

# strbee: Randomisation-based efficacy estimator

Ian R. White<sup>1,3</sup>      Sarah Walker<sup>2</sup>  
Abdel Babiker<sup>2</sup>

<sup>1</sup> MRC Biostatistics Unit  
Institute of Public Health  
Robinson Way  
Cambridge CB2 2SR

<sup>2</sup> MRC Clinical Trials Unit  
222 Euston Road  
London NW1 2DA

<sup>3</sup> Author for correspondence.

Tel: +44-1223 330399

Fax: +44-1223 330388

Email: ian.white@mrc-bsu.cam.ac.uk

March 14, 2001

## **Abstract**

**strbee** analyses a two-group clinical trial with a survival outcome, in which some subjects may ‘cross over’ to receive the treatment of the other arm. Adjustment for treatment cross-over is done by a randomisation-respecting method which preserves the intention-to-treat P-value.

Keywords: clinical trials, treatment changes, randomisation-respecting.

## Description

`strbee` analyses a two-group clinical trial with a survival outcome, in which some subjects may “cross over” to receive the treatment of the other arm. The method is due to Robins and Tsiatis (1991) and has been developed by White *et al.* (1997, 1999). The observed event time  $T$  is related to an underlying event time  $U$  that would have been observed in the absence of treatment through an accelerated life model. The parameter  $\psi$  of this model represents the factor by which life is accelerated by treatment and is estimated as the value at which  $U$  is balanced between the treatment groups (on a user-specified test).

Estimation is performed by searching over a range of values of  $\psi$  and computing the test statistic  $Z(\psi)$ . Censoring that is non-informative on the  $T$ -scale becomes informative on the  $U$ -scale: to cope with this, `strbee` performs recensoring using *potential censoring times*.

## Syntax

Syntax 1:

```
strbee treatvar [if exp] [in range],  
[test(logrank|wilcoxon|cox|weibull|exponential)  
xo0(timevar eventvar) xo1(timevar eventvar) endstudy(varname)  
savedta(filename[,append|replace])  
psimin(#) psimax(#) psistep(#) tol(#)  
trace list graph graph_options level(#)]
```

Syntax 2:

```
strbee using filename, [list graph graph_options level(#)]
```

`strbee` is an `st` command - data must be `stset` before running `strbee`.

Each subject must have a single record starting at time 0.

`treatvar` is the treatment arm, which must have values 0 and 1.

Estimation is performed by searching over a range of values of  $\psi$  and computing the test statistic  $Z(\psi)$ .

Censored data present an extra problem for **strbee**, since non-informative censoring of T implies informative censoring of U. To avoid bias, users must specify the potential censoring time for all subjects: **strbee** then computes a re-censoring time, which may be earlier than the actual censoring time, and recensors the data.

Syntax 1 is the standard syntax. Results for  $\psi$  and  $Z(\psi)$  may be stored or appended to an existing results file.

Syntax 2 works from previously stored results for  $\psi$  and  $Z(\psi)$ .

## Options

### Specifying the model

**xo0(timevar eventvar)** specifies time to cross-over and a cross-over indicator in arm 0. If not specified then it is assumed that there is no cross-over from arm 0 to arm 1.

**xo1(timevar eventvar)** does the same for arm 1. If not specified then it is assumed that there is no cross-over from arm 1 to arm 0.

**endstudy(varname)** specifies the time of the end of study (the potential censoring time). U-values are then recensored at the minimum possible potential censoring time on the U-scale, where the minimisation is carried out over all treatment profiles possible for the subject's randomised arm.

Note that the potential censoring time must be specified both for censored and uncensored subjects.

Censoring due to random events (e.g. competing risk or loss to follow-up) should ideally be treated differently from censoring at end of study. However, if small amounts of random censoring are present, a reasonable approximation

is to set the potential censoring time for subjects who are randomly censored equal to the actual time of random censoring.

`test(logrank|wilcoxon|cox|weibull|exponential)` specifies the test used to compare U in the two randomised arms. The default is the log rank test. Cox, weibull and exponential use the Wald tests from `stcox`, `streg`, `dist(wei)` and `streg`, `dist(exp)` respectively.

## Specifying the estimation procedure

`psimin(#)` and `psimax(#)` specify extreme values of the parameter  $\psi$ . Defaults are -1 and 1. With syntax 2 they only affect the values plotted on the graph.

`psistep(#)` with  $#>0$  specifies the step size between the extreme values for a grid search.

`psistep(0)` specifies an interval-bisection approach for  $\psi$ .  $Z(\psi)$  is evaluated at `psimin` and `psimax`. If  $Z(\text{psimin}) < 0$  or  $Z(\text{psimax}) > 0$  then  $\psi$  can't be estimated - a wider range of  $\psi$  is probably required. Otherwise,  $Z(\psi)$  is evaluated at  $(\text{psimin} + \text{psimax}) / 2$  and  $\psi$  is narrowed down to the appropriate half of the interval. After convergence, the procedure is repeated to find the upper and then the lower confidence limit.

Interval-bisection may give wrong answers if the test statistic is non-decreasing in  $\psi$ , and should always be checked using the graph option.

`tol(#)` specifies the convergence criterion for interval-bisection estimation. `strbee` searches until bounds for each solution differ by less than  $10^{-\text{tol}}$ . `tol` also defines the number of decimal places reported. The default is `tol(3)`.

`noci` suppresses searching for the confidence limits. `strbee` still reports what limited information it has about them.

## Controlling storage of results

`savedta(filename[, append|replace])` directs the values of  $\psi$  and the test statistic to the specified file.

If `append` is specified, `strbee` checks that the current test statistics and the stored test statistics were computed using the same model. In this case the results reported only use test statistics computed in this run of `strbee`: to get full results, run `strbee` using `filename`.

## Controlling output

`trace` gives details on recensoring and prints  $\psi$  and the test statistic at each step. It also saves a file `rbeetrace.dta` containing variables `u du z0 dz0 z1 dz1` and `recens` evaluated at the last value of  $\psi$  used - note that with grid search this is the value of `psimax`, and with interval-bisection search it is the upper confidence limit without `noci` or the point estimate with `noci`.

`graph` graphs the test statistic against  $\psi$ . All graph options are allowed.

`list` lists the values of  $\psi$  and the test statistic.

## Example

We will illustrate `strbee` using simulated data based a randomised controlled trial (Concorde Coordinating Committee, 1994). Patients with asymptomatic HIV infection were randomised to immediate treatment with zidovudine or to deferred treatment with zidovudine. The outcome is time to disease progression or death.

An intention to treat analysis estimates the effect of applying the policy of immediate zidovudine compared with deferred zidovudine. However, some patients in the deferred arm started zidovudine before symptoms of HIV on the basis of low CD4 cell counts (a prognostic marker for disease progression).

Therefore the intention to treat estimate of treatment effect is biased towards the null as an estimate of the efficacy of zidovudine.

## Data

The data are stored in file `immdef.dta`.

```
. use immdef, clear
(Example data for strbee)

. summ

Variable |      Obs        Mean    Std. Dev.       Min       Max
-----+-----+-----+-----+-----+-----+
      id |     1000      500.5    288.8194          1      1000
     def |     1000        .5    .5002502          0          1
     imm |     1000        .5    .5002502          0          1
censyrs |     1000      2.2452   .4529483          1.5          3
      xo |     1000       .189    .3917045          0          1
  xoysrs |     1000     .7770897   .9340944          0          3
     prog |     1000       .312    .4635417          0          1
 progyrs |     1000      1.9325   .6633275     .0215206          3

. l id imm xo xoysrs prog progyrs in 1/10, noobs nodis
```

id	imm	xo	xoysrs	prog	progyrs
1	1	0	0	0	3
2	0	1	2.652797	0	3
3	1	0	0	1	1.737838
4	1	0	0	1	2.166291
5	0	1	2.1221	1	2.884646
6	0	1	.557392	0	3
7	0	0	2.18947	1	2.18947
8	1	0	0	1	.9226239
9	1	0	0	0	3
10	1	0	0	0	3

For example, subject 5 in the deferred arm started treatment (“crossed over”) at 2.12 years and progressed at 2.88 years. Subject 2 crossed over at 2.65 years and was censored for progression at 3 years. Subject 3 had

cross-over censored at time 0 (as did all subjects in the immediate arm) and progressed at 1.74 years.

## Intention-to-treat analysis

Firstly we estimate the intention to treat estimate of the effect of the policy of giving immediate zidovudine from the accelerated failure time model, ignoring treatment changes. An unweighted logrank test using `sts test` gives a borderline significant P-value of 0.056. Using `strbee`, the default method to test for equality of underlying survival times ( $U$ ) is also the unweighted logrank test (equivalent to the Cox proportional hazards score test).

```
. stset progyrs, failure(prog)
(output omitted)

. strbee imm
imm=0: 500 subjects, 0 cross-overs
imm=1: 500 subjects, 0 cross-overs
Search method: interval bisection
.....
Estimating psi from accelerated life model
Method: Robins-Tsiatis method
Test: logrank
Cross-over in arm 0: none
Cross-over in arm 1: none
Treatment variable: imm
End of study variable: none
-----
          psi      95% confidence interval
-----
best      -0.147      -0.302       0.002
lower bound -0.147      -0.302       0.001
upper bound -0.146      -0.301       0.002
-----
```

The ITT estimate is -0.147, meaning that lifetime is used up  $\exp(-0.147) = 0.863$  times as fast when on treatment as when off treatment. This has been called an acceleration factor.

Note that the confidence interval just spans 1, consistent with the P-value from `sts test`.

## Adjusting for cross-overs

The effect of patients in the deferred zidovudine arm starting open label zidovudine is to dilute the estimated efficacy of zidovudine from the intention-to-treat analysis. Estimating the efficacy of zidovudine adjusting for these changes from the control group (`imm=0`) increases the size of the effect as expected.

```
. strbee imm, xo0(xoys xo)
imm=0: 500 subjects, 189 cross-overs
imm=1: 500 subjects, 0 cross-overs
Search method: interval bisection
.....
Estimating psi from accelerated life model
Method: Robins-Tsiatis method
Test: logrank
Cross-over in arm 0: xoys xo
Cross-over in arm 1: none
Treatment variable: imm
End of study variable: none
-----
          psi      95% confidence interval
-----
best      -0.185      -0.367       0.004
lower bound -0.186      -0.367       0.004
upper bound -0.185      -0.366       0.005
-----
```

The point estimate is now -0.185, further from the null than the intention-to-treat estimate, but the upper confidence limit remains just above 0, reflecting the fact that the P-value is unchanged by this procedure.

In some situations, we might want to adjust for treatment crossovers from both arms, which can be specified separately using `xo0()` and `xo1()`.

## Allowing for censoring

The trial lasted 3 years with staggered entry over the first 1.5 years. Variable `censyrs` gives the time from entry to the end of the 3 years.

So far, censoring indicators on the progression time scale in the trial (T) scale have been passed over to the underlying failure time (U) scale. Unfortunately, uninformative censoring on the T scale may be informative on the underlying U time scale, because patients receiving the superior treatment have their underlying lifetime U extended so that they are censored, whereas patients with the same underlying failure time U receiving the inferior treatment may have the event observed. Thus, on the U scale, censoring is informative with respect to the treatment group. One solution is to “recensor” the underlying failure times, U, by the minimum underlying failure time that could have been observed for each patient across all possible history of treatment changes. This requires passing to `strbee` the potential censoring time for each patient, which is the time from trial entry to the planned end of study. Recensoring is therefore associated with a loss of information, but a smaller bias in estimating the treatment effect.

For our data, the trial lasted 3 years with staggered entry over the first 1.5 years. The potential censoring time is variable `censyrs` which ranges from 3 years for the first patients to 1.5 years for the last patients.

```
. l id imm censyrs xo xoysr prog progyrs in 1/10, noobs nodis
```

<b>id</b>	<b>imm</b>	<b>censyrs</b>	<b>xo</b>	<b>xoyrs</b>	<b>prog</b>	<b>progyrs</b>
1	1	3	0	0	0	3
2	0	3	1	2.652797	0	3
3	1	3	0	0	1	1.737838
4	1	3	0	0	1	2.166291
5	0	3	1	2.1221	1	2.884646
6	0	3	1	.557392	0	3
7	0	3	0	2.18947	1	2.18947
8	1	3	0	0	1	.9226239
9	1	3	0	0	0	3
10	1	3	0	0	0	3

The result of **strbee** with recensoring is shown below. In this case the point estimate is little changed. We also used the **savedta()** option which stores the results for future use.

```
. strbee imm, xo0(xoyrs xo) endstudy(censyrs) savedta(recens)
imm=0: 500 subjects, 189 cross-overs
imm=1: 500 subjects, 0 cross-overs
Search method: interval bisection
.....
Estimating psi from accelerated life model
Method: Robins-Tsiatis method
Test: logrank
Cross-over in arm 0: xoyrs xo
Cross-over in arm 1: none
Treatment variable: imm
End of study variable: censyrs
-----
          psi      95% confidence interval
-----
best      -0.181      -0.350       0.002
lower bound -0.182      -0.351       0.002
upper bound -0.181      -0.350       0.003
-----
file recens.dta saved
```

## Advanced use and checking for multiple cross-overs

In practice we recommend exploring the graph of  $Z(\psi)$  against  $\psi$  in greater detail, especially to check that there really is a unique solution to  $Z(\psi)=0$ .

In doing this we will illustrate some more advanced features of **strbee**.

We will explore the range -0.5 to 0.1 which contains the 95% confidence interval reported above. We will use a grid search, evaluating  $Z(\psi)$  at  $\psi = -0.5(0.02)0.1$ , and draw a graph. We will append the results to those saved above.

```
. strbee imm, xo0(xoys xo) endstudy(censyrs) savedta(recens,append) /*
>      */ psimin(-0.5) psimax(0.1) psistep(0.02) graph /*
>      */ title(RBEE with recensoring)
imm=0: 500 subjects, 189 cross-overs
imm=1: 500 subjects, 0 cross-overs
Search method: grid search
.....  
  
Estimating psi from accelerated life model
Method: Robins-Tsiatis method
Test: logrank
Cross-over in arm 0: xoys xo
Cross-over in arm 1: none
Treatment variable: imm
End of study variable: censyrs
-----  
          psi      95% confidence interval
-----  
best        -0.184      -0.351       0.003
lower bound   -0.200      -0.360       0.000
upper bound   -0.180      -0.340       0.020
-----  
Appending results from recens.dta  
  
2 duplicate records deleted
file recens.dta saved
```

The point estimate is different from before only because of the coarse grid

used —  $\psi$  is really only estimated as lying between -0.2 and -0.18. We can get the best results from the combined analysis:

```
. strbee using recens, graph psimin(-0.5) psimax(0.1)
(Example data for strbee)
```

Estimating psi from accelerated life model

Method: Robins-Tsiatis method

Test: logrank

Cross-over in arm 0: xoysr xo

Cross-over in arm 1: none

Treatment variable: imm

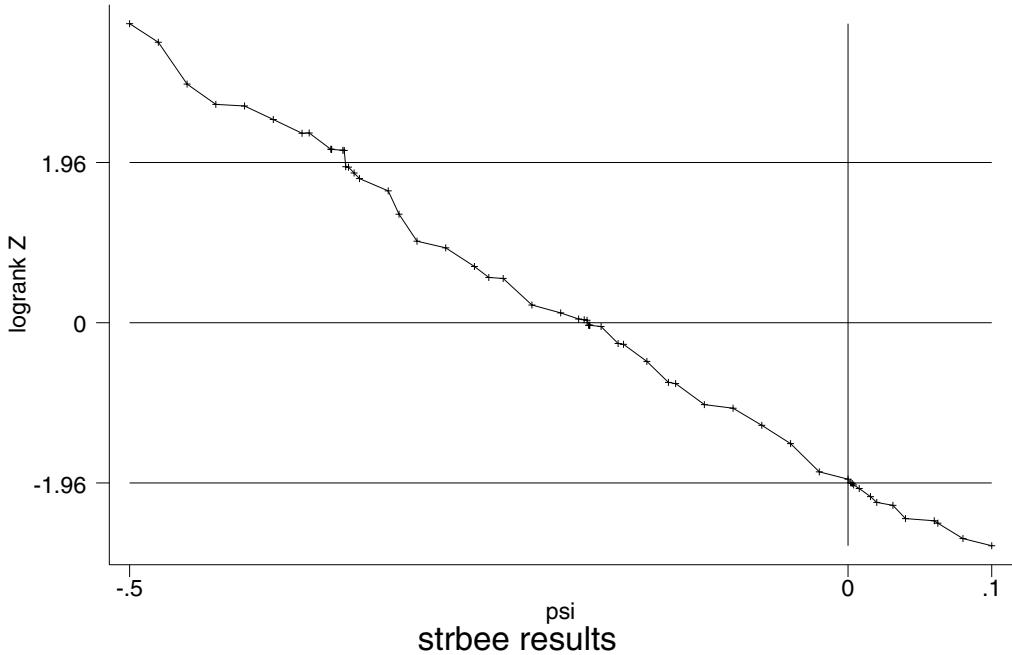
End of study variable: censyrs

	psi	95% confidence interval	
best	-0.181	-0.350	0.002
lower bound	-0.182	-0.351	0.002
upper bound	-0.181	-0.350	0.003

The point estimate is the same as before. Of greater interest is the graph of  $Z(\psi)$  against  $\psi$ , which shows no evidence of non-monotonicity.

## Technical notes

The method used for the interval-bisection search is as follows. **strbee** first evaluates  $Z(\psi)$  at  $\psi = \text{psimin}$  and  $\text{psimax}$ . If they have the same sign then the procedure stops because there is probably no solution in this range. Otherwise **strbee** next evaluates  $Z(\psi)$  at  $\psi = (\text{psimin} + \text{psimax})/2$ , and it continues to bisect intervals until the interval width is narrow enough. It then uses the already computed values of  $Z(\psi)$  to put bounds on the lower confidence limit, and uses another interval-bisection search between these bounds. Finally it repeats the same procedure for the upper confidence limit.



If repeated solutions of  $Z(\psi) = 0$  exist then there is considerable uncertainty about the best point estimate. **strbee** uses an approach proposed by White *et al.* (1999). For example, suppose  $Z(\psi) < 0$  for  $\psi < 0.2$  and for  $0.3 < \psi < 0.5$ . The total length of the range for which  $Z(\psi) < 0$  is the same as if  $Z(\psi)$  only crossed at  $\psi = 0.4$ , so the point estimate is taken as 0.4.

**strbee** censors any cross-overs which occur after the event at the event time.

**strbee** assumes that censored cross-overs occurring before the event represent no cross-over.

## Limitations

**strbee** can account only for a single change in treatment to the treatment of the opposite trial arm. For methods for data with multiple treatment changes, see White *et al.* (1999). For methods for data with changes to

non-trial treatments, see White and Goetghebeur (1998).

## Saved results

`strbee` saves the following scalars in `r()`:

<code>r(psi)</code>	estimate of $\psi$
<code>r(psiupp)</code>	estimate of upper confidence limit around $\psi$
<code>r(psilow)</code>	estimate of lower confidence limit around $\psi$

## References

- Concorde Coordinating Committee (1994) Concorde: MRC/ANRS randomised double-blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection. *Lancet*, **343**, 871–81.
- Robins, J. M. and Tsiatis, A. A. (1991) Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Commun Statist - Theory Meth*, **20**(8), 2609–2631.
- White, I. R. and Goetghebeur, E. J. T. (1998) Clinical trials comparing two treatment policies: which aspects of the treatment policies make a difference? *Statistics in Medicine*, **17**, 319–339.
- White, I. R., Walker, S., Babiker, A. G. and Darbyshire, J. H. (1997) Impact of treatment changes on the interpretation of the Concorde trial. *AIDS*, **11**, 999–1006.
- White, I. R., Babiker, A. G., Walker, S. and Darbyshire, J. H. (1999) Randomisation-based methods for correcting for treatment changes: examples from the Concorde trial. *Statistics in Medicine*, **18**, 2617–2634.