Treatment-Effects for Survival-Time Outcomes: Theory and Applications using Stata 14

Rebecca Pope

Health Econometrician
StataCorp

24 September 2015
Oceania Stata Users Group
Canberra, Australia
A research question

- Do post-release support programs decrease recidivism?
  - Does not identify an intervention or the effect we are interested in
- What is the effect of participation in a supported work program after release from prison on the time until a subsequent arrest?
  - Identified a treatment
  - Question is about the time to an event
  - Implies that we want our effect estimate in easy-to-understand units of time
- We could choose a randomized design or an observational design
  - Not always feasible
Random-assignment case

- When the treatment is randomly assigned, the outcome we observe is independent of the treatment assignment process.
- This simplifies our calculation of the treatment effect.
  - The average time until rearrest for ex-prisoners who were assigned to the supported-work program would be a good estimator of what we would expect to happen for all the participating ex-prisoners in the population.
  - The average time until rearrest for ex-prisoners who were not assigned to the supported-work program would be a good estimator of what we would expect to happen for all nonparticipating ex-prisoners in the population.
- The average treatment effect is just the difference between two observed averages:

\[
ATE = \mathbb{E}[t_i(\text{treatment})] - \mathbb{E}[t_i(\text{notreatment})]
\]
Observational case

► Many questions require using observational data, because experimental data would be unethical
  ► We could not ask a random selection of women to smoke after having a heart attack to see if smoking decreased the time to a second heart attack
► Other questions may use observational data because while assignment could be random, participation will not be
  ► We could study time to rearrest between prisoners who chose to participate and those who did not
► We may wish to understand the relationship between variables that select individuals into treatment if we could affect those variables
► A randomized study may be cost- or time-prohibitive
Survival-time data
Right-censored survival-time data
Right-censored survival-time data
Time as we see it

Duration Observed and Censoring Status

- **Observed failure (rearrest=1)**
- **Lost to follow-up**
- **Administratively censored**

Administrative censoring shown for illustration purposes only
stset the data

\[
\text{. stset rtime, failure(rearrest) noshow}
\]

\[
\begin{align*}
\text{failure event:} & \quad \text{rearrest} \neq 0 \& \text{rearrest} < . \\
\text{obs. time interval:} & \quad (0, \text{rtime}] \\
\text{exit on or before:} & \quad \text{failure}
\end{align*}
\]

\begin{center}
\begin{tabular}{ll}
4260 & total observations \\
0 & exclusions \\
4260 & observations remaining, representing \\
3766 & failures in single-record/single-failure data \\
70570.174 & total analysis time at risk and under observation \\
\end{tabular}
\end{center}

at risk from \( t = 0 \) \\
earliest observed entry \( t = 0 \) \\
last observed exit \( t = 308.702 \)

3,766 of the 4,260 observations record actual time to rearrest; remaining 494 were censored (11.6%)
The data at hand

. describe

Contains data from recidivism.dta

| obs: 4,260 | Rearrest data (simulated) |
| vars: 10 | 11 Sep 2015 11:30 |
| size: 80,940 |

<table>
<thead>
<tr>
<th>variable name</th>
<th>storage</th>
<th>display</th>
<th>value</th>
<th>label</th>
<th>variable label</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>int</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>Subject ID</td>
</tr>
<tr>
<td>age</td>
<td>byte</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>Age (years)</td>
</tr>
<tr>
<td>married</td>
<td>byte</td>
<td>%11.0g</td>
<td>mstat</td>
<td></td>
<td>Marital status</td>
</tr>
<tr>
<td>educ</td>
<td>byte</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>Years of education</td>
</tr>
<tr>
<td>wrkhis</td>
<td>int</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>Longest job duration (months)</td>
</tr>
<tr>
<td>parrest</td>
<td>byte</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>Number of prior arrests</td>
</tr>
<tr>
<td>crime</td>
<td>byte</td>
<td>%14.0g</td>
<td>crime</td>
<td></td>
<td>Type of crime</td>
</tr>
<tr>
<td>partic</td>
<td>byte</td>
<td>%9.0g</td>
<td>treat</td>
<td></td>
<td>Participation status</td>
</tr>
<tr>
<td>rearrest</td>
<td>byte</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>Rearrested (1=yes)</td>
</tr>
<tr>
<td>rtime</td>
<td>double</td>
<td>%10.0g</td>
<td></td>
<td></td>
<td>Months to rearrest (w/ rand. cens.)</td>
</tr>
</tbody>
</table>

Sorted by: id
A Cox model for the treatment effect

- Many researchers would start by fitting a Cox model
- The Cox model leaves the baseline hazard unspecified
  - Increases the flexibility of the model
- Estimate the probability that the event will occur in the next moment, given that it has not yet happened
- Our estimated effect is a ratio of the hazard for those in the treatment group to the hazard for those in the control group
Our question answered by the Cox PH model

- Include age at the time of release (age) and the number of prior arrests (parrest) as covariates in our model
- The estimated hazard ratio is

\[
\frac{\lambda(t|x, \text{partic} = 1)}{\lambda(t|x, \text{partic} = 0)} = \frac{\lambda_0(t) \exp(\beta_{partic} + x_0 \beta_o)}{\lambda_0(t) \exp(x_0 \beta_o)} = \exp(\beta_{partic})
\]

where \(x_0 \beta_o = \text{age} \beta_{age} + \text{parrest} \beta_{parrest}\)

Our research question

What is the effect of participation in a supported work program after release from prison on the time until a subsequent arrest?

- Effect we are estimating is the ratio of the hazard of rearrest for those who participate in the program to the hazard of rearrest those who do not participate in the program
Example: Cox PH model

```
. stcox i.partic age parrest, nolog
Cox regression -- no ties
No. of subjects =  4,260 Number of obs   =  4,260
No. of failures =  3,766
Time at risk    =  70570.17407
LR chi2(3)      =  405.10
Log likelihood  =  -27517.455 Prob > chi2    =  0.0000

         _t  Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------- -------- -------- -------- -------- ------------------
    partic     .5662226  .019774   -16.29  0.000       .528763   .606336
        age      1.038272  .0035322   11.04  0.000       1.031372   1.045218
     parrest     1.081966  .0082901   10.28  0.000       1.065839   1.098337
```

- Participating decreases the hazard of rearrest by a factor of about 0.57 relative to not participating
- But how many months of delay is that?
Problems to solve

1. We wanted our effect in terms of units of time, but we got a hazard ratio instead
   - It can be hard for nontechnical audiences to understand the units of the hazard ratio

2. The interpretation of the hazard ratio for the treatment variable as the effect of the treatment is only true if the treatment enters the $x\beta$ term linearly
   - If the treatment is interacted with other covariates, the effect of the treatment varies over individuals
A parametric survival model for the treatment effect

- PH models can also be estimated by specifying a distribution for the hazard function
  - Gompertz, Weibull, exponential
- Using a parametric method allows us to also obtain an estimate of the baseline hazard
- We can make predictions about mean survival time that rely on this baseline hazard and use them in subsequent calculations
- The hazard ratio has the same interpretation as in the Cox model
## Example: Weibull PH model

```
. streg i.partic age parrest, distribution(weibull) nolog
Weibull regression -- log relative-hazard form
No. of subjects = 4,260 Number of obs = 4,260
No. of failures = 3,766
Time at risk = 70570.17407
Log likelihood = -7242.5323 Prob > chi2 = 0.0000
```

|   | Haz. Ratio | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|---|------------|-----------|-----|-----|----------------------|
| _t |            |           |     |     |                      |
| 1.partic | .5642147  | .0195953  | -16.48 | 0.000 | .5270867 - .603958  |
| age | 1.038531   | .0035222  | 11.15 | 0.000 | 1.031651 - 1.045458 |
| parrest | 1.082521  | .0082692  | 10.38 | 0.000 | 1.066434 - 1.09885 |
| _cons | .0414343   | .0042275  | -31.20 | 0.000 | .0339245 - .0506065 |
| /ln_p | -.2314783  | .0125867  | -18.39 | 0.000 | -.2561477 - -.2068089 |
| p   | .7933599   | .0099857  |     |     | .7740276 - .813175  |
| 1/p | 1.260462   | .015865   |     |     | 1.229748 - 1.291944 |

- Participating decreases the hazard of rearrest by a factor of about 0.56 relative to not participating
Example: Weibull PH model with an interaction

```bash
.streg i.partic##c.parrest c.age, distribution(weibull) nolog
```

Weibull regression -- log relative-hazard form

|                       | Haz. Ratio | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|-----------------------|------------|-----------|------|------|----------------------|
| _t                    | 1.partic   | 1.110348  | 0.0133167 | 8.73 | 0.000 | 1.084552 1.136757   |
|                       | parrest    | 1.110348  | 0.0133167 | 8.73 | 0.000 | 1.084552 1.136757   |
|                       | partic#c.parrest | .9588852       | 0.0148206 | -2.72 | 0.007 | .9302729 .9883775  |
|                       | age        | 1.038525  | 0.0035235 | 11.14 | 0.000 | 1.031641 1.045453  |
|                       | _cons      | 0.0368276 | 0.0040948 | -29.69 | 0.000 | .0296162 .0457948  |
|                       | /ln_p      | -.2300976 | 0.0125948 | -18.27 | 0.000 | -.2547829 -.2054122 |
|                       | p          | .7944561  | 0.010006 |                | .7750847 .8143116  |
|                       | 1/p        | 1.258723  | 0.0158534 |                | 1.228031 1.290182  |

Hazard ratio must now be interpreted conditional on the number of prior arrests
Conditional vs. marginal

- Conditional interpretation is helpful for questions that focus on individuals
  - What is the probability that a patient will live 15 years past diagnosis if she gets the treatment given that she is 45 years old?
  - What is the probability that a former prisoner will be rearrested before 18 months if he participates in the supported work program given that he has 3 prior convictions?

- Marginal interpretation is helpful for questions that focus on populations
  - What is the average increase in life expectancy for patients that get this new treatment?
  - What is the average increase in time-to-rearrest for former prisoners who participate in a supported work program?
By switching to a potential-outcomes framework for analysis of observational survival-time data we gain

1. Estimation techniques that deal with inherently missing information
   - What would have happened had an individual who was not treated been treated instead
   - What would have happened had an individual who was treated not been treated instead

2. Effects that are easier to interpret
   - The estimated effect will be the average difference in time until the event occurs when everyone gets the treatment instead of when no one gets the treatment

3. Effects that are population-averaged, even when the treatment is interacted with a covariate
The “fundamental problem of causal inference” (Holland (1986)) is that we only observe one of the potential outcomes

The other potential outcome is missing

1. We only see $t_{partic}$ for ex-prisoners who participated in the supported-work program
2. We only see $t_{nopartic}$ for ex-prisoners who did not participate in the supported-work program

We can use classic tools for missing-data analysis to estimate treatment effects

- Regression adjustment
- Weighting observations for the probability that they were observed
Goal: Find the missing potential outcome

- For each treatment level, there is a potential outcome that we would observe if a subject received that treatment level.
- Potential outcomes are the data that we wish we had to estimate causal treatment effects.
- Suppose that we could see
  1. the time to rearrest for each ex-prisoner when he participated in the post-release supported-work program, and
  2. the time to rearrest for each ex-prisoner when he did not participate in the post-release supported-work program.
Potential outcomes

For example, we wish we had data like

![Graph showing potential time to rearrest for each treatment status.](image-url)
Average treatment effect

- The average difference in the (observed) potential outcomes would estimate the population average treatment effect.
- The average of a potential outcome in the population is known as the potential-outcome mean (POM) for a treatment level.
  - The ATE is a difference in POMs

\[
ATE = POM_{participate} - POM_{noparticipate} \\
= E[t_i(\text{partic})] - E[t_i(\text{nopartic})]
\]

- \(t_i(\text{partic})\) is the time to rearrest when person \(i\) participates and
- \(t_i(\text{nopartic})\) is the time to rearrest when person \(i\) does not participate.

- The ATE provides a measure of the effect in the units of time in which the time to event is measured.
As good as random

- The conditional independence (CI) assumption says that the potential outcomes must be independent of the treatment assignment process once we condition on observable covariates.
  - The random-assignment methods used with experimental data are useful, because observational-data methods build on them.
  - Instead of assuming that the treatment is randomly assigned, we assume that the treatment is as good as randomly assigned after conditioning on covariates.

- More formally, we only need conditional mean independence which says that after conditioning on covariates, the treatment does not affect the means of the potential outcomes.
A positive probability of treatment

- The predicted probability of treatment must be sufficiently greater than 0 and cannot be too close to certainty
  - Formally, for each possible $x_i$ in the population and each treatment level $t$, $0 < P(t_i = t| x) < 1$
- Could weaken this assumption when calculating the average treatment effect on the treated
  - Individuals cannot be treated with certainty, or formally, $P(t_i = t| x) < 1$
- When the overlap assumption is not satisfied, the estimators perform poorly because
  - For estimators that use regression adjustment, we’re making predictions that are based on little or no data
  - For estimators that use weighting, our weights become very large or very small and thus unstable
Correct censoring model

If our data is censored, then we must assume

1. The censoring time is fixed or that the process that determines when an observation is censored is independent of the outcome, conditional upon covariates
   - Standard assumption in survival analysis
   - For example, we rule out ex-prisoners who anticipate being rearrested soon dropping out of our study early

2. That the method used to adjust for censoring is correct
   - For estimators that do not require an explicit model of the censoring process, this is not any more restrictive than what we assumed for conditional mean independence
   - For estimators that do require a model of the censoring process, this means that we must correctly specify that model
We can broadly divide the survival-time treatment effects estimators by

1. Whether we want to model the outcome, the treatment, or both

2. Whether or not we must specify a model for the censoring process

<table>
<thead>
<tr>
<th>No censoring model</th>
<th>Model</th>
<th>Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>outcome</td>
<td>Regression adjustment (RA)</td>
</tr>
<tr>
<td></td>
<td>outcome &amp; treatment</td>
<td>Inverse-probability weighted RA with likelihood-adjusted censoring (LAC-IPWRA)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Censoring model required</th>
<th>Model</th>
<th>Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>outcome</td>
<td>→</td>
<td>Weighted regression adjustment (WRA)</td>
</tr>
<tr>
<td>treatment</td>
<td>→</td>
<td>Inverse-probability weighted (IPW)</td>
</tr>
<tr>
<td>outcome &amp; treatment</td>
<td>→</td>
<td>Inverse-probability weighted RA with weighting-adjusted censoring (WAC-IPWRA)</td>
</tr>
</tbody>
</table>
Auxiliary models

- Recall that the potential-outcomes framework formulates the estimation of the ATE as a missing-data problem
- We use the parameters of auxiliary models to solve the missing-data problem
  - How we condition on covariates so that the treatment is as good as randomly assigned
  - Also handles the data lost to censoring
- For the following examples, our auxiliary models will be

  **Outcome**  
  Weibull model with covariates $\text{parrest}$ (number of prior arrests) and age

  **Treatment**  
  Logit model with covariates $\text{married}$ (whether or not the former prisoner is married at the time of release), $\text{wrkhis}$ (number of months of work history prior to incarceration), and $\text{educ}$ (years of completed schooling)

  **Censoring**  
  Weibull model with covariates $\text{crime}$ (type of crime for most recent arrest) and age, which we allow to be nonlinear
Regression adjustment estimators

- RA estimators fit separate survival models for each treatment level and then
  1. Estimate each $POM_i$ as the predicted time to event using the estimated coefficients from the model for those who got treatment level $i$
  2. Use differences between the $POM$s to estimate the ATE

- RA estimators model the outcome without any assumptions about the functional form for the probability of treatment model
- Censoring is handled in the log-likelihood functions of the survival models
Example: RA – modeling the outcome

```
. stteffects ra (parrest age) (partic), nolog
Survival treatment-effects estimation                           Number of obs = 4,260
Estimator : regression adjustment
Outcome model : Weibull
Treatment model: none
Censoring model: none

|                  | Coef.  | Robust Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|------------------|--------|------------------|-------|-------|----------------------|
| _t               |        |                  |       |       |                      |
| ATE              |        |                  |       |       |                      |
| partic (1 vs 0)  | 12.2334| .8485777         | 14.42 | 0.000 | 10.57016              |
|                   |        |                  |       |       | 13.89652              |
| POmean partic 0   | 12.34105| .4430541         | 27.85 | 0.000 | 11.47268              |
|                   |        |                  |       |       | 13.20942              |

If no ex-prisoners participated in the supported-work program, then the average time to rearrest would be 12.3 months.
If all ex-prisoners participated in the program, then the average time to rearrest would increase by 12.2 months.
Inverse-probability-weighted estimators

- Inverse-probability-weighted (IPW) estimators
  1. Account for the missing potential outcome by estimating the treatment-assignment model and creating a weight for each observation equal to the inverse of the probability that it is observed.
  2. Account for data lost to censoring by estimating the censoring model and creating a weight that is equal to the inverse of the probability that the observation is censored.
  3. Use both weights to create weighted averages of the outcomes for each treatment level (the POMs).
  4. Use differences between the POMs to estimate the ATE.

- Observations that are not likely to contain missing data get a weight close to one; and the converse.

- IPW do not make any assumptions about the functional form for the outcome model.
Special properties of estimators with a censoring model

- Fixed censoring processes, such as administrative censoring, are precluded
- Time to censoring must be random conditional on covariates
  - We must correctly specify our censoring model
- The random censoring process cannot vary by treatment level
  - These estimators are not appropriate, for example, with risky treatments where the risk is modified by some covariate
- These estimators require the overlap assumption to be extended so that the probability of not being censored is also sufficiently greater than 0 and less than 1
Example: IPW – modeling treatment assignment and censoring

```
. stteffects ipw (partic i.married age wrkhis educ)
   (i.crime c.age##c.age), nolog
```

Survival treatment-effects estimation
Number of obs = 4,260
Estimator : inverse-probability weights
Outcome model : weighted mean
Treatment model: logit
Censoring model: Weibull

| _t  | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|-----|--------|-----------|-------|------|----------------------|
| ATE | partic (1 vs 0) | 12.36087  | 1.119836 | 11.04 | 0.000 | 10.16603 - 14.55571 |
| P0mean | partic 0 | 12.95968  | .5959098 | 21.75 | 0.000 | 11.79172 - 14.12764 |

If no ex-prisoners participated in the supported-work program, then the average time to rearrest would be 13 months.
If all ex-prisoners participated in the program, then the average time to rearrest would increase by 12.4 months.
IPW RA estimators

- Inverse probability weighted regression adjustment (IPWRA) estimators combine the methods of RA with weights derived from a model for treatment assignment and possibly a censoring model, just like IPW.

- IPWRA estimators come in two varieties, distinguished by the method used for censoring:
  1. Likelihood-adjusted censoring IPWRA (LAC-IPWRA) estimators
  2. Weighted-adjusted censoring IPWRA (WAC-IPWRA) estimators

- LAC-IPWRA estimators handle censoring like the RA estimators, by including a term in the log-likelihood function for the outcome.

- WAC-IPWRA estimators handle censoring like the IPW estimators, by adjusting the weights for data lost to censoring.
Fitting the IPWRA model

- To use the **LAC-IPWRA** estimator, we just specify the outcome and treatment model
  
  . stteffects ipwra (parrest age)  
  (partic i.married wrkhis educ)

- To use the **WAC-IPWRA** estimator, we need to also specify the censoring model
  
  . stteffects ipwra (parrest age)  
  (partic i.married wrkhis educ)  
  (i.crime c.age##c.age)

- Stata automatically switches between the two methods for handling censoring for IPWRA estimators
Example: IPWRA – modeling the outcome and treatment assignment

```
.ssteffects ipwra (parrest age) (partic i.married age wrkhis educ), nolog
Survival treatment-effects estimation Number of obs = 4,260
Estimator : IPW regression adjustment
Outcome model : Weibull
Treatment model: logit
Censoring model: none

|     | Coef.      | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|-----|------------|-----------|-------|-------|----------------------|
| _t  |            |           |       |       |                      |
| ATE |            |           |       |       |                      |
| partic (1 vs 0) | 12.41034 | .8772714 | 14.15 | 0.000 | 10.69092 14.12976   |
| POmean |            |           |       |       |                      |
| partic 0 | 12.41359 | .4667908 | 26.59 | 0.000 | 11.4987  13.32849   |
```

If no ex-prisoners participated in the supported-work program, then the average time to rearrest would be 12.4 months.

If all ex-prisoners participated in the program, then the average time to rearrest would increase by 12.4 months.
Example: Our various model specifications

```
. estimates table RA IPW IPWRA_LAC IPWRA_WAC,
    keep(ATE: POmean:) b(%8.1f) se(%8.1f) modelwidth(9)
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>RA</th>
<th>IPW</th>
<th>IPWRA_LAC</th>
<th>IPWRA_WAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r.partic</td>
<td>12.2</td>
<td>12.4</td>
<td>12.4</td>
<td>12.6</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>1.1</td>
<td>0.9</td>
<td>1.1</td>
</tr>
<tr>
<td>POmean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r.partic</td>
<td>12.3</td>
<td>13.0</td>
<td>12.4</td>
<td>12.8</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.6</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

legend: b/se
Testing the overlap assumption: Predictions

Recall that we want the predicted probability of treatment at each treatment level to be sufficiently greater than 0 and sufficiently less than 1.

```stata
. quietly stteffects ipw (partic i.married age wrkhis educ) (i.crime c.age##c.age)
. predict ps0 ps1, ps
. summarize ps0 if partic==0
    Variable |    Obs    Mean    Std. Dev.    Min    Max
-------------|--------|-----------|----------------|--------|--------
      ps0 |   1,605  .4178441   .1242475  .1202521  .8840557
. summarize ps1 if partic==1
    Variable |    Obs    Mean    Std. Dev.    Min    Max
-------------|--------|-----------|----------------|--------|--------
      ps1 |   2,655  .6480752   .1187267  .2051178  .9346817
```

There is no evidence that the overlap assumption is violated.
Testing the overlap assumption: Plots

- We can also examine treatment probabilities using an overlap plot
  . teffects overlap

What we want is for the densities to span the $x$ axis without massing near 0 or 1 and to overlap, as we see here
Testing the overlap assumption: Censoring

- Also recall that we want the predicted probability of treatment at each treatment level to be sufficiently greater than 0 and less than 1

```
predict prnotcens, censurv
summarize prnotcens if rearrest==1
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>prnotcens</td>
<td>3,766</td>
<td>.9036242</td>
<td>.1128106</td>
<td>.2329026</td>
<td>.9999843</td>
</tr>
</tbody>
</table>

The smallest probability of not being censored is 0.233
We might be a little concerned about how close maximum is to 1
Summing up

- The potential-outcomes framework gives us
  1. Estimation techniques that deal with counterfactuals, the inherently missing information about what would have happened had a subject received a different treatment
  2. Effects that are on the same time scale as the time in which the outcome is measured
  3. Population-averaged effects

- Using Stata for treatment-effects analysis of survival-time data gives you
  1. Access to a variety of estimators
  2. A range of postestimation commands to check the assumptions of the models
  3. The world’s greatest manuals for a statistical software package should you have any questions
Now what?

- Go to [TE] tteffects intro advanced for more information and lots of links to literature and examples

- Also check out [TE] stteffects intro for more information about survival-time treatment-effects estimators, including additional specification tests and multivalued treatments
  http://www.stata.com/manuals14/testteffectsintro.pdf
Thank you
The ratio of hazards of unconditional (marginal) distributions is not the same as an average of conditional hazard ratios

\[
\frac{\lambda_{\text{partic}}(t)}{\lambda_{\text{nopartic}}(t)} = \frac{f_{\text{partic}}(t)}{S_{\text{partic}}(t)} \neq \mathbb{E} \left[ \frac{\lambda_{\text{partic}}(t \mid x \beta_{\text{partic}})}{\lambda_{\text{nopartic}}(t \mid x \beta_{\text{nopartic}})} \right]
\]

- $\lambda_{\text{partic}}(t)$ is the unconditional hazard when everyone participates
- $\lambda_{\text{nopartic}}(t)$ is the unconditional hazard when no one participates
- $f_{\text{partic}}(t)$ is the unconditional density when everyone participates
- $f_{\text{nopartic}}(t)$ is the unconditional density when no one participates
- $S_{\text{partic}}(t)$ is the unconditional survival function when everyone participates
- $S_{\text{nopartic}}(t)$ is the unconditional survival function when no one participates
- Have a multistep estimator
- Example based on RA, same logic works for IPW and IPWRA
  1. Model outcome conditional on covariates for treated observations
  2. Model outcome conditional on covariates for not treated observations
  3. Estimate predicted mean survival time of all observations given covariates from treated model estimates
  4. Estimate predicted mean survival time of all observations given covariates from not-treated model estimates
  5. Difference in means of predicted means estimates ATE
Each step can be obtained by solving moment conditions yielding a method of moments estimator known as an estimating equation (EE) estimator

- $m_i(\theta)$ is vector of moment equations and $m(\theta) = 1/N \sum_{i=1}^{N} m_i(\theta)$

The estimator for the variance-covariance matrix of the estimator has the form $1/N(DMD')$ where

$D = \left( \frac{1}{N} \frac{\partial m(\theta)}{\partial \theta} \right)^{-1}$ and

$M = \frac{1}{N} \sum_{i=1}^{N} m_i(\theta)m_i(\theta)$

Stacked moments do not yield a symmetric $D$, so no simplification under correct specification


