneo, M. D., D. M. Drukker, and A. D. Holland. 2013. Estimation of multivalued treatment effects under conditional independence. *Stata Journal* 13: 407–450.

Cleves, M. A., W. W. Gould, and Y. V. Marchenko. 2016. *An Introduction to Survival Analysis Using Stata*. Rev. 3rd ed. College Station, TX: Stata Press.

Guo, S., and M. W. Fraser. 2015. *Propensity Score Analysis: Statistical Methods and Applications*. 2nd ed. Thousand Oaks, CA: SAGE.

Heckman, J. J. 1997. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources* 32: 441–462. https://doi.org/10.2307/146178.

Hirano, K., G. W. Imbens, and G. Ridder. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71: 1161–1189. https://doi.org/10.1111/1468-0262.00442.

Holland, P. W. 1986. Statistics and causal inference. *Journal of the American Statistical Association* 81: 945–960. https://doi.org/10.2307/2289064.

Imbens, G. W. 2000. The role of the propensity score in estimating dose–response functions. *Biometrika* 87: 706–710. https://doi.org/10.1093/biomet/87.3.706.

Imbens, G. W., and J. M. Wooldridge. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47: 5–86. https://doi.org/10.1257/jel.47.1.5.

Kalbfleisch, J. D., and R. L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*. 2nd ed. New York: Wiley.

Robins, J. M., and A. Rotnitzky. 2006. Inverse probability weighting in survival analysis. In *Survival and Event History Analysis*, ed. N. Keiding and P. K. Andersen, 266–271. Chichester, UK: Wiley.

Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41–55. https://doi.org/10.2307/2335942.

Royston, P. 2015. Estimating the treatment effect in a clinical trial using difference in restricted mean survival time. *Stata Journal* 15: 1098–1117.

Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688–701. https://doi.org/10.1037/h0037350.

Vittinghoff, E., D. V. Glidden, S. C. Shiboski, and C. E. McCulloch. 2012. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. 2nd ed. New York: Springer.

Wooldridge, J. M. 2007. Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics* 141: 1281–1301. https://doi.org/10.1016/j.jeconom.2007.02.002.

——. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Also see

[CAUSAL] **stteffects postestimation** — Postestimation tools for stteffects

[CAUSAL] **teffects intro advanced** — Advanced introduction to treatment effects for observational data

[ST] **streg** — Parametric survival models

[ST] **stset** — Declare data to be survival-time data

[U] 20 Estimation and postestimation commands