Estimating efficacy in trials with selective crossover

Adam R. Brentnall,*† ⓘ   Peter Sasieni and Jack Cuzick

When one arm in a trial has a worse early endpoint such as recurrence, a data-monitoring committee might recommend that all participants are offered the apparently superior treatment. The resultant crossover makes it difficult to measure differences between arms thereafter, including for longer-term endpoints such as mortality and disease-specific mortality. In this paper, we consider estimators of the efficacy of treatment on those who would not cross over if randomised to the apparently inferior arm. Binomial and proportional hazards maximum likelihood estimators are developed. The binomial estimator is applied to analysis of a breast cancer treatment trial and compared with intention-to-treat and inverse probability weighting estimators. Full and partial likelihood proportional-hazard model estimators are assessed through computer simulations, where they had similar bias and variance. The new efficacy estimators extend those for all-or-none compliance to this important problem. © 2017 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

Keywords: binomial model; causal inference; compliance; proportional hazards model; switching

1. Introduction

Unplanned crossover occurs in randomised trials when participants decide to switch treatment arms. It makes it more difficult to measure a difference between arms, especially when there is selective crossover because those who switch have a different background risk than others. In this paper, we consider trials with selective crossover in one arm arising due to results from early efficacy endpoints, such as when initial results from a trial or other concurrent trials lead a data-monitoring committee to recommend that all participants are offered the apparently superior treatment. This form of crossover is represented using a lexis diagram in Figure 1. It occurred in the BIG-1 98 trial, where women were randomised to receive 5 years of either letrozole or tamoxifen to prevent breast cancer recurrence. Initial results [1] led the organisers to inform women in the tamoxifen monotherapy arms of their treatment, in order to allow an informed decision about future care; one quarter of those in the tamoxifen arm switched to letrozole [2].

The traditional inferential method for a trial with selective crossover is an intention-to-treat (ITT) analysis. This maintains the randomised balance of all causal factors other than the treatment, but it is likely to attenuate the estimated effect of treatment, because some participants in the control arm receive the intervention. A per-protocol approach censors those who cross over in the analysis, but it is unreliable when crossover is informative because bias may occur in either direction and inference is compromised [3]. Instead of these, we focus in this article on the effect of treatment on those who would not cross over when offered if randomised to control. Although this may also be biased for the effect of treatment on everyone in the trial, it is still of interest to identify the effect of treatment on those who would receive it for the duration, and these are the only individuals who provide information on treatment effects over the complete follow-up period.

Efficacy is defined more formally using a model with two latent strata. The first latent stratum contains insistors, who would cross over to treatment when offered if randomised to control; the second consists of 'ambivalents', who would not cross over when offered if randomised to control. The strata are shown

Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, Charterhouse square, London EC1M 6BQ, U.K.
*Correspondence to: Adam R. Brentnall, Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, Charterhouse square, London EC1M 6BQ, U.K.*
†E-mail: a.brentnall@qmul.ac.uk
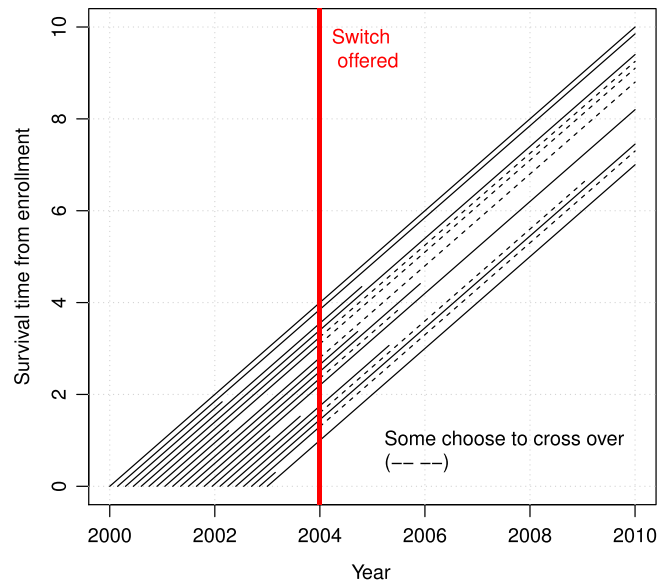
Figure 1. A lexis diagram representing survival and crossover. Each diagonal line is a person followed until they have the event and the line stops; censoring only occurs after 10-year calendar time for clarity. [Colour figure can be viewed at wileyonlinelibrary.com]
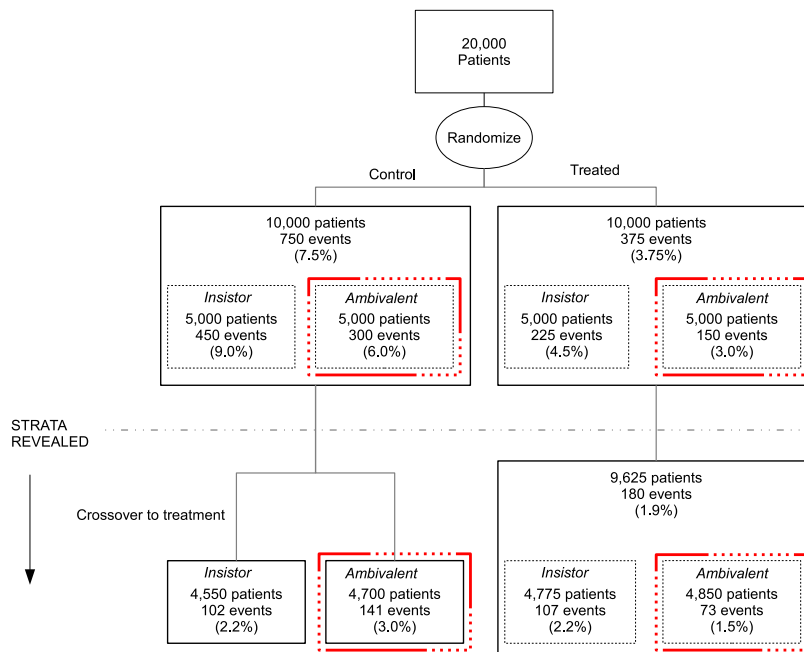
Figure 2. Hypothetical example to show the assumed latent strata in both trial arms when the endpoint is disease-free survival. Each solid box contains observable data; efficacy compares ambivalents in the line-dot boxes (— - - -). [Colour figure can be viewed at wileyonlinelibrary.com]

using an example in Figure 2. No one in the treatment arm switches to control, so there are assumed to be no individuals who defy randomisation by taking the opposite treatment, or who would always switch from treatment to control. The latent strata are unobserved in both arms prior to crossover, and only partially revealed in the control arm. Efficacy measures the average difference between the ambivalents in the two treatment arms.

Figure 2 is a suitable schematic for analysis of disease-free survival such as in the BIG-1 98 trial, but it does not adequately show subtle complications that arise for a mortality endpoint. Figure 3 illustrates the flow of patients under the model in the control arm when death is also included. The additional issue is that the underlying strata of participants in the control arm are not revealed after unblinding for those
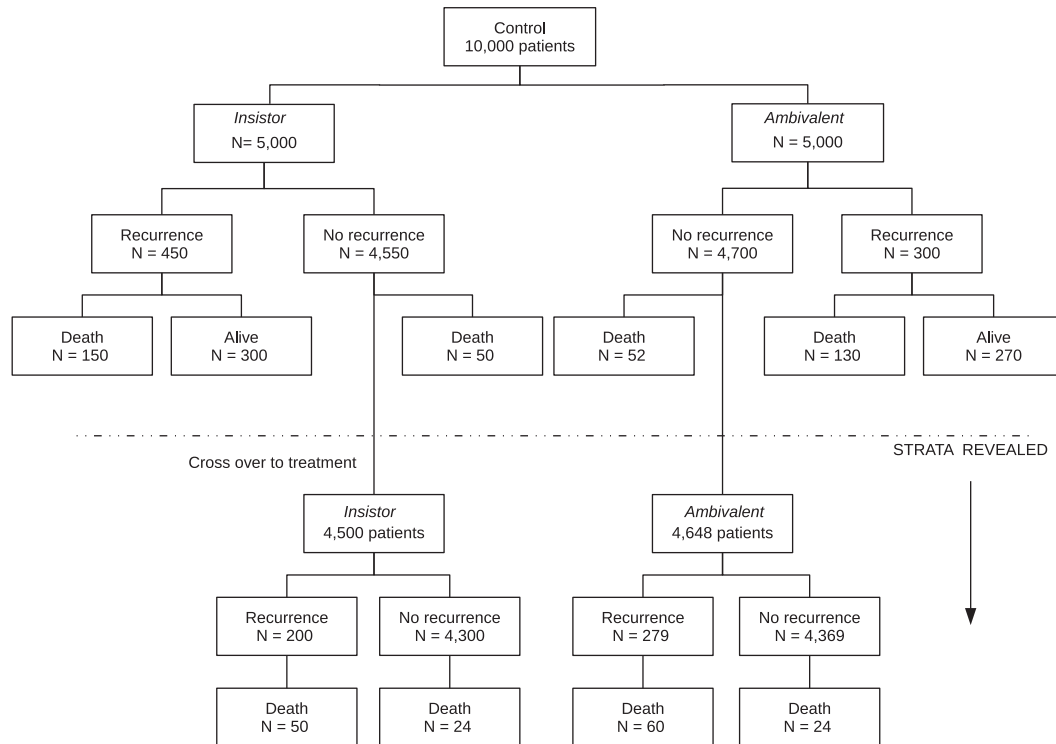
Figure 3. Hypothetical example to show assumed latent class structure in control when the endpoint is mortality.

still alive but whose disease progressed beforehand. If this information was ignored, then the insistors in control with a recurrence before the offer to switch would be incorrectly identified as ambivalent.

We view the model as a plausible description of the trial population when there is a difference in underlying risk after crossover between those who switch treatment and the entire treatment arm. Models with similar latent strata have been applied to different problems [4], including when they are revealed at baseline [5–7] and when they are revealed after baseline [8]. A consequence of the two-strata model is that the proportion of insistors at risk is likely to be different at baseline from the time of potential crossover.

The unobserved boxes in Figure 2 show why the estimation problem is difficult. The marginal effect of treatment may be estimated before crossover by comparing the two arms without adjustment. After crossover, insistors in control receive the treatment. If the ambivalents were observed in both arms after the offer, then estimation of efficacy would be straightforward. The problem is that they are missing data before crossover in control and all the time in treatment. The remainder of the article considers this estimation problem.

2. Binomial model

A binary model may be applied to analysis of a trial when events are rare and little information is contained in the follow-up or event time of patients. It also provides insight into assumptions and parameter identifiability aspects. We next focus on the model illustrated by Figure 2, but the methods may be adapted to mortality endpoints such as in Figure 3 by ensuring that the latent strata are unknown throughout follow-up for those with a recurrence before the opportunity to cross over.

2.1. Assumptions

The first assumption is as follows:

(A1) The potential time of crossover (time when the patient would cross over if an insistor and randomised to control) is observable for all participants in both arms.

This holds for the type of crossover shown by Figure 1. Using A1, we may take two time periods before and after the offer (potential crossover time) and parameterise the probability p_{trk}

of an event in period $t = 0, 1$ (before, after the offer), treatment arm $r = 0, 1$ (control, treatment) and latent insistor stratum $k = 0, 1$ (ambivalent, insistor) as follows. For ambivalents ($k = 0$) in control ($r = 0$) and periods $t = 0, 1$

$$p_{t00} = \alpha_t,$$

where $0 \leqslant \alpha_t \leqslant 1$ is an unknown probability of the event (if a constant rate $\alpha^* \geqslant 0$ is assumed and the time at risk T_t is known, then $\alpha_t = 1 - \exp(-\alpha^* T_t)$, so that α_t in both periods only depends on a single unknown parameter). Efficacy γ_t in period $t = 0, 1$ is defined as the ratio between the treated and untreated ambivalents

$$\gamma_t = p_{t10}/p_{t00}.$$

Insistors in the treatment arm are taken to differ from the ambivalents in period $t = 0, 1$ through

$$\omega_t = p_{t11}/p_{t10}.$$

We also assume the following:

(A2) The treatment effect prior to the offer is the same for the ambivalents and insistors, so that $p_{011}/p_{001} = \gamma_0$, or equivalently

$$p_{001}/p_{000} = \omega_0.$$

A pragmatic reason for taking assumption A2 is that it ensures that the conditional treatment effect among ambivalents before the offer matches the marginal effect (seen hereafter in Equation (2)), which would be the primary measure for an ITT analysis of a trial without selective crossover.

After the offer to switch, we allow a different treatment effect for the insistors in control through

$$p_{101} = \alpha_1 \omega_1 \gamma_1^*,$$

where γ_1^* need not equal γ_1. For example, a treatment crossover effect may exist, such as resulting from receiving a drug in control then the different drug from the treatment arm (which would not have been the case at baseline for those randomised to treatment). Alternatively, this might represent a difference in the treatment effect for insistors and ambivalents, and so testing whether $\gamma_1^* = \gamma_1$ is a way to partially verify A2, albeit after the offer. We require two further assumptions:

(A3) The probability π of being in strata $k = 1$ (insistor) at baseline is the same in both arms because of randomisation.

(A4) Censoring is independent of strata $k = 0, 1$.

A4 may be partly verified by using data in the control arm after the offer to switch and inspecting censoring in the other observed groups prior to and after the offer. Equivalent versions of A3 and A4 have been used by all-or-none compliance estimators [7].

2.2. Identifiability

Six observable statistics depend on the model parameters. These are the total number of events in arm $r = 0, 1$ in the first period y_{0r}, the number of insistors n_{101} (or ambivalents n_{100}) in control after the offer, the number of events in the second period in control ambivalents y_{100} and insistors y_{101}, and the number of events in the treatment arm in the second period y_{11}. The number censored d_{0r} in the first period in each arm r is an ancillary statistic, and binomial distributions are conditional upon the number n_{tr} in each arm r at risk at the start of each period t. The expected values of the observable statistics relate to the model as follows. Before the offer, the expected proportion of events in control is

$$E(Y_{00})/n_{00} = \pi \alpha_0 \omega_0 + (1 - \pi)\alpha_0, \tag{1}$$

and in the treatment arm it is

$$E(Y_{01})/n_{01} = \gamma_0 E(Y_{00})/n_{00}. \tag{2}$$

At the offer to cross over, we observe the number of insistors still at risk in control n_{101}, where

$$E(N_{101}) = \pi n_{00} - E(Y_{00})\left\{ \frac{\pi \omega_0}{(1 - \pi) + \pi \omega_0} \right\} - d_{00}\pi, \tag{3}$$

which subtracts the expected number of events and censored (assumption A4) in insistors from the expected number enrolled in control at baseline. The term compounded by $\mathrm{E}(Y_{00})$ is the probability that the events are insistors. In the treatment arm, the number of insistors at risk at the offer N_{111} is unobserved, but the expected value has the same form as Equation (3) with $r = 1$ in the subscripts rather than $r = 0$. Thus, in the first period, there are three equations but four unknowns $(\pi, \alpha_0, \omega_0, \gamma_0)$. The only parameter that is uniquely identifiable without further constraints is the treatment effect γ_0, replacing expectations by observed values in Equation (2). After the offer to switch the expected proportion of events in three observed groups (cf. Figure 2) is

$$\mathrm{E}(Y_{100})/n_{100} = \alpha_1 \tag{4}$$

$$\mathrm{E}(Y_{101})/n_{101} = \gamma_1^* \alpha_1 \omega_1 \tag{5}$$

$$\mathrm{E}(Y_{11})/n_{11} = \gamma_1 \{\pi_{11}\alpha_1\omega_1 + (1 - \pi_{11})\alpha_1\}, \tag{6}$$

where π_{11} is the probability that individuals at risk in the treatment arm after the offer are insistors.

These equations help to assess identifiability of the model. Replacing the expectation with the observed value in Equation (4) provides an estimate of α_1. There is one degree of freedom for $\hat{\omega}_0$ and $\hat{\pi}$ (π may be expressed as a polynomial function of ω_0 via Equation (3) and $\hat{\pi}_{11} = \hat{\mathrm{E}}(N_{111})/n_{11}$). It might be important to allow for different rates in both periods, so that α_0 and α_1 cannot be tied together through an constant-rate assumption. For example, risk of death increases with age. After incorporating $(\alpha_0, \alpha_1, \omega_0, \gamma_0)$, we have three unknowns $(\omega_1, \gamma_1, \gamma_1^*)$ and two degrees of freedom. However, this is not very restrictive because our a priori assumption is that $\gamma_0 = \gamma_1 = \gamma_1^*$, partly because we are interested in an overall measure of treatment efficacy over the entire period of follow-up, and $\omega_0 = \omega_1$ is also our primary model. The ability to check these assumptions contrasts favourably with an ITT analysis, where it would be very difficult to check whether the treatment effect is constant because any difference might only be caused by crossover.

2.3. Estimation

Estimation might be based on equating observed to expected in Equations (1)–(6), but we recommend to use maximum likelihood estimation, which will be covered in Section 4. For insight into the target of estimation, consider when $\gamma_1^* = \gamma_1$ and rearrange Equation (6) using Equations (4) and (5) so that

$$\gamma_1 = \frac{\mathrm{E}(Y_{11})/n_{11} - \pi_{11}\mathrm{E}(Y_{101})/n_{101}}{(1 - \pi_{11})\mathrm{E}(Y_{100})/n_{100}}. \tag{7}$$

The numerator is a total proportion of events in treatment after crossover minus the proportion of events in insistors after crossover, scaled by the proportion at risk at crossover in treatment. The denominator is the proportion of events in the ambivalents in control after crossover, scaled by the proportion of ambivalents at risk in treatment at crossover. This structure is very similar to for all-or-none compliance [6]. Essentially, estimation of efficacy after crossover subtracts out the insistors from both arms by using the observed insistors from the control arm.

3. Proportional-hazards model

The binomial model extends straightforwardly to an exponential model, but a proportional-hazards model is more flexible, and so we focus on it next. Let the results from the trial for independent individuals $i = 1, \ldots, n$ be (t_i, δ_i) where t_i is the continuous observation time and δ_i is equal to 1 if the event occurred and 0 if right censored at the last follow-up. Denote treatment randomisation indicator r_i to be 1 if subject i was randomised to treatment and 0 if control, and $c_i = 1$ if an insistor and 0 if ambivalent. The potential crossover time is s_i (cf. Figure 1). Then the hazard at time $t \geqslant 0$, conditional on covariates $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{im})$ at baseline, is taken to be of form

$$\lambda(t \mid r_i, s_i, c_i, \boldsymbol{x}_i; \boldsymbol{\theta}) = \lambda_{0i}(t)\psi(t, r_i, s_i, c_i, \boldsymbol{x}_i; \boldsymbol{\theta}), \tag{8}$$

where $\lambda_{0i}(t)$ is a baseline hazard function (with an i subscript because it might be stratified so that $\lambda_{0i} = \lambda_{0j}$ if i and j are in the same strata, or not with $\lambda_{0i} = \lambda_0$), and

$$\psi(t, r_i, s_i, c_i, \boldsymbol{x}_i; \boldsymbol{\theta}) = \exp[\gamma\{r_i + I(t \geqslant s_i)(1 - r_i)c_i\} + \omega c_i + \boldsymbol{x}_i\boldsymbol{\beta}'], \tag{9}$$

where $I(.)$ is the indicator function, prime denoting the transpose, and parameters $\theta = (\gamma, \omega, \beta_1, \ldots, \beta_m)$. The treatment effect parameter is γ, the effect of being an insistor ω, and covariate effects $\beta = (\beta_1, \ldots, \beta_m)$. The formulation is an extension of [7], but where the insistors receive the effect of treatment only after crossing over. It is quite different than treating compliance as a time-dependent covariate, which does not respect randomisation and leads to bias, because it allows for the latent strata in both arms. The model does not include time-varying covariates, but this is unlikely to be restrictive for randomised trials, where even static covariates are not commonly used other than by stratifying the baseline hazard function.

The estimators that we develop in the next section require assumptions A1–A3 from Section 2 and the following.

(A4b) Censoring is independent given covariates and independent of latent strata given covariates.
(A5a) Latent strata are independent of covariates, or
(A5b) latent strata are independent of covariates within defined groups.
(A6) Latent strata are independent of time at entry within defined groups.

A5a is needed for the partial likelihood estimator in the Section 4; the more relaxed A5b is for the full likelihood estimator. Both are not an issue if no covariates are used in the model, and A5a was also used by [7] in their partial-likelihood estimator. A6 is needed for both estimators. The assumptions A5 and A6 may be checked using the data after the offer to switch. If A5b and A6 are not met without groups, then groups might be sought where they are met. As for the binomial estimator, one may test if the treatment effect is constant before and after the offer.

An outline of a full maximum likelihood estimation algorithm is as follows; see Section 4 for technical details. Firstly, the probability each individual is an insistor at baseline is computed given $\hat{\theta}$; then the baseline hazard may be fitted conditional upon this and $\hat{\theta}$; finally, θ are fitted given the baseline hazard and insistor probability estimates. The approach used in this algorithm to estimate the proportion of insistors at baseline (full likelihood) and through time (partial likelihood) is the biggest difference compared with [7]. This is because in the all-or-none compliance case the proportion of switchers at risk through time in the unobserved arm may be estimated simply as the observed number of switchers at risk. However, the new algorithm could also be applied in the all-or-none case where the main practical difference is that the estimate of the number of non-compliers at risk in the unobserved arm would change each time there is an event in that arm, rather than only each time there is an event in the observed arm. The overall profile-likelihood algorithm continues until convergence in the likelihood [9]. The partial likelihood is asymptotically concave because it is the same form as for all-or-none compliance [7], so it offers a good starting point for full maximum likelihood.

Inference may be based on profile likelihood. One approach that is not recommended is to use Wald confidence intervals from the partial likelihood; these would underestimate variability by treating the proportion of insistors at risk through time as known.

4. Maximum likelihood estimation

4.1. Binary model

In the first period in control ($t = 0$; $r = 0$) and the treatment arm in both periods ($t = 0, 1$; $r = 1$), the probability of an event is

$$p_{tr} = \pi_{tr} p_{tr1} + (1 - \pi_{tr}) p_{tr0}, \tag{10}$$

where p_{trk} was defined in Section 2 and $\pi_{0r} = \pi$ is assumed constant for $t = 0$. There are five unknown parameters in the primary model: the baseline probabilities α_0, α_1 (potentially different in both periods), the insistor effect $\omega = \omega_0 = \omega_1$ (common in both periods), the treatment effect $\gamma = \gamma_0 = \gamma_1 = \gamma_1^*$ (common in both periods) and the proportion of insistors at baseline π. The probability of being an insistor at risk in treatment after the offer (π_{11}) is estimated as a function of the data and parameter estimates ($\alpha_0, \omega, \gamma, \pi$), as discussed in Section 2. For the treatment arm ($t = 0, 1$; $r = 1$) and the control arm before the offer ($t = 0$; $r = 0$), the likelihood is

$$L_{tr} \propto p_{tr}^{y_{tr}} (1 - p_{tr})^{n_{tr} - y_{tr}} \tag{11}$$

and after the offer in control

$$L_{10} \propto \prod_{k=0}^{1} p_{10k}^{y_{10k}}(1 - p_{10k})^{n_{10k}-y_{10k}}. \tag{12}$$

The complete likelihood

$$L \propto \prod_{t=0}^{1} \prod_{r=0}^{1} L_{rt} \tag{13}$$

may be used to form maximum likelihood estimates and profile-likelihood confidence intervals in the usual manner.

4.2. Proportional-hazards model

4.2.1. Likelihood function. Before crossover and in the treatment group, we do not observe the latent strata. Thus, if $t_i < s_i$ and $r_i = 0$, or when $r_i = 1$, then the likelihood for individual i is

$$L_i(\boldsymbol{\theta} \mid r_i, s_i, \boldsymbol{x}_i) = \pi_i L_i(\boldsymbol{\theta} \mid r_i, c_i = 1, \boldsymbol{x}_i) + (1 - \pi_i) L_i(\boldsymbol{\theta} \mid r_i, c_i = 0, \boldsymbol{x}_i) \tag{14}$$

where $\pi_i = P(C_i = 1 \mid r_i, \boldsymbol{x}_i)$ is the probability a person is an insistor conditional on baseline covariates $\boldsymbol{x}_i$ and

$$L_i(\boldsymbol{\theta} \mid r_i, c_i, \boldsymbol{x}_i) = \lambda(t_i \mid r_i, c_i, \boldsymbol{x}_i; \boldsymbol{\theta})^{\delta_i} \exp\left\{-\int_0^{t_i} \lambda(u \mid r_i, c_i, \boldsymbol{x}_i; \boldsymbol{\theta}) du\right\}.$$

If the crossover is observed ($t_i \geqslant s_i$ and $r_i = 0$), then L_i is the product of two terms given later in Equation (20). The first contribution is a survivor function $P(t_i > s_i \mid r_i = 0, s_i, \boldsymbol{x}_i)$, which is of the same form as Equation (14); the second contribution when c_i is known ($t_i \geqslant s_i$) is $\lambda(t_i \mid r_i, s_i, c_i, \boldsymbol{x}_i; \boldsymbol{\theta})^{\delta_i} \exp\left\{-\int_{s_i}^{t_i} \lambda(u \mid r_i, c_i, \boldsymbol{x}_i; \boldsymbol{\theta}) du\right\}$. The full likelihood for everyone is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta} \mid r_i, s_i, \boldsymbol{x}_i),$$

and we will write $l(\boldsymbol{\theta})$, or just l, for the log likelihood.

4.2.2. Partial likelihood function. Consider when $\lambda_{0i}(t) = \lambda_0(t)$ so that there are no baseline-hazard groups. Then the partial likelihood is denoted

$$PL(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \frac{\eta(t_i, r_i, s_i, \boldsymbol{x}_i)}{\sum_{j=1}^{n} I(t_j \geqslant t_i) \eta(t_i, r_j, s_i, \boldsymbol{x}_j)} \right\}^{\delta_i}. \tag{15}$$

When the latent stratum is unknown ($r_j = 0$ and $s_j > t_i$, or $r_j = 1$)

$$\eta(t_i, r_j, s_j, \boldsymbol{x}_j) = \pi(t_i \mid r_j)\psi(t_i, r_j, s_j, c_j = 1, \boldsymbol{x}_j) + \{1 - \pi(t_i \mid r_j)\}\psi(t_i, r_j, s_j, c_j = 0, \boldsymbol{x}_j),$$

where $\pi(t_i \mid r_j)$ is the conditional probability of being an insistor ($c_j = 1$) still at risk in arm r_j at time t_i. Note that we follow assumption A5a and assume $\pi(t_i \mid r_j, \boldsymbol{x}_j) = \pi(t_i \mid r_j)$. When crossover status is known ($r_j = 0$ and $s_j \leqslant t_i$), then $\eta(t_i, r_j, s_j, \boldsymbol{x}_j) = \psi(t_i, r_j, c_j, \boldsymbol{x}_j)$. For a model with different baseline hazard functions for defined groups, the product of different partial likelihoods of form (15) for each baseline-hazard group is taken.

4.2.3. Proportion of insistors. In this section, we develop a consistent estimator of the proportion of insistors at risk through time given the event times, insistor effect ω and the number of insistors at baseline u, without a full probability survival model.

We take a group where C is independent of X (assumption A5b) and time at entry in the group (A6); the following may be applied separately to all such groups. Then consider arm 0 where latent strata are observed for $t_j \geqslant s_j$. The proportion of insistors at risk is to be updated at $t'_j = \min(t_j, s_j)$ for $j = 1, \ldots, n_0$ individuals in arm 0, which is the crossover time s_j if it is before the survival time t_j, or t_j otherwise. The combination of censoring indicator δ_j and whether the individual was at risk at s_j is classified using

$$e_j = \begin{cases} 0 & s_j > t_j \text{ and } \delta_j = 1 \text{ (event before offer to switch)} \\ 1 + c_j & s_j \leqslant t_j & \text{(at risk after offer)} \\ 3 & s_j > t_j \text{ and } \delta_j = 0 \text{ (censored before the offer).} \end{cases}$$

We take the order of individuals to be sorted ascending by t_j', so that for indices $j \in (1, \ldots, n_0 - 1)$, $t_j' \leqslant t_{j+1}'$. For a given $u_0 = u$ number of insistors in arm 0 at baseline and insistor effect $\omega' = \exp(\omega)$, the number u_j and proportion of insistors at risk for $j = 1, \ldots, n_0$ are estimated iteratively using

$$\xi_j = \begin{cases} \hat{\pi}(t_{j-1}')\omega'/\{(1 - \hat{\pi}(t_{j-1}')) + \hat{\pi}(t_{j-1}')\omega'\} & e_j = 0 \text{ (unknown stratum, event before the offer)} \\ 0 & e_j = 1 \text{ (ambivalent)} \\ 1 & e_j = 2 \text{ (insistor)} \\ \pi(t_{j-1}') & e_j = 3 \text{ (unknown stratum, censored before the offer)} \end{cases}$$

(16)

so that $\hat{\pi}(t_j') = \hat{u}_j/(n_0 - j)$ where $\hat{u}_j = \hat{u}_{j-1} - \xi_j$. The rationale is as follows. When $e_j = 1, 2$, then we know whether the individual crossed over, and the number still at risk $\hat{u}_j$ at time t_j' is updated accordingly. When $e_j = 3$, then $\hat{\pi}(t_j')$ does not change because censoring is assumed independent of insistor status. If $e_j = 0$, then c_j is unobserved and ξ_j estimates the probability $P(c_j = 1 \mid e_j)$; compare with Equation (3) for the binomial model. It arises from model (8) where

$$E\{dN(t_j' \mid c_j = 1)\} = \frac{\pi(t_j')\omega'}{1 - \pi(t_j')} E\{dN(t_j' \mid c_j = 0)\},$$

(17)

using $N(t \mid c_j)$ to denote a counting process that jumps by one when there is an event. For example, if the arms are balanced with $\pi(t_j') = 0.5$, but the insistor effect $\omega' = 2$, then for each event among the ambivalents, $(c_j = 0)$ $\omega' = 2$ events are expected in the insistors $(c_j = 1)$, that is, $E\{dN(t_j' \mid c_j = 1)\} = 2 \times E\{dN(t_j' \mid c_j = 0)\}$. Similarly, if $\omega' = 1$ and there are twice as many insistors as ambivalents $(\pi(t_j') = 2/3)$, then for each event among ambivalents, there will be $\pi(t_j')\{1 - \pi(t_j')\}^{-1} = 2$ events expected in the insistors. Although the survival model is in continuous time, a few ties are sometimes tolerated. If $t_j' = t_{j+1}' = \ldots = t_{j+k}'$ are tied, then the suitable adjustment is to set the associated $\hat{u}_l = \hat{u}_{j+k}$ for $l = j, \ldots, j + k - 1$.

Secondly, the number of insistors at baseline u may be estimated by maximum likelihood, given ω. The data are occasions when an individual crossed over or not, that is, when $e_j = 1, 2$. For $j = 1, \ldots, n_0$, if $e_j = 0$ or 3, then the log-likelihood $l_j(u \mid \omega) = 0$, else

$$l_j(u \mid \omega) = (e_j - 1)\log\{\pi^*(t_j')\} + (2 - e_j)\log\{1 - \pi^*(t_j')\},$$

(18)

where the $\pi^*(t_j') = \pi(t_j')$ if $t_j' > t_{j-1}'$, otherwise $\pi^*(t_j') = \pi(t_{j-k}')$ where k selects the first crossover event time equal to t_j'. The $\pi^*(t_j')$ definition is needed for ties when more than one event occurs at each point t_j'; the same estimate of the proportion of insistors should be used for all of them. The proportion of insistors at the start may be estimated by maximising the overall log-likelihood $l(u \mid \omega) = \sum_{j=1}^{n_0} l_j(u \mid \omega)$.

Finally, given u and ω, one may estimate the proportion of insistors at risk in the observed arm 0. If the proportion of insistors at the start of the trial in arm 0 is estimated to be the same as in arm 1, then one may use Equation (16) but for the unobserved arm 1.

4.2.4. Baseline hazard function. A baseline hazard function may be estimated for each baseline-hazard group, so for simplicity take that $\lambda_{i0}(t) = \lambda_0(t)$ for $i = 1, \ldots, n$, where $\lambda_0(t)$ is a step function with unknown jumps Δ_i at each observed event time t_i and that the times are sorted ascending. To estimate it, we use the partial derivative

$$\frac{\partial l}{\partial \Delta_i} = \sum_{j=1}^n \frac{\partial l_j}{\partial \Delta_i}.$$

In the following, we drop conditioning arguments from the setup except c in reduce notational burden, but they are still present. Firstly, consider $j = 1, \ldots, n$ where $r_j = 1$, that is, those in the treatment arm whose crossover status c_j is not observed. Let $\pi_j = P(C_j = 1 \mid x_j)$, then in this case

$$L_j = \pi_j \lambda_j^{\delta_j}(c_j = 1)\exp\{-\Lambda(c_j = 1)\} + (1 - \pi_j)\lambda_j^{\delta_j}(c_j = 0)\exp\{-\Lambda(c_j = 0)\}$$

so that

$$l_j = \log(\Delta_j^{\delta_j}) + \log(A_j)$$

where

$$A_j = \pi_j \psi(c_j = 1)^{\delta_j} \exp\{-\Lambda(c_j = 1)\} + (1 - \pi_j)\psi(c_j = 0)^{\delta_j} \exp\{-\Lambda(c_j = 0)\}.$$

For $j = i, \dots, n$, because

$$\partial \exp\{-\Lambda(t_j)\}/\partial \Delta_i = -\psi \exp\{-\Lambda(t_j)\},$$

we have that, for $j = i, \dots, n$ such that $r_j = 1$,

$$
\begin{aligned}
\frac{\partial l_j}{\partial \Delta_i} &= \delta_i/\Delta_i - \sum_{j=i}^{n} [\pi_j \psi(c_j = 1)^{(1+\delta_j)} \exp\{-\Lambda(t_j \mid c_j = 1)\} \\
&\quad + (1 - \pi_j)\psi(c_j = 0)^{(1+\delta_j)} \exp\{-\Lambda(t_j \mid c_j = 0)\}]/A_j \\
&:= \delta_i/\Delta_i - \sum_{j=i}^{n} a_{1j} \\
&:= \delta_i/\Delta_i + a_1,
\end{aligned}
\tag{19}
$$

where $:=$ denotes definition. Now consider $j = 1, \dots, n$ in arm 0. When insistor status is not observed because $t_j < s_j$, then the same form as Equation (19) applies; denote it a_2. If $t_j \geqslant s_j$, then

$$
\begin{aligned}
L_j &= [\pi_j \exp\{-\Lambda(s_j \mid c_j = 1)\} + (1 - \pi_j)\exp\{-\Lambda(s_j \mid c_j = 0)\}] \\
&\quad \times \lambda(t_j \mid c_j)^{\delta_j} \exp\{-\Lambda(s_j, t_j \mid c_j)\}
\end{aligned}
\tag{20}
$$

where $\Lambda(s_j, t_j \mid c_j) = \int_{s_j}^{t_j} \lambda_j(v \mid c_j)\mathrm{d}v$, so that

$$l_j = \log\{A_j^*\} + \log(\Delta_j^{\delta_j}) - \Lambda(s_j, t_j \mid c_j) + \log(\psi^{\delta_j})$$

where

$$A_j^* = \pi_j \exp\{-\Lambda(s_j \mid c_j = 1)\} + (1 - \pi_j)\exp\{-\Lambda(s_j \mid c_j = 0)\}.$$

Then for $j = i, \dots, n$ such that $s_j \leqslant t_i$

$$
\begin{aligned}
\frac{\partial l_j}{\partial \Delta_i} &= -\psi(c_j) + I(j = i)\delta_j/\Delta_i \\
&:= a_{3j}^{(1)}
\end{aligned}
$$

and for when $s_j > t_i$

$$
\begin{aligned}
\frac{\partial l_j}{\partial \Delta_i} &= -[\pi_j \psi(c_j = 1)\exp\{-\Lambda(s_j \mid c_j = 1)\}+ \\
&\quad (1 - \pi_j)\psi(c_j = 0)\exp\{-\Lambda(s_j \mid c_j = 0)\}]/A_j^* + I(j = i)\delta_j/\Delta_i \\
&:= a_{3j}^{(2)},
\end{aligned}
$$

so that overall for those observed to crossover ($t_j > s_j$)

$$
\begin{aligned}
\frac{\partial l}{\partial \Delta_i} &= \sum_{j=i}^{n} a_{3j}^{(1)} I(s_j < t_i) + a_{3j}^{(2)} I(s_j \geqslant t_i) \\
&:= a_3.
\end{aligned}
$$

Overall, we have

$$\frac{\partial l}{\partial \Delta_i} = I(r_i = 1)a_1 + I(r_i = 0)\{I(t_i < s_i)a_2 + I(t_i \geqslant s_i)a_3\}. \tag{21}$$

Setting Equation (21) to zero provides a route to estimation of Δ_i for $i = 1, \ldots, n$, conditional on θ and π. If $\delta_i = 1$ and $\delta_{i+1} = 1$ and it is not true that $t_i < s_j < t_{i+1}$ for $j = i + 1, \ldots, n$, then

$$
\Delta_i^{-1} - \Delta_{i+1}^{-1} = I(r_i = 1)a_{1i} + I(r_i = 0)\{I(t_i < s_i)a_{2i} + I(t_i \geq s_i)a_{3i}\} \\
:= D_i^*
\tag{22}
$$

say. This simple form arises because the $i + 1, \ldots, n$ terms in the a_1, a_2, a_3 summations over j cancel out. Extra terms are needed with censoring, and if $t_i < s_j < t_{i+1}$ for $j = i + 1, \ldots, n$ because the derivatives of those j's likelihood contribution with respect to Δ_i do not cancel out. Thus, more generally for two events at times i and $i + k$ with censored observations between from $i + 1, \ldots, i + k - 1$

$$
\Delta_i^{-1} - \Delta_{i+k}^{-1} = \sum_{j=i}^{i+k-1} D_j^* + \sum_{j=i+k}^{n} I(r_j = 0)I(t_i < s_j \leq t_{i+k})(a_{3j}^{(2)} - a_{3j}^{(1)}) \\
:= D_i
\tag{23}
$$

say, where the first summation is for the censored observations and the second summation is for those observed to crossover in arm 0 between t_i and t_{i+k}. Now $\Delta_{i+1}^{-1} = \Delta_i^{-1} - D_i$, and the problem of estimating the baseline given θ is reduced to determining Δ_i for the first i (i.e. smallest t_i) with $\delta_i = 1$. This may be achieved using a root-finding algorithm for the score function.

The above extends when there are ties. Suppose that w_i times are tied following t_i ($t_i = t_{i+1}, \ldots, t_{i+w-1}$), and w_{i+k} following t_{i+k}. Then the left-hand side of Equation (23) is $w_i \Delta_i^{-1} - w_{i+k} \Delta_{i+k}^{-1}$.

5. Example

We next apply the binomial estimator to the data from two monotherapy arms in the BIG1-98 trial, where postmenopausal women who had been diagnosed with hormone-receptor positive invasive breast cancer were randomised to receive 5 years of tamoxifen (control) or letrozole (treatment). Following the primary analysis [1], women who were still receiving tamoxifen were unblinded [2]. In this example, one might question the model assumption that the treatment effect of letrozole is that same after taking tamoxifen, as at baseline (i.e. $\gamma_1 = \gamma_1^*$ in the notation of Section 2). This issue was actually investigated by planned crossover arms in the BIG-1 98 trial, where little difference was found [10], so we proceed under the assumption.

Table I shows a summary of data previously reported [10, 11]. The results before unblinding are in period 0; period 1 is the time after that until 12-year follow-up. Of 2459 women randomised to tamoxifen, some 25% received letrozole after unblinding, with most crossing over between 3 and 5 years since randomisation.

Table II shows published results from fitting ITT and inverse probability of censoring weighting (IPCW) stratified proportional hazards models using individual time-to-event data [10]. The per-protocol

Table I. Disease-free survival results from BIG-1 98 trial [10, 11].

	Period	At risk	Number (%) events
Letrozole	0	2463	352 (14.3)
Tamoxifen	0	2459	418 (17.0)
Letrozole	1	2045	294 (13.0)
Tamoxifen	1	1975	309 (13.8)
– do not cross over	1	1356	251 (15.5)
– cross over	1	619	58 (9.4)

The data in period 0 were reported by [11]. The number of events in period 1 are the 12 year follow-up analysis [10] minus events in period 0. The number at risk in the second period was calculated as the number randomised minus the number (i) of disease-free survival events in the first period, (ii) lost to follow up and (iii) withdrawn (Table A2 in [11]), split by the reported 619 who chose to cross over [10].

Table II. Disease-free survival treatment effect estimates (95% CI) from BIG-1 98 trial 12-year update [10].

		Estimate	(95%CI)
Hazard ratio	ITT	0.86	(0.78, 0.96)
	IPCW	0.82	(0.74, 0.92)
Relative risk	ITT	0.89	(0.81, 0.97)
	Efficacy ($\hat{\gamma}$)	0.86	(0.77, 0.96)
	$-t = 0$ ($\hat{\gamma}_0$)	0.84	(0.74, 0.96)
	$-t = 1$ ($\hat{\gamma}_1$)	0.90	(0.74, 1.07)
	$-$ [Het. test]		[$\chi_1^2 = 0.32$]

Het. test, heterogeneity likelihood-ratio test.
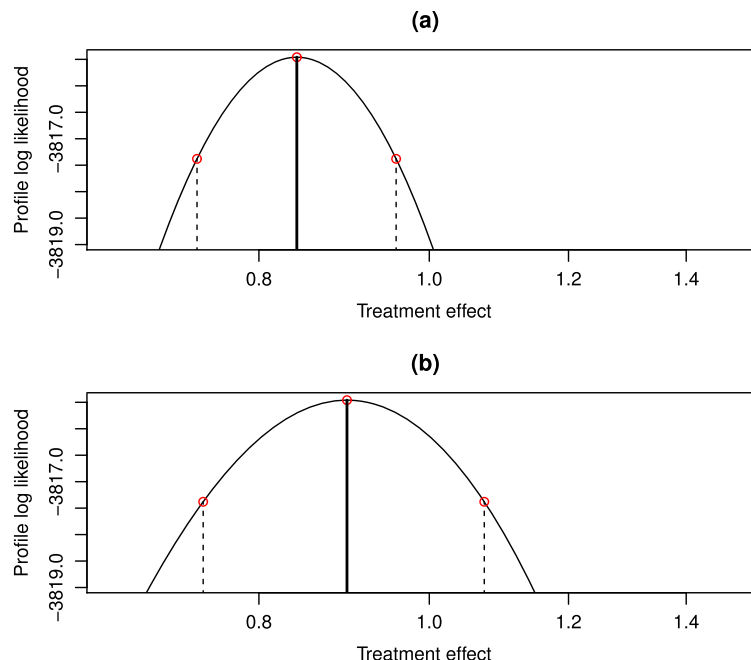
Figure 4. Profile likelihood surface to obtain confidence intervals for separate treatment effects in (a) the first period and (b) the second period, for disease-free survival in the example. The dotted lines and points are the 95% confidence interval. [Colour figure can be viewed at wileyonlinelibrary.com]

approach that censors women who ceased to comply with their randomised allocation is not suitable for these data because there was strong evidence that those who crossed over had a different baseline risk than those who did not; the group who crossed over from tamoxifen in Table I had a lower proportion of events (9.4%) than the letrozole arm (13.0%). Thus, the randomised balance of risk factors between comparison groups is almost certainly lost in a per-protocol analysis.

An ITT analysis was applied using just the number of events and women in both arms. In general, relative risks are closer to unity than hazard ratios, so it is not surprising that the relative risks for ITT were closer to unity than the hazard ratio in Table II. However, differences were small and ITT inference was unchanged. The binomial efficacy estimate for disease-free survival was of a similar order less than the ITT estimate, in comparison with the IPCW estimate. This provides some support to the IPCW estimate for disease-free survival.

We also used the model to allow for different treatment effects in each period. There was little heterogeneity between the point estimates of efficacy before and after unblinding (Table II; Figure 4). However, the wide confidence intervals after crossover indicate that further follow-up would be useful to determine whether disease-free survival was better in the first period. There was also almost no evidence for a difference in ω_0 and ω_1 (data not shown).

A limitation of the example is that it is best viewed as a demonstration of the method. Firstly, the actual number at risk in the second period was only estimated on the basis of published data (although small

changes to these numbers will not materially affect results). Secondly, application of the proportional-hazards model on individual data would also be a better comparison with IPCW. Finally, we were unable to apply either estimator for mortality endpoints because the subdivision of data illustrated by Figure 3 would be needed.

6. Simulations

The following simulations consider the performance of the full and partial likelihood estimators for different treatment and insistor effects, as the level of information on the treatment effect changes from censoring. Nelder–Mead simplex algorithms were used for both full and partial likelihood estimation, but very little difference was seen to a Broyden–Fletcher–Goldfarb–Shanno algorithm (not reported); Brent–Dekker algorithms estimated the insistor proportions and baseline hazard function [12].

Survival times were simulated from an exponential distribution. The simulation was set up so that there were three periods in follow-up time: before, during and after crossover. The period during crossover was the same length as the period before it. We considered two censoring scenarios (i) with censoring where an expected 5% events occurred before the first crossover in the lowest hazard group, and 5% events after the last crossover, and (ii) no censoring with an expected 10% of events before the first crossover in the lowest hazard group. For example, if the hazard was less for insistors and the treatment group, then the lowest hazard group was insistors randomised to treatment. One thousand individuals were in each arm, and the chance of being an insistor was 25% in both arms, but the number at baseline was sampled independently in both arms from a binomial distribution. The treatment effect (γ) was 7/10 or 10/7; the effect of being an insistor (ω) was either 0.2 or 2.0.

The pseudo algorithm used to generate the data is as follows. For $r = 0, 1$ and $i = 1, \ldots, n_r$ individuals in arm r with maximum follow-up F and an event-rate pre-crossover $\lambda_{ar}(c_i)$ and post-crossover $\lambda_{pr}(c_i)$ conditional on underlying crossover trait c_i: sample (i) crossover trait c_i at baseline from $P(c_i = 1) = 0.25$; (ii) crossover time s_i from a uniform distribution between 0 and 1 $\{U(0, 1)\}$ plus 1; (iii) survival time t_{i1} from an exponential distribution with pre-crossover rate $\lambda_{ar}(c_i)$; (iv) survival time t_{i2} from an exponential distribution with post-crossover rate $\lambda_{pr}(c_i)$. Then calculate (i) unobserved survival time (no censoring) $t_i^u = t_{i1}$ if $t_{i1} \leqslant s_i$; else $t_i^u = s_i + t_{i2}$; (ii) observed survival time (with censoring) $t_i = \min(t_i^u, F)$; and (iii) event status $d_i = 1$ if $t_i < F$; else $d_i = 0$.

Full and partial likelihood estimators are compared in Table III. Mean bias was very small; for example, 0.467 for $\hat{\gamma}$ from the partial likelihood in scenario 4 implies that the mean estimated treatment effect was $0.699 = \exp\{(1 + 0.00467) \times \log(0.7)\}$. The full likelihood bias was less than the partial likelihood for both the treatment effect γ and insistor effect ω, but their variances were very similar. There was also very little difference in the estimated proportion of insistors at baseline. Thus, overall, the full and partial

Table III. Simulation results of FL and PL estimators.								
Scenario	1	2	3	4	5	6	7	8
Treatment $\exp(\gamma)$	7/10	7/10	7/10	7/10	10/7	10/7	10/7	10/7
Switching $\exp(\omega)$	0.2	0.2	2.0	2.0	0.2	0.2	2.0	2.0
Censoring	Y	N	Y	N	Y	N	Y	N
(a) Mean bias (%) of log estimates								
γ FL	−0.002	0.004	−0.004	0.003	0.006	−0.003	0.007	0.001
γ PL	−0.231	0.340	0.533	0.467	0.666	−0.654	−0.037	−0.275
ω FL	0.009	−0.003	−0.019	0.002	0.015	0.002	−0.003	0.001
ω PL	0.835	−0.527	−1.078	0.130	1.522	−0.069	−0.259	−0.033
(b) Variance of log estimates								
γ FL	0.485	0.377	1.388	0.289	0.515	0.449	1.281	0.274
γ PL	0.485	0.407	1.405	0.289	0.519	0.484	1.284	0.282
ω FL	6.597	2.295	5.944	0.994	5.067	1.194	3.831	1.409
ω PL	6.612	2.066	6.005	0.993	5.078	1.210	3.849	1.399
(c) Percent (%) of insistors at baseline estimates $\hat{\pi}$								
Mean FL	25.056	25.116	24.971	25.042	24.998	25.073	25.016	25.034
Mean PL	25.059	25.118	24.989	25.041	24.997	25.091	25.020	25.022
SD FL	1.579	1.910	1.732	1.848	1.447	1.589	1.786	2.013
SD PL	1.580	1.741	1.736	1.848	1.447	1.603	1.787	2.025

FL, full likelihood; PL, partial likelihood.

Table IV. Bootstrap estimates of parameter uncertainty using partial likelihood estimator.

Scenario	1	2	3	4	5	6	7	8
Var($\hat{\gamma}$)	0.485	0.407	1.405	0.289	0.519	0.484	1.284	0.282
Mean bootstrap	0.503	0.412	1.450	0.295	0.548	0.476	1.181	0.275
SD bootstrap	0.053	0.049	0.235	0.033	0.063	0.067	0.161	0.029
Var($\hat{\omega}$)	6.612	2.066	6.005	0.993	5.078	1.210	3.849	1.399
Mean bootstrap	7.296	2.003	6.581	1.005	5.439	1.209	3.683	1.319
SD bootstrap	2.111	0.473	1.376	0.173	1.120	0.229	0.568	0.258

likelihood estimators had similar performance in a setting where the model assumptions were met, and assumption A5a was appropriate.

To evaluate the use of a simple bootstrap for inference, Table IV presents some results of resampling with replacement individuals in each arm in each simulation sample. Two hundred bootstrap resamples were made in each sample, and an estimate of the variance of the parameter estimate was obtained. The table shows that this matched the observed variance from the 1000 simulation replicates at each scenario quite closely. Because the variance of the full likelihood matched that of the partial likelihood quite closely in Table III, one might just bootstrap the partial likelihood as computational time becomes more of a consideration.

7. Conclusion

The methods in this paper are applicable for analysis of trials with selective crossover arising due to results from early efficacy endpoints and preserve randomisation. Our binary estimator is relatively simple to implement, but the proportional-hazards model is more involved. An R package [13] with code for both is available on request. We end by discussing the model assumptions in relation to other methods, and comment on other breast cancer trials where treatment efficacy might be estimated.

The most common methods to account for selective crossover in practice are IPCW [14] and g-estimation [15, 16]. IPCW censors individuals when they cease to comply with randomised allocation, but up-weights those who do comply and have similar characteristics to the non-compliers, as judged by a chosen statistical model. The problem is that the assumptions required to guarantee that randomisation is maintained cannot be verified because it is not possible to check the randomised balance of factors that have not been measured. G-estimation involves predicting the outcome for each person who switches treatment to be that expected if they had not switched. It is often accomplished through an accelerated failure-time model [17]. In contrast with IPCW, randomisation is maintained. A disadvantage is that it impossible to assess the assumed effect of treatment on the insistors after the offer to switch because there is no comparison group: all insistors receive treatment after the offer. It is also not possible to allow for time-dependent treatment effects. Other advantages and disadvantages of these methods have been discussed more fully elsewhere [18, 19].

Although inverse probability weighting and g-estimation make important untestable assumptions that are not required by our method, they may have better statistical performance than our estimators when their assumptions are most closely met. For instance, assumptions A5b and A6 might be restrictive for a particular data set. We also note that assumption A1 precludes a stochastic time of crossover, perhaps arising due to the health of a participant though time when randomised to control, but inverse probability weighting and g-estimation do not require this assumption. An important example of this in oncology is crossover after progression-free survival in advanced disease.

Other oncology trials appear to be suitable for our estimators. The open-label ABCSG-8 addressed a similar issue to BIG-1 98 and after publication of initial results approximately 9% crossed over from the tamoxifen arm [20]. The MA.17 trial showed superiority of letrozole over placebo for postmenopausal women with hormone receptor-positive breast cancer following about 5 years of tamoxifen [21] and almost two-thirds of the patients in the control arm crossed over after unblinding. In the open-label Herceptin Adjuvant (HERA) trial, more than half of patients in the 'control' group chose to cross over after a first interim analysis [22]. With such substantial crossover, methods such as those in this article are clearly needed to help assess longer-term differences between treatments.

In conclusion, we recommend that treatment efficacy and our estimators thereof be considered for trials with selective crossover in one arm arising due to results from early efficacy endpoints.

Acknowledgements

References

1. The Breast International Group (BIG) 1-98 Collaborative Group. A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer. *The New England Journal of Medicine* 2005; **353**(26):2747–2757.

2. The BIG 1-98 Collaborative Group. Letrozole therapy alone or in sequence with tamoxifen in women with breast cancer. *The New England Journal of Medicine* 2009; **361**(8):766–776.

3. Campigotto F, Weller E. Impact of informative censoring on the Kaplan–Meier estimate of progression-free survival in phase II clinical trials. *Journal of Clinical Oncology* 2014; **32**(20):3068–3074.

4. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; **58**(1):21–29.

5. Sommer A, Zeger SL. On estimating efficacy from clinical trials. *Statistics in Medicine* 1991; **10**(1):45–52.

6. Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. *Statistics in Medicine* 1997; **16**(9):1017–1029.

7. Cuzick J, Sasieni P, Myles J, Tyrer J. Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2007; **69**(4):565–588.

8. Rubin DB. Causal inference through potential outcomes and principal stratification: application to studies with "censoring" due to death. *Statistical Science* 2006; **21**(3):299–309.

9. Zeng D, Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2007; **69**(4):507–564.

10. Regan MM, Neven P, Giobbie-Hurder A, Goldhirsch A, Ejlertsen B, Mauriac L, Forbes JF, Smith I, Láng I, Wardley A, Rabaglio M, Price KN, Gelber RD, Coates AS, Thürlimann B, BIG 1-98 Collaborative Group, and International Breast Cancer Study Group (IBCSG). Assessment of letrozole and tamoxifen alone and in sequence for postmenopausal women with steroid hormone receptor-positive breast cancer: the BIG 1-98 randomised clinical trial at 81 years median follow-up. *Lancet Oncology* 2011; **12**(12):1101–1108.

11. Coates AS, Keshaviah A, Thurlimann B, Mouridsen H, Mauriac L, Forbes JF, Paridaens R, Castiglione-Gertsch M, Gelber RD, Colleoni M, Lang I, Del Mastro L, Smith I, Chirgwin J, Nogaret JM, Pienkowski T, Wardley A, Jakobsen EH, Price KN, Goldhirsch A. Five years of letrozole compared with tamoxifen as initial adjuvant therapy for postmenopausal women with endocrine-responsive early breast cancer: update of study BIG 1-98. *Journal of Clinical Oncology* 2007; **25**(5):486–492.

12. Galassi M, Davies J, Theiler J, Gough B, Jungman G, Alken P, Booth M, Rossi F. *GNU Scientific Library Reference Manual* 3rd revised edition edn. Network Theory Ltd: Bristol, 2009.

13. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria, 2010.

14. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; **56**(3):779–788.

15. Robins JM, Greenland S. Adjusting for differential rates of prophylaxis therapy for PCP in high-versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association* 1994; **89**(427):737–749.

16. Joffe MM, Brensinger C. Weighting in instrumental variables and g-estimation. *Statistics in Medicine* 2003; **22**(8):1285–1303.

17. Walker AS, White IR, Babiker AG. Parametric randomization-based methods for correcting for treatment changes in the assessment of the causal effect of treatment. *Statistics in Medicine* 2004; **23**(4):571–590.

18. Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP, Akehurst RL, Campbell MJ. Adjusting survival time estimates to account for treatment switching in randomized controlled trials - an economic evaluation context: Methods, limitations, and recommendations. *Medical Decision Making* 2014; **34**(3):387–402.

19. Latimer NR, Abrams KR, Lambert PC, Morden JP, Crowther MJ. Assessing methods for dealing with treatment switching in clinical trials: a follow-up simulation study. *Statistical Methods in Medical Research* 2016. 0962280216642 264+doi:10.1177/0962280216642264. http://dx.doi.org/10.1177/0962280216642264.

20. Dubsky PC, Jakesz R, Mlineritsch B, Pöstlberger S, Samonigg H, Kwasny W, Tausch C, Stöger H, Haider K, Fitzal F, Singer CF, Stierer M, Sevelda P, Luschin-Ebengreuth G, Taucher S, Rudas M, Bartsch R, Steger GG, Greil R, Filipcic L, Gnant M. Tamoxifen and anastrozole as a sequencing strategy: a randomized controlled trial in postmenopausal patients with Endocrine-Responsive early breast cancer from the Austrian breast and colorectal cancer study group. *Journal of Clinical Oncology* 2012; **30**(7):722–728.

21. Jin H, Tu D, Zhao N, Shepherd LE, Goss PE. Longer-term outcomes of letrozole versus placebo after 5 years of tamoxifen in the NCIC CTG MA.17 trial: analyses adjusting for treatment crossover. *Journal of Clinical Oncology* 2012; **30**(7):718–721.

22. Gianni L, Dafni U, Gelber RD, Azambuja E, Muehlbauer S, Goldhirsch A, Untch M, Smith I, Baselga J, Jackisch C, Cameron D, Mano M, Pedrini JL, Veronesi A, Mendiola C, Pluzanska A, Semiglazov V, Vrdoljak E, Eckart MJ, Shen Z, Skiadopoulos G, Procter M, Pritchard KI, Piccart-Gebhart MJ, Bell R. Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. *Lancet Oncology* 2011; **12**(3):236–244.