STATISTICS IN MEDICINE
Statist. Med. 2006; **25**:1619–1622
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2453

LETTER TO THE EDITOR

Estimating treatment effects in randomized trials with treatment switching

From: *Ian R. White*
MRC Biostatistics Unit
Institute of Public Health
Robinson Way
Cambridge CB2 2SR, U.K.

Shao, Chang and Chow (henceforth SCC) propose a new method for estimating the causal effect of treatment in a randomized trial with a survival outcome and time-dependent treatment switches [1]. Their method builds on the work of Branson and Whitehead (henceforth BW) [2] which itself builds on the work of Robins and Tsiatis (henceforth RT) [3]. I will demonstrate that SCC implicitly assume that treatment switchers are comparable with non-switchers, while BW inadequately deal with censoring, so the best available method therefore remains that of RT.

METHOD OF BRANSON AND WHITEHEAD

I will compare the methods of BW and RT for a trial with an immediate treatment arm, in which all subjects start treatment at randomization, and a deferred treatment arm, in which some subjects start treatment after randomization.

Both methods use the same accelerated life model. In this model, if the actual event time is T_R and the time of starting treatment is U, then the (counterfactual) event time in the absence of treatment is

$$T_L = \begin{cases} U + \mathrm{e}^{-\eta_0}(T_R - U) & \text{if} U < T_R \\ T_R & \text{if} U \geqslant T_R \end{cases} \quad (1)$$

where η_0 expresses the true treatment effect. T_L is assumed independent of randomized group.

Both methods estimate the treatment effect by a randomization-based procedure. Let $T_L(\eta)$ be the value of (1) when the unknown η_0 is replaced by a candidate value η. RT estimate η_0 as the value of η for which a rank test statistic for equality of $T_L(\eta)$ across treatment arms has value 0. BW use an iterative parameter estimation procedure (IPE) to find a value of η at which a parametric accelerated life model fitted to the perfect-compliance event times ($T_L(\eta)$ in the deferred arm and $T_L(\eta)\mathrm{e}^{\eta}$ in the immediate arm) yields estimated treatment effect η. This condition implies that a parametric accelerated life model fitted to $T_L(\eta)$ yields

estimated treatment effect 0. Thus the estimands of the two methods differ only by the use of a parametric test instead of a rank test.

The methods do differ in their estimation algorithm. RT recommend a grid-based search, whereas the IPE algorithm typically converges much faster. In my view, this important innovation is BW's major contribution.

The methods also differ in their treatment of censoring. Censoring of T_R at a censoring time C_R induces censoring of $T_L(\eta)$ at a time $C_L(\eta)$ which may be computed by replacing T_L with $C_L(\eta)$, T_R with C_R, and η_0 with η in equation (1). RT demonstrated that non-informative censoring of T_R induces informative censoring of $T_L(\eta)$ if U is associated with $T_L(\eta)$, which is highly plausible in practice. They therefore proposed *recensoring* $T_L(\eta)$ in the deferred treatment arm at its earliest possible censoring time. We consider three cases:

1. If $\eta<0$ (treatment is harmful, so $T_L(\eta)\geqslant T_R$) then the smallest possible value of $C_L(\eta)$ is C_R. In this case, values of $T_L(\eta)$ may be projected beyond C_R, and such values should be recensored at C_R.
2. If $\eta=0$ then $T_L(\eta)=T_R$, $C_L(\eta)=C_R$, and no recensoring is needed.
3. If $\eta>0$ (treatment is beneficial, so $T_L(\eta)\leqslant T_R$) then the smallest possible value of $C_L(\eta)$ is $C_R e^{-\eta}$. In this case, values of $T_L(\eta)$ between $C_R e^{-\eta}$ and C_R should be recensored at $C_R e^{-\eta}$.

BW implement this recensoring if $\eta<0$ but not if $\eta>0$. Their method is therefore potentially affected by informative censoring whenever the point estimate $\hat{\eta}>0$, and this is likely to induce bias when $\eta_0>0$. Such bias is not seen in their simulation study (their Table IV), possibly because only 10% of subjects were censored.

To explore the bias in the IPE procedure, I simulated data from the same model as in BW's Table IV, with probability of switching $=0.7$, $a=2$, and $b=4$. Following BW, I explored the bias and variance of $e^{\hat{\eta}}$ with 10% censoring (Table I). In agreement with the results of BW, there is no evidence of bias under any approach to recensoring. However, under recensoring only for $\eta<0$, my variance estimates are substantially larger than those of BW. Two observations suggest that BW's simulated variances are wrong in this case. Firstly, they are smaller than the corresponding variances for uncensored data (reported in BW's Table I). Secondly, they are smaller than the variances without recensoring even when $e^{\eta_0}=2$, despite the fact that all simulated $\hat{\eta}>0$ so that recensoring is never actually done.

I then repeated the simulation with 80% censoring, as is common in many clinical trials (Table II). Here, failing to recensor causes substantial bias towards the null, and this bias is

Table I. Simulation study with 10% censoring: results for $e^{\hat{\eta}}$ (results of Branson and Whitehead in brackets).

True e^{η_0}	Method	Mean	Variance
2	IPE uncensored	2.02 (2.00)	0.069 (0.066)
	IPE recensored if $\eta<0$	2.02 (2.01)	0.069 (0.041)
	IPE recensored if $\eta\neq 0$	2.03	0.073
0.5	IPE uncensored	0.508 (0.504)	0.0045 (0.0041)
	IPE recensored if $\eta<0$	0.501 (0.506)	0.0044 (0.0024)
	IPE recensored if $\eta\neq 0$	0.501	0.0044

Table II. Simulation study with 80% censoring: results for $e^{\hat{\eta}}$ and $\hat{\eta}$.

		$e^{\hat{\eta}}$		$\hat{\eta}$	
True e^{η_0}	Method	Mean	Bias	Mean	Bias
2	Intention-to-treat	1.73	−0.27	0.53	−0.17
	IPE unrecensored	1.83	−0.17	0.58	−0.11
	IPE recensored if $\eta<0$	1.83	−0.17	0.58	−0.11
	IPE recensored if $\eta\neq 0$	2.08	0.08	0.69	0.00
0.5	Intention-to-treat	0.60	0.10	−0.53	0.17
	IPE unrecensored	0.66	0.16	−0.43	0.26
	IPE recensored if $\eta<0$	0.53	0.03	−0.69	0.00
	IPE recensored if $\eta\neq 0$	0.53	0.03	−0.69	0.00

of similar magnitude to the bias in intention-to-treat estimation. The bias in failing to recensor is towards the null because the simulation assumes a positive association between T_L and U (switchers have worse prognosis); a negative association between T_L and U would induce a bias away from the null [4]. Full recensoring has a small bias on the e^{η} scale, but this bias disappears on the more appropriate η scale. Recensoring only when $\eta<0$, as recommended by BW, is the same as full recensoring (and hence unbiased) when $e^{\eta_0}=0.5$, but is the same as not recensoring (and hence biased) when $e^{\eta_0}=2$. Previous work has also demonstrated the need for full recensoring [4].

METHOD OF SHAO, CHANG AND CHOW

SCC claim that the model of BW 'does not take into account the fact that treatment switch is typically based on prognosis and/or investigator's judgment'. This is a misleading statement. The estimation procedures of RT and BW are free of assumptions about the association between treatment switching and prognosis (defined as T_L). However, SCC appear to mean that the model of BW does not allow treatment switches to be associated with the individual benefit to be gained from treatment. They therefore extend the model by adding a term $w_{0,\eta}(s)$ which allows the causal effect of treatment to differ between those who receive treatment by switching from control and those who receive treatment by randomization.

The problem with SCC's approach is their method of estimation. Both their parametric and semi-parametric methods are based on the likelihood *conditional on the time of switching*. This conditional likelihood approach is only valid if switching is ignorable (that is, if switching treatment is independent of prognosis). Their method is thus potentially biased when switching is associated with prognosis, as commonly occurs [5].

SCC's simulation study does not reveal a problem with their estimation method because they generate switching times independently of prognosis. A hint of the possible problems is seen in the last column of their Table I, where incorrect model specification yields an estimated effect of the opposite sign to the correct value.

CONCLUSIONS

BW present an efficient algorithm for fitting the model of RT, but their recensoring algorithm is flawed. The method of SCC is potentially highly biased unless switching is ignorable. For data analysis, the best available method therefore remains that of RT with full recensoring. A Stata program `strbee.ado` implements the RT method [6], and an updated version which optionally implements the IPE algorithm (including the correct recensoring) is available from http://www.mrc-bsu.cam.ac.uk/pub/software/stata.

The problems of the BW and SCC methods were obscured by simulation studies. Future simulation studies of new methods for dealing with treatment switches must always consider the situation where switching is non-ignorable.

REFERENCES

1. Shao J, Chang M, Chow SC. Statistical inference for cancer trials with treatment switching. *Statistics in Medicine* 2005; **24**:1783–1790.
2. Branson M, Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. *Statistics in Medicine* 2002; **21**:2449–2463.
3. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics—Theory and Methods* 1991; **20**(8):2609–2631.
4. White IR, Babiker AG, Walker S, Darbyshire JH. Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. *Statistics in Medicine* 1999; **18**:2617–2634.
5. Horwitz RI, Horwitz SM. Adherence to treatment and health outcomes. *Annals of Internal Medicine* 1993; **153**:1863–1868.
6. White IR, Walker S, Babiker A. strbee: randomisation-based efficacy estimator. *Stata Journal* 2002; **2**:140–150.

