Published in final edited form as:

Biom J. 2009 December ; 51(6): 932–945. doi:10.1002/bimj.200800244.

An Accelerated Failure Time Mixture Cure Model with Masked Event

Jenny J. Zhang^{1,*} and Molin Wang^{1,2}

¹Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA

²Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, 3 Blackfan Circle, Boston, MA 02115, USA

Abstract

We extend the Dahlberg and Wang (*Biometrics* 2007, 63, 1237–1244) proportional hazards (PH) cure model for the analysis of time-to-event data that is subject to a cure rate with masked event to a setting where the PH assumption does not hold. Assuming an accelerated failure time (AFT) model with unspecified error distribution for the time to the event of interest, we propose rank-based estimating equations for the model parameters and use a generalization of the EM algorithm for parameter estimation. Applying our proposed AFT model to the same motivating breast cancer dataset as Dahlberg and Wang (*Biometrics* 2007, 63, 1237–1244), our results are more intuitive for the treatment arm in which the PH assumption may be violated. We also conduct a simulation study to evaluate the performance of the proposed method.

Keywords

Accelerated failure time model; Cure rate; EM algorithm; Masked event; Rank-based estimating equations

1 Introduction

There are certain clinical studies where any of a number of different potential events may lead to an observed failure, of which only one is the event of interest and the exact event may not always be identifiable. We refer to the unidentifiable failures as masked. In addition, a portion of the patient population is cured, *i.e.* they do not experience the event of interest. In this paper, we propose a semiparametric method to model the cure rate (incidence), the failure time distribution of the event of interest (latency), and the covariate effects on both when the proportional hazard (PH) assumption does not hold.

*Correspondence author: jjzhang6@gmail.com, Phone: 11-240-8998098.

Supporting Information for this article is available from the author or on the WWW under <http://dx.doi.org/10.1002/bimj.200800244>.

Conflict of Interests Statement

The authors have declared no conflict of interest.

Cure rate models were first presented by Berkson and Gage (1952) as an appropriate method to model data where a portion of the subjects may not experience the event of interest. Since then, many parametric and semiparametric mixture cure models have been proposed (*e.g.* Peng, Dear, and Denham, 1998; Sy and Taylor, 2000; Peng and Dear, 2000; Li and Taylor, 2002; Zhang and Peng, 2007). Estimation methods for survival data that account for masked event have also been studied in a number of situations (*e.g.* Goetghebeur and Ryan, 1995; Flehinger, Reiser, and Yashchin, 1998; Craiu and Duchesne, 2004). The accelerated failure time (AFT) model was first advocated as a useful alternative to the PH model for censored time-to-event data by Wei (1992). Although the PH model specifies that the effects of the covariates act multiplicatively on the hazard function, the AFT model regresses the logarithm of the failure times over the covariates, postulating a direct relationship between failure time and the covariates. Despite the theoretical advances made in the last decade (*e.g.* Tsiatis, 1990; Ritov, 1990; Fyngenson and Ritov, 1994; Jin *et al.*, 2003), semiparametric methods for the AFT model have rarely been used in applications due to the scarcity of efficient and reliable computational methods.

Very few estimation methods are available for time-to-event data subject to a cure rate with masked event. Dahlberg and Wang 2007, proposed a semiparametric PH mixture cure model for such data (for ease of reference, Dahlberg and Wang, 2007, will be referred to as DW hereafter). Assuming a PH model for the latency and a logistic model for the incidence, DW used the EM algorithm to conduct likelihood maximization and parameter estimation. Their motivating example came from International Breast Cancer Study Group (IBCSG) Trial VIII, where premenopausal, node-negative breast cancer patients were randomized to four treatment arms, and stratified according to estrogen-receptor (ER) status, whether radiotherapy was planned after surgery, and institution. We use the same example here to motivate our work; please refer to DW and Castiglione-Gertsch *et al.* (2003) for details on IBCSG Trial VIII.

It has been shown that some breast cancer adjuvant therapies for premenopausal women may interrupt menses, or induce amenorrhea (*i.e.* treatment-induced amenorrhea or TIA). As the number of young women diagnosed with breast cancer continues to increase, so does the demand for information about the impact of adjuvant therapies on menses and fertility. Such information strongly influences the treatment decisions of these patients (Partridge *et al.*, 2004). The underlying process of TIA, however, is not well understood and the observed data is complicated by the fact that menopause may also occur. Since both TIA (event I) and menopause (event II) result in an observed cessation of menses (the failure), the event leading to the observed failure is masked unless the patient recovers her menses (event III) after treatment. Moreover, a cured proportion is assumed to exist in the population since not all patients on treatment may experience TIA.

When DW applied their PH model to the goserelin arm of the IBCSG data, they found that older patients take a significantly longer time to experience TIA than younger patients, which is counter-intuitive since older patients are expected to have higher degrees of ovarian function suppression. Figure 1 plots the estimated log-integrated hazard *versus* log(time) for categorical age for the goserelin arm; a possible violation of the PH assumption is suggested by the crossing of the curves. Thus, DW's PH model may be inappropriate for this treatment

arm. As an alternative, the AFT model does not require the PH assumption and has parameter interpretations that bring different insights into the problem.

In this paper, we extend DW's PH mixture cure model for time-to-event data with masked event to the setting where the PH assumption does not hold through use of an AFT model. Sections 2 and 3 describe, respectively, the model and estimation method in detail. In Section 4, we discuss results from a simulation study performed to evaluate the proposed method, and in Section 5, we apply the method to the IBCSG Trial VIII data. We close with some discussion in Section 6.

2 AFT Mixture Cure Model

We use the same notation and definitions as DW, and refer the reader to Section 2.1 of their paper for details. Both Szwarc and Bonetti (2006) and DW assume that TIA may occur after menopause (and before treatment end) given the subject is uncured. We adopt the same assumption. We argue that such an assumption is valid since, although TIA would be an unobservable event in this case, the underlying, treatment-induced, biological changes would still occur. Moreover, menopause may also occur after TIA, with or without recovery first. Unlike chemotherapy, which may permanently damage ovarian function (deHaes *et al.*, 2003), there is little evidence to suggest that hormonal therapy (*e.g.* goserelin) has any effect on the natural process of menopause.

For the goserelin arm of the IBCSG trial that motivated this work, Szwarc and Bonetti (2006) showed that the occurrence of TIA due to goserelin does not affect time to menopause. Thus, we assume that time to TIA (event I) and time to menopause (event II) are independent conditional on the covariates. Note that our interest lies only in TIA, and menopause is essentially regarded as a censoring event. The only difference between menopause in our setting and a censoring event in the standard time-to-event data setting is that, given a failure, we may not know whether it is TIA or menopause. We also assume that the other censoring events are independent conditional on the covariates. In a doctoral thesis (Zhang and Wang, 2008), we investigated the relationship between TIA and disease recurrence or death for the goserelin arm; no notable association was found, controlling for patient characteristics.

We propose the following mixture cure model for the time to the event of interest (event I), T_1 ,

$$S_{T_1}(t_{1i}) = \alpha_i S_{T_1}(t_{1i} | \tau_i = 1) + (1 - \alpha_i),$$

where $\alpha_i = \mathbf{P}(\tau_i = 1)$. The incidence, α_i , is modeled using a logistic regression and the latency, $S_{T_1}(t_{1i} | \tau_i = 1)$, is modeled using an AFT model with unspecified error distribution. Specifically, the model for the incidence is

$$\alpha_i(\mathbf{Z}_i) = \frac{\exp(\mathbf{Z}_i \bar{\alpha}^T)}{1 + \exp(\mathbf{Z}_i \bar{\alpha}^T)}, \quad (1)$$

where $\bar{\mathbf{a}}$ is a vector of regression parameters for the vector of covariates $\mathbf{Z}_i$ and ^{T} denotes the transpose. The model for the latency is

$$\log(T_{1i}) = \mathbf{Z}_i \boldsymbol{\beta}^T + \varepsilon_i, \quad (2)$$

where T_{1i} is the time to event I for the i -th subject, $\boldsymbol{\beta}$ is a vector of regression parameters for the covariates $\mathbf{Z}_i$, and the distribution of the independent error terms, ε_i , is unspecified. Let $f_\varepsilon(\cdot)$ and $F_\varepsilon(\cdot)$ represent the probability and cumulative density functions of ε , respectively. For simplicity of exposition, we assume that time to event II, T_2 , follows a parametric distribution with parameters $\boldsymbol{\Omega}$, where $f_{T_2}(\cdot)$ and $F_{T_2}(\cdot)$ are the corresponding probability and cumulative density functions, respectively. The method can be easily extended such that T_2 follows a semiparametric distribution.

3 Estimation

3.1 Complete-data likelihood

Since direct maximization of the observed likelihood is difficult, we use an iterative method analogous to the EM algorithm to estimate the set of parameters $\boldsymbol{\Theta} = \{\bar{\mathbf{a}}, \boldsymbol{\beta}, \boldsymbol{\Omega}\}$. Let $\varepsilon_i(\boldsymbol{\beta}) = \log(y_i) - \mathbf{z}_i \boldsymbol{\beta}^T$, where y_i is the minimum of the failure time (x_i) and the censoring time (c_i). The complete-data likelihood, $L(\boldsymbol{\Theta}|\mathcal{P}_i)$, can be written and factored into three distinct components, $L_1(\bar{\mathbf{a}}|\mathcal{P}_i)$, $L_2(\boldsymbol{\beta}|\mathcal{P}_i)$, and $L_3(\boldsymbol{\Omega}|\mathcal{P}_i)$, as in Section 3.1 of DW with the following notational exceptions due to model differences: (i) $F_{T_1}(\cdot)$ in DW is $F_\varepsilon(\cdot)$ in our setting and (ii) $t_{(j)}$ in DW is $\varepsilon_{(j)}(\boldsymbol{\beta})$, where $\varepsilon_{(1)}(\boldsymbol{\beta}) < \varepsilon_{(2)}(\boldsymbol{\beta}) < \dots < \varepsilon_{(k)}(\boldsymbol{\beta})$ are the k distinct, ordered, uncensored failure residuals. Formulations of $L_1(\cdot)$, $L_2(\cdot)$, and $L_3(\cdot)$ are given in Appendix A for ease of reference.

3.2 Complete-data estimating functions

For $L_1(\bar{\mathbf{a}}|\mathcal{P}_i)$ and $L_3(\boldsymbol{\Omega}|\mathcal{P}_i)$, the corresponding complete-data estimating functions are simply the score functions, denoted by $\mathbf{U}_1(\bar{\mathbf{a}}, \mathcal{P}_i)$ and $\mathbf{U}_3(\boldsymbol{\Omega}, \mathcal{P}_i)$, respectively. Specifically,

$$\mathbf{U}_1(\bar{\mathbf{a}}, \mathcal{P}_i) = \sum_{i:\delta_i=6} \tau_i^\dagger \mathbf{z}_i - \sum_{i:\delta_i=1} \zeta_i (1 - \tau_i) \mathbf{z}_i + \sum_{i:\delta_i \in \{1,2,3\}} \mathbf{z}_i - \sum_{i:\delta_i \in \{4,5,6\}} \mathbf{z}_i \alpha_i \quad (3)$$

and

$$\begin{aligned} \mathbf{U}_3(\boldsymbol{\Omega}, \mathcal{P}_i) = & \frac{\partial}{\partial \boldsymbol{\Omega}} \left[\sum_{i:\delta_i=1} \zeta_i \log\{f_{T_2}(x_i)\} \right] - \frac{\partial}{\partial \boldsymbol{\Omega}} \left[\sum_{i:\delta_i=1} \zeta_i \log\{1 - F_{T_2}(x_i)\} \right] \\ & + \frac{\partial}{\partial \boldsymbol{\Omega}} \left[\sum_{i:\delta_i \in \{1,2,4,6\}} \log\{1 - F_{T_2}(x_i)\} \right] + \frac{\partial}{\partial \boldsymbol{\Omega}} \left[\sum_{i:\delta_i \in \{3,5\}} \log\{f_{T_2}(t_{2i})\} \right]. \end{aligned} \quad (4)$$

A score function for $\boldsymbol{\beta}$ based on $L_2(\boldsymbol{\beta}|\mathcal{P}_i)$ requires the correct specification of a parametric distribution for the error, ε , in the AFT model. As an alternative, we propose a complete-data estimating function, $\mathbf{U}_2(\boldsymbol{\beta}, \mathcal{P}_i)$, which does not require knowledge of the distribution of ε .

Fyngenson and Ritov (1994) proposed the following estimating function for censored time-to-event data in the standard framework (*i.e.* no cure rate or masked event), which is easily shown to be monotone in each component of β :

$$U(\beta) = n^{-1} \sum_{i=1}^n \sum_{h=1}^n \varphi_i(\mathbf{z}_i - \mathbf{z}_h) 1(\varepsilon_h(\beta) \geq \varepsilon_i(\beta)), \quad (5)$$

where φ_i is the censoring indicator (0 = censored, 1 = uncensored), and $1(\cdot)$ denotes the indicator function. In our setting, we have two possible events leading to a failure (event I and event II) and we are only interested in failures due to event I. Moreover, some subjects are cured with respect to event I. These complications will lead to a modified form of estimating function (5).

Let D_i denote the event that the i -th subject is observed to fail and the underlying event leading to failure (whether masked or unmasked) is event I, and $E_{i,h}$ denote the event that the h -th subject is uncured and in the risk set for event I at $\varepsilon_i(\beta)$. Also, let γ_i be the censoring indicator of whether (1) or not (0) a failure is observed before treatment end (U), the upper bound of T_1 . An estimating function analogous to (5) is then

$n^{-1} \sum_{i=1}^n \sum_{h=1}^n 1(D_i \cap E_{i,h})(\mathbf{z}_i - \mathbf{z}_h)$. It is straightforward to show that $1(D_i \cap E_{i,h})$ can be written as $v_i w_h 1(\varepsilon_h(\beta) \geq \varepsilon_i(\beta))$, where $v_i = \gamma_i \{ (1 - \gamma_i)(1 - \zeta_i) + \gamma_i \}$ and $w_h = \gamma_h [(1 - \gamma_h) \{ (\tau_h \zeta_h) + (1 - \zeta_h) \} + \gamma_h] + (1 - \gamma_h)(\tau_h)$.

Thus, we propose the following estimating function for β :

$$U_2(\beta, \mathcal{P}_i) = n^{-1} \sum_{i=1}^n \sum_{h=1}^n v_i w_h (\mathbf{z}_i - \mathbf{z}_h) 1(\varepsilon_h(\beta) \geq \varepsilon_i(\beta)). \quad (6)$$

The unbiasedness of $U_2(\beta, \mathcal{P}_i)$ follows directly from the fact that it can be rewritten as a U -statistic with a symmetric kernel:

$$U_2(\beta, \mathcal{P}_i) = n^{-1} \sum_{i=1}^n \sum_{h=1}^{i-1} (\mathbf{z}_i - \mathbf{z}_h) \{ v_i w_h 1(\varepsilon_h(\beta) \geq \varepsilon_i(\beta)) - v_h w_i 1(\varepsilon_i(\beta) \geq \varepsilon_h(\beta)) \}.$$

It follows that, the set of complete-data estimating equations for Θ is

$$U_c(\mathcal{P}_i, \Theta) = (U_1(\bar{\alpha}, \mathcal{P}_i), U_2(\beta, \mathcal{P}_i), U_3(\Omega, \mathcal{P}_i)) = \mathbf{0}.$$

3.3 The ES algorithm

We apply the ES algorithm, an iterative estimation method analogous to the EM algorithm (Dempster, Laird, and Rubin, 1977) proposed by Elashoff and Ryan (2004) for parameter estimation. The ES algorithm accommodates missing data in cases where a set of estimating equations can be found for the complete-data setting. In essence, it can be seen as a generalization of the EM algorithm for estimating equations; if the estimating equations

arise as score functions from a standard likelihood, then the ES algorithm reduces to the EM algorithm.

To employ the ES algorithm, we need to rewrite $\mathbf{U}_q(\mathcal{P}_i, \boldsymbol{\Theta})$, $q = 1, 2, 3$, as the sum of a function involving the complete data, $\mathbf{S}_q(\mathcal{P}_i)$, and a function involving only the observed data, $\mathbf{b}_q(\mathbf{O}_i)$. That is, $\mathbf{U}_q(\mathcal{P}_i, \boldsymbol{\Theta}) = \mathbf{S}_q(\mathcal{P}_i) + \mathbf{b}_q(\mathbf{O}_i)$. For $\mathbf{U}_1(\boldsymbol{\alpha}, \mathcal{P}_i)$ and $\mathbf{U}_3(\boldsymbol{\Omega}, \mathcal{P}_i)$, $\mathbf{S}_1(\mathcal{P}_i)$ and $\mathbf{S}_3(\mathcal{P}_i)$ are the first two terms of (3) and (4), respectively, while $\mathbf{b}_1(\mathbf{O}_i)$ and $\mathbf{b}_3(\mathbf{O}_i)$ are the last two terms of (3) and (4), respectively. For $\mathbf{U}_2(\boldsymbol{\beta}, \mathcal{P}_i)$,

$$\mathbf{S}_2(\mathcal{P}_i) = n^{-1} \sum_{i=1}^n \sum_{h=1}^n \vartheta_{ih} (\mathbf{z}_i - \mathbf{z}_h) 1(\varepsilon_h(\boldsymbol{\beta}) \geq \varepsilon_i(\boldsymbol{\beta}))$$

and

$$\mathbf{b}_2(\mathbf{O}_i) = n^{-1} \sum_{i=1}^n \sum_{h=1}^n \varpi_{ih} (\mathbf{z}_i - \mathbf{z}_h) 1(\varepsilon_h(\boldsymbol{\beta}) \geq \varepsilon_i(\boldsymbol{\beta})),$$

which are obtained by noting that $v_i w_h$ in estimating function (6) can be re-expressed as $\vartheta_{ih} + \varpi_{ih}$, where $\vartheta_{ih} = \frac{1}{n} \{ (1 - \gamma_i) \{ \frac{1}{n} (1 - \gamma_h) [\zeta_h (1 - \tau_h) + \zeta_i (1 - \zeta_h + \zeta_h \tau_h)] \} + \gamma_i \{ (1 - \tau_h) - \frac{1}{n} (1 - \tau_h) (1 - \gamma_h) \zeta_h \} \}$ involves the complete data and $\varpi_{ih} = \frac{1}{n} \{ 1 - \gamma_i (1 - \gamma_h) \}$ involves only the observed data.

We can then replace $\mathbf{S}(\mathcal{P}_i) = (\mathbf{S}_1(\mathcal{P}_i), \mathbf{S}_2(\mathcal{P}_i), \mathbf{S}_3(\mathcal{P}_i))$ in $\mathbf{U}_c(\mathcal{P}_i, \boldsymbol{\Theta})$ with $\mathbf{E}[\mathbf{S}(\mathcal{P}_i) | \mathbf{O}_i, \boldsymbol{\Theta}]$, the expectation of $\mathbf{S}(\mathcal{P}_i)$ conditional on the observed data, $\mathbf{O}_i$, and the unknown parameters, $\boldsymbol{\Theta}$, resulting in new estimating equations that involve only the observed data:

$$\mathbf{U}_{\text{obs}}(\mathbf{O}_i, \boldsymbol{\Theta}) = \mathbf{E}[\mathbf{S}(\mathcal{P}_i) | \mathbf{O}_i, \boldsymbol{\Theta}] + \mathbf{b}(\mathbf{O}_i) = \mathbf{0},$$

where $\mathbf{b}(\mathbf{O}_i) = (\mathbf{b}_1(\mathbf{O}_i), \mathbf{b}_2(\mathbf{O}_i), \mathbf{b}_3(\mathbf{O}_i))$. Following the arguments in Elashoff and Ryan (2004), $\mathbf{U}_{\text{obs}}(\cdot) = \mathbf{0}$ is an unbiased estimating equation for $\boldsymbol{\Theta}$. The estimate of $\boldsymbol{\Theta}$ can be obtained by solving $\mathbf{U}_{\text{obs}}(\cdot) = \mathbf{0}$ using the ES algorithm, which iterates between an E-step and a S-step. The E-step computes $\hat{\mathbf{S}} = \mathbf{E}[\mathbf{S}(\mathcal{P}_i) | \mathbf{O}_i, \boldsymbol{\Theta}^{(m)}]$, the conditional expectation of $\mathbf{S}(\mathcal{P}_i)$ with respect to the latent variables τ_i and ζ_i given the observed data and the latest updated parameter estimates, $\boldsymbol{\Theta}^{(m)}$. The S-step substitutes the conditional expectations calculated in the E-step into the complete-data estimating equations and solves $\mathbf{U}_{\text{obs}}(\cdot) = \hat{\mathbf{S}} + \mathbf{b}(\mathbf{O}_i) = \mathbf{0}$.

To solve the component of $\mathbf{U}_{\text{obs}}(\cdot)$ that involves the AFT model parameter $\boldsymbol{\beta}$, we note that $\mathbf{U}_2(\boldsymbol{\beta}, \mathcal{P}_i)$ can be taken as the gradient of

$$\mathcal{H}_2(\boldsymbol{\beta}, \mathcal{P}_i) = n^{-1} \sum_{i=1}^n \sum_{h=1}^n \vartheta_{ih} |\varepsilon_i(\boldsymbol{\beta}) - \varepsilon_h(\boldsymbol{\beta})| 1(\varepsilon_i(\boldsymbol{\beta}) < \varepsilon_h(\boldsymbol{\beta})) + n^{-1} \sum_{i=1}^n \sum_{h=1}^n \varpi_{ih} |\varepsilon_i(\boldsymbol{\beta}) - \varepsilon_h(\boldsymbol{\beta})| 1(\varepsilon_i(\boldsymbol{\beta}) < \varepsilon_h(\boldsymbol{\beta})),$$

which is a convex function, and finding the root of $\mathbf{U}_2(\boldsymbol{\beta}, \mathcal{P}_i) = \mathbf{0}$ is equivalent to minimizing $\mathcal{H}_2(\boldsymbol{\beta}, \mathcal{P}_i)$. Thus, the BFGS algorithm (Press *et al.*, 1992), a quasi-Newton method where the Hessian is updated by analyzing successive gradient vectors, is used to solve the estimating equations in the S-step. This algorithm is implemented with the `optim()` function in the statistical software package R given that you provide the gradient. Since the rank-based estimating function for $\boldsymbol{\beta}$ given in (6) is not continuous, a sandwich variance estimator is difficult to calculate. Therefore, we consider the case resampling bootstrap method (Efron and Tibshirani, 1993) for standard error estimation.

3.4 Expectations

For the E-step of the ES algorithm, it is necessary to calculate and update similar conditional expectations as in Section 3.1 of DW with respect to the latent variables τ_i and ζ_i given the observed data, $\mathbf{O}_i$, and the latest updated parameter estimates, $\boldsymbol{\Theta}^{(m)}$. For ease of reference, the formulation of these conditional expectations are given in Appendix B; please refer to DW for derivation details keeping in mind the notational differences outlined previously in Section 3.1.

Updating the conditional expectations involves the conditional survival function $S_\varepsilon(\varepsilon_i(\boldsymbol{\beta})|\tau_i = 1, \mathbf{Z}_i)$. We assume a nonparametric form for the conditional survival function,

$$S_\varepsilon(\varepsilon_i(\boldsymbol{\beta})|\tau_i=1, \mathbf{Z}_i) = \prod_{j: \varepsilon_{(j)}(\boldsymbol{\beta}) \leq \varepsilon_i(\boldsymbol{\beta})} \nu_j, \quad (7)$$

where $\nu_j \geq 0$, $\nu_0 = 1$, $S_\varepsilon(\varepsilon_{(j)}^-(\boldsymbol{\beta})|\tau_i=1, \mathbf{Z}_i) = S_\varepsilon(\varepsilon_{(j-1)}(\boldsymbol{\beta})|\tau_i=1, \mathbf{Z}_i)$, $\lambda_\varepsilon(\varepsilon_{(j)}(\boldsymbol{\beta})|\tau_i = 1, \mathbf{Z}_i) = 1 - \nu_j$ (Kalbfleisch and Prentice, 2002, Section 4.3), and $\lambda_\varepsilon(\cdot) = f_\varepsilon(\cdot)/S_\varepsilon(\cdot)$ is the hazard function for ε . Let $\mathbf{v} = (\nu_1, \dots, \nu_k)$.

Let R_j denote the risk set at $\varepsilon_{(j)}(\boldsymbol{\beta})$ and $\kappa_l = \{(\zeta_l \tau_l) - \zeta_l\}1(\delta_l=1) + (\tau_l^\dagger)1(\delta_l=6) + 1(\delta_l \neq 6)$. Following similar derivations as Kalbfleisch and Prentice (2002) and plugging in the nonparametric form of the conditional survival function (7), $L_2(\boldsymbol{\beta}|\mathcal{P}_i)$ can be rewritten as

$$\begin{aligned} & \prod_{j=1}^k \prod_{l \in A_j} S_\varepsilon(\varepsilon_{(j)}(\boldsymbol{\beta})|\tau_l=1, \mathbf{Z}_l)^{\tau_l \zeta_l} \{ \lambda_\varepsilon(\varepsilon_{(j)}(\boldsymbol{\beta})|\tau_l=1, \mathbf{Z}_l) S_\varepsilon(\varepsilon_{(j)}^-(\boldsymbol{\beta})|\tau_l=1, \mathbf{Z}_l) \}^{1-\zeta_l} \\ & \times \prod_{l \in B_j} \lambda_\varepsilon(\varepsilon_{(j)}(\boldsymbol{\beta})|\tau_l=1, \mathbf{Z}_l) S_\varepsilon(\varepsilon_{(j)}^-(\boldsymbol{\beta})|\tau_l=1, \mathbf{Z}_l) \prod_{l \in C_j} S_\varepsilon(\varepsilon_{(j)}(\boldsymbol{\beta})|\tau_l=1, \mathbf{Z}_l)^{\tau_l^\dagger} \quad (8) \\ & = \prod_{j=1}^k \prod_{l \in A_j} \nu_j^{\tau_l \zeta_l} (1 - \nu_j)^{1-\zeta_l} \prod_{l \in B_j} (1 - \nu_j) \prod_{l \in R_j - A_j - B_j} \nu_j^{\kappa_l}, \end{aligned}$$

where A_j , B_j , and C_j are as defined in Appendix A.

The closed-form maximum likelihood estimate (MLE) of $\mathbf{v}$, based on (8), leads to the following formula for updating $\mathbf{v}$ in the conditional expectations.

$$\hat{\mathbf{v}}^{(m+1)} = \left(1 - \frac{\sum_{l \in A_j} (1 - \zeta_l^\dagger) + d_{B_j}}{\sum_{l \in A_j} (1 - \zeta_l^\dagger + (\tau_l \zeta_l)^\dagger) + d_{B_j} + \sum_{l \in R_j - A_j - B_j} \kappa_l^\dagger}, j=1, \dots, k \right),$$

where d_{B_j} is the number of unmasked failures at $\varepsilon_{(j)}(\boldsymbol{\beta})$ and the superscript $\dagger$ denotes the conditional expectation of those indicators given $\mathbf{O}_i$, $\boldsymbol{\Theta}^{(m)}$, and $\mathbf{v}^{(m)}$. Note, if there is no cure rate or masked event, $\hat{\mathbf{v}}^{(m+1)}$ would simply be equal to $(1 - d_j/R_j, j = 1, \dots, k)$, where d_j and R_j are the number of failures and the risk set at $\varepsilon_{(j)}(\boldsymbol{\beta})$, respectively.

4 Simulation Study

To evaluate the performance of the proposed method, a simulation study is conducted where sample generation is based on the incidence and latency models specified in (1) and (2) with $\beta = -0.05$ or 0 , and $\bar{\mathbf{a}} = (-6, 0.2)$ or $(-10, 0.3)$ corresponding to respective average cure rates of around 27 and 39%. We generate the residuals, ε , in (2) from $N(0.2, 0.01)$ and the covariate Z from $N(35, 9)$. The simulated samples are modeled after the IBCSG Trial VIII data, where Z is a random variable for age at entry, T_1 is the time from study entry to TIA, and M is age at menopause, which is assumed to follow $N(51, 71)$. Since the patient population is premenopausal, for patients who enter the study, the conditional CDF for M is

$$F_M(m|Z=z, Z < M) = \frac{F_M(m) - F_M(z)}{1 - F_M(z)}, z \leq m.$$

The Kaplan–Meier survival estimate, assuming that all masked events are TIA, is used as the initial estimate for the conditional survival function. For each simulated sample, the logistic regression parameter estimates using only the unmasked data (*i.e.* the naive estimates) are used as starting values for $\bar{\mathbf{a}} = (p_0, p_1)$. Four different simulations are conducted, denoted A, B, C, and D in Table 1. Simulations A and B contain both cure rate and masked event while simulations C and D contain only masked event (*i.e.* no cure rate). The number of subjects in each sample (n) is 100, and all simulation results are for 100 replicates. To mimic the IBCSG data, there is a large proportion of masked event in all samples with and without a cure rate. We see that the results converged reliably to the true parameter values.

Simulations were also conducted to investigate the accuracy of the bootstrap standard errors mentioned at the end of Section 3.3; 100 bootstrap samples of size $n = 100$ were used for each simulation replicate. As seen in Table 1, the relative difference (*i.e.* ratio of [bootstrap SE – empirical SE]/empirical SE) between the bootstrap and empirical SEs is around 5%.

5 Breast Cancer Example

We now fit our proposed AFT model and apply our estimation method to the IBCSG Trial VIII data, with particular focus on the goserelin arm for which, as was shown in Fig. 1, the PH assumption may be violated. There are 304 (47%), 316 (80%), and 302 (83%) patients eligible for analysis (with corresponding percentages of masked events) in the goserelin,

CMF (chemotherapy), and CMF+goserelin arms, respectively. The objective of this analysis is to estimate the time to TIA, where we are interested in the effect of the single continuous covariate, age at entry (Z). As in DW, we assume that age at menopause follows $N(51.02, 71.23)$ in this study population, and eliminate the piece of the likelihood from the model for T_2 (time to menopause) to reduce the number of parameters in the estimation procedure and thus gain some efficiency. Given the pharmacodynamics of goserelin, it is assumed that the treatment induces amenorrhea in all patients eventually (Castiglione-Gertsch *et al.*, 2003), thus, $\alpha_i(z_i)$ is fixed to equal 1 for all patients on this treatment arm and the corresponding logistic regression parameters are not estimated.

Using our proposed estimation method, the parameter estimates and corresponding bootstrap standard errors (SE) by treatment arm are shown in Table 2. The bootstrap SEs are based on 100 bootstrap replicates for each arm. As in the simulations, we use the naive estimates of $\bar{\mathbf{a}}$ as starting values. The estimates of p_1 for the CMF and combination therapy arms are both positive with corresponding p -values of <0.001 and 0.056 , respectively. Thus, both the CMF and combination therapy arms show that older patients have a higher probability of experiencing TIA than younger patients (Fig. 2). However, this age effect is only significant for the CMF arm. This relationship between age at entry and probability of TIA was also shown in DW, and are consistent with those reported in the IBCSG Trial VIII clinical paper (Castiglione-Gertsch *et al.*, 2003).

The estimates of β for the latency are statistically significant for the CMF and combination therapy arms, where the corresponding p -values are both <0.0001 . In contrast to the PH model, the β estimates from the AFT model have interpretations as acceleration factors.

Specifically, if we let $T_1^{(z)}$ denote the time from study entry to TIA for subjects age z , then the acceleration factor for age z_1 relative to age z_2 can be calculated as

$AF_{(z_1, z_2)} = T_1^{(z_2)} / T_1^{(z_1)} = \exp\{-\beta(z_1 - z_2)\}$, where values greater than 1 denote that subjects age z_1 have a more accelerated (*i.e.* shorter) time to TIA than subjects age z_2 and vice-versa for values less than 1.

For example, in the CMF arm, $AF_{(35,44)} = 0.75$ (0.041) and $AF_{(55,44)} = 1.41$ (0.093), where values in parentheses are standard errors and 44 is the mean age. In other words, on the CMF arm, subjects age 35 will progress to TIA 0.25 times slower than subjects age 44, whereas subjects age 55 will progress to TIA 1.4 faster than subjects age 44. Thus, older patients on the CMF arm have a greater accelerated risk of TIA than younger patients. Similar conclusions can be drawn for the combination therapy arm, where $AF_{(35,44)} = 0.58$ (0.026) and $AF_{(55,44)} = 1.95$ (0.107), and the mean age is also 44. These conclusions are reflected in the corresponding estimated survival curves with respect to TIA for ages 35, 45, and 55 in Fig. 3, and are similar to those obtained in DW.

The age effect on the cure rate for the CMF and combination therapy arms can also be seen in the leveling off of the estimated marginal survival curves at the bottom of Fig. 3. The estimated marginal survival curves for the younger patients level off sooner than older patients in both arms, implicating that younger patients have a lower probability of experiencing TIA. For patients on the goserelin arm, the risk of TIA does not depend on age

at entry, which “corrects” the counter-intuitive effect found by DW and agrees with what is clinically expected. Given that goserelin injections were administered for 24 months in IBCSG Trial VIII, we see from Fig. 3 that goserelin induces amenorrhea very quickly (median time of 3 months), which is indicative of the pharmacodynamics of the treatment. As discussed in Section 2, we believe that the assumption of independent events (TIA and menopause) is justified for our treatment arm of interest (goserelin), in which the PH assumption may be violated. We acknowledge, however, that this assumption may not hold for the chemotherapy-containing arms of IBCSG trial VIII.

Because of the discrepancy between some aspects of our proposed AFT model and DW’s PH model when applied to the goserelin arm, it is useful to investigate the goodness-of-fit. We thank a referee for suggesting the method used in Li and Taylor (2002), where a quantity with an approximate uniform (0,1) distribution if the model fits is estimated from each observation. We then graphically compare the empirical distribution of this quantity across the observations to a uniform distribution. More specifically, when we fit a time-to-event model (AFT or PH), we obtain the estimated survival probabilities, denoted by $\{\hat{G}_i(T_i), i = 1, \dots, n\}$, based on our parameter estimates. This quantity should have an approximate uniform (0,1) distribution. In Fig. 4, we plot the Kaplan–Meier curve where $\{1 - \hat{G}_i(T_i), i = 1, \dots, n\}$ are regarded as the survival times with corresponding indicators of whether or not T_i is censored for the AFT and PH models for the goserelin arm, and compare it with the diagonal line representing the uniform distribution. We see a much better approximation to a uniform for the proposed AFT model than DW’s PH model. The gaps in the PH model plot are due to the large drops in estimated survival probabilities over time for the goserelin arm in DW (e.g. from 0.688 at $t_1 = 1$ to 0.219 at $t_1 = 2$).

6 Discussion

The proposed method for the analysis of time-to-event data subject to a cure rate with masked event is an extension of the work by DW to the setting where the PH assumption does not hold. Our estimation method is based on the AFT model with unspecified error distribution and does not require any distributional assumptions on T_1 or T_2 . In contrast to the results from DW’s PH model, our results for the goserelin arm in the IBCSG data example follow clinical intuition.

A well-known issue of the mixture cure model is that correct estimation of the cure rate (α_i) is contingent on the identifiability of the model, *i.e.* the ability to distinguish between cured patients and long-term uncured survivors. Yu *et al.* (2004) investigated the identifiability of mixture cure models in the setting of parametric models and no masked event through extensive simulations. They showed that cure rate estimates could differ greatly if the latency distribution is incorrectly specified and/or if the length of follow-up is less than the median survival time of uncured patients. The semiparametric AFT mixture cure models proposed in this paper and Zhang and Peng’s (2007) do not require specification of the latency distribution. Zhang and Peng (2007) focused on a setting without masked event and found, through a simulation study, that their semiparametric AFT mixture cure model performed favorably compared to the parametric AFT mixture cure model with correct latency distribution specification, implying that their model has good identifiability.

Considering event II (menopause) as a form of censoring for our event of interest (event I: TIA), our proposed estimating function (6) reduces to Zhang and Peng's (2007).

As pointed out by a referee, since TIA can only occur during treatment in our breast cancer example, if $S_{T_1}(t_1 = U | \tau_i = 1, \mathbf{Z}_i) > 0$, where U denotes treatment end, there may be identifiability problems. In the case where there is no finite upper limit for T_1 , if the survival probability as time goes to infinity is non-zero, Taylor (1995) and Sy and Taylor (2000) suggested imposing a “zero-tail constraint,” where the survival probabilities for all times past the last failure time are forced to equal 0. We employ a similar constraint in our case; that is, let $\hat{S}_\varepsilon(\varepsilon_{(j)}(\beta) | \tau_i = 1, \mathbf{Z}_i) = 0$ for all j such that $\varepsilon_{(j)}(\beta) > \varepsilon_U(\beta)$, where $\varepsilon_U(\beta) = \log(U) - (\mathbf{Z}\beta^T)^*$ and $(\mathbf{Z}\beta^T)^* = \max(\mathbf{Z}_i\beta^T, i = 1, \dots, n)$. It is easy to show that, under the above constraint, the survival probability of T_1 conditional on $\tau_i = 1$ is zero at $t_1 = U$. The use of such a constraint was avoided in our IBCSG data example.

It is worthwhile to note that the ES algorithm proposed by Elashoff and Ryan (2004) described in Section 3.3 does not involve an infinite dimensional parameter that needs to be estimated iteratively. Our application of the algorithm involves such a parameter in the estimation of the conditional survival function, $S_\varepsilon(\varepsilon_i(\beta) | \tau_i = 1, \mathbf{Z}_i)$, discussed in Section 3.4. However, from our simulation study in Section 4, we see that the results converged reliably to the true parameter values.

In both our simulation study and breast cancer data application, we assumed that the incidence and latency are affected by the same covariate (*i.e.* age at entry); however, our methodology would also apply when the incidence and the latency are affected by different covariates; a covariate that is important for the incidence may not be important for the latency and vice versa. In addition, although the proposed method is dependent on the presence of some unmasked events in order to give accurate estimates of the cure rate and survival distribution, our simulations and IBCSG data analysis show that our method is able to accommodate fairly substantial proportions of masked event.

Acknowledgements

We thank the patients, physicians, nurses, and data managers who participated in the International Breast Cancer Study Group (IBCSG) Trial VIII. We further acknowledge support from the United States National Cancer Institute (CA-75362) and the United States National Institute of Health Cancer Training Grant (Jing J. Zhang). We express our gratitude to Richard Gelber, Robert Gray, Ann Partridge, and Zhi-Min Yuan for their assistance throughout. We also thank the editor, associate editor, and referee for their insightful comments and suggestions.

Appendix

A Components of the complete-data likelihood referred to in Section 3.1

$$L_1(\bar{\alpha} | \mathcal{D}_i) = \prod_{i: \delta_i=1} \pi_i^{\tau_i \zeta_i + (1-\zeta_i)} (1 - \pi_i)^{(1-\tau_i) \zeta_i} \prod_{i: \delta_i \in \{2,3\}} \pi_i \prod_{i: \delta_i \in \{4,5\}} (1 - \pi_i) \times \prod_{i: \delta_i=6} \pi_i^{\tau_i^\dagger} (1 - \pi_i)^{(1-\tau_i^\dagger)}$$

$$L_2(\boldsymbol{\beta}|\mathcal{P}_i)=\prod_{j=1}^k \prod_{l \in A_j} \{1 - F_{\varepsilon}(\varepsilon_{(j)}(\boldsymbol{\beta}))\}^{\tau_l \zeta_l} f_{\varepsilon}(\varepsilon_{(j)}(\boldsymbol{\beta}))^{1-\zeta_l} \prod_{l \in B_j} f_{\varepsilon}(\varepsilon_{(j)}(\boldsymbol{\beta})) \times \prod_{l \in C_j} \{1 - F_{\varepsilon}(\varepsilon_{(j)}(\boldsymbol{\beta}))\}^{\tau_l^{\dagger}}$$

$$L_3(\boldsymbol{\Omega}|\mathcal{P}_i)=\prod_{i:\delta_i=1} f_{T_2}(x_i)^{\zeta_i} \{1 - F_{T_2}(x_i)\}^{1-\zeta_i} \prod_{i:\delta_i \in \{3,5\}} f_{T_2}(t_{2i}) \prod_{i:\delta_i \in \{2,4,6\}} \{1 - F_{T_2}(c_i)\}$$

Note that A_j denotes the set of all subjects who experienced a masked event at $\varepsilon_{(j)}(\boldsymbol{\beta})$, B_j denotes the set of all subjects who experienced an unmasked failure at $\varepsilon_{(j)}(\boldsymbol{\beta})$, and C_j denotes the set of subjects censored in the interval $[\varepsilon_{(j)}(\boldsymbol{\beta}), \varepsilon_{(j+1)}(\boldsymbol{\beta}))$, $j = 1, \dots, k$.

B Conditional expectations referred to in Section 3.4

For notational simplicity, we let $\boldsymbol{\beta}$ denote $\boldsymbol{\beta}^{(m)}$.

$$E(\zeta_i|\mathbf{O}_i, \boldsymbol{\theta}^{(m)}, \tau_i=1)=\frac{\{1 - F_{\varepsilon}(\varepsilon_i(\boldsymbol{\beta}))\}f_{T_2}(x_i)}{\{1 - F_{\varepsilon}(\varepsilon_i(\boldsymbol{\beta}))\}f_{T_2}(x_i)+\{1 - F_{T_2}(x_i)\}f_{\varepsilon}(\varepsilon_i(\boldsymbol{\beta}))}$$

$$E(\tau_i|\mathbf{O}_i, \boldsymbol{\theta}^{(m)})=\frac{\{1 - F_{T_1}(\varepsilon_i(\boldsymbol{\beta}))\}\alpha_i f_{T_2}(x_i)+\{1 - F_{T_2}(x_i)\}\alpha_i f_{\varepsilon}(\varepsilon_i(\boldsymbol{\beta}))}{f_{T_2}(x_i)\{1 - \alpha_i F_{\varepsilon}(\varepsilon_i(\boldsymbol{\beta}))\}+\{1 - F_{T_2}(x_i)\}\alpha_i f_{\varepsilon}(\varepsilon_i(\boldsymbol{\beta}))}$$

$$E(\tau_i^{\dagger}|\mathbf{O}_i, \boldsymbol{\theta}^{(m)})=\frac{\alpha_i\{1 - F_{\varepsilon}(\varepsilon_i(\boldsymbol{\beta}))\}}{(1 - \alpha_i)+\alpha_i\{1 - F_{\varepsilon}(\varepsilon_i(\boldsymbol{\beta}))\}}$$

$$E(\zeta_i \tau_i|\mathbf{O}_i, \boldsymbol{\theta}^{(m)})=E(\zeta_i|\mathbf{O}_i, \boldsymbol{\theta}^{(m)}, \tau_i=1)E(\tau_i|\mathbf{O}_i, \boldsymbol{\theta}^{(m)})$$

$$E(\zeta_i|\mathbf{O}_i, \boldsymbol{\theta}^{(m)})=E(\zeta_i \tau_i|\mathbf{O}_i, \boldsymbol{\theta}^{(m)})+1 - E(\tau_i|\mathbf{O}_i, \boldsymbol{\theta}^{(m)})$$

References

- Berkson J, Gage RP. Survival curve for cancer patients following treatment. Journal of the American Statistical Association. 1952; 47:501–515.
- Castiglione-Gertsch M, O'Neill A, Price KN, Goldhirsch A, Coates AS, Colleoni M, Nasi ML, Bonetti M, Gelber RD. on behalf of the International Breast Cancer Study Group. Adjuvant chemotherapy followed by goserelin versus either modality alone for premenopausal lymph node-negative breast cancer: a randomized trial. Journal of the National Cancer Institute. 2003; 95:1833–1846. [PubMed: 14679153]
- Craiu RV, Duchesne T. Inference based on the EM algorithm for the competing risks model with masked causes of failure. Biometrika. 2004; 91:543–558.

- Dahlberg S, Wang M. A proportional hazards cure model for the analysis of time to event with frequently unidentifiable causes. *Biometrics*. 2007; 63:1237–1244. [PubMed: 18078485]
- deHaes H, Olschewski M, Kaufmann M, Schumacher M, Sauerbrei W. Quality of life in goserelin-treated versus cyclophosphamide + methotrexate 6 fluorouracil-treated premenopausal and perimenopausal patients with node-positive, early breast cancer: the zoladex early breast cancer research association trialists group. *Journal of Clinical Oncology*. 2003; 21:4510–4516. [PubMed: 14610048]
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*. 1977; 39:1–22.
- Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall; 1993.
- Elashoff M, Ryan L. An EM algorithm for estimating equations. *Journal of Computational and Graphical Statistics*. 2004; 13:48–65.
- Flehinger JB, Reiser B, Yashchin E. Survival with competing risks and masked causes of failures. *Biometrika*. 1998; 85:151–164.
- Fygenson M, Ritov Y. Monotone estimating equations for censored data. *The Annals of Statistics*. 1994; 22:732–746.
- Goetghebuer E, Ryan L. Analysis of competing risks survival data when some failure types are missing. *Biometrika*. 1995; 82:821–833.
- Jin Z, Lin DY, Wei LJ, Ying Z. Rank-based inference for the accelerated failure time model. *Biometrika*. 2003; 90:341–353.
- Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. 2nd edn. New York: Wiley; 2002.
- Li CS, Taylor JMG. A semiparametric accelerated failure time cure model. *Statistics in Medicine*. 2002; 21:3235–3247. [PubMed: 12375301]
- Partridge AH, Gelber S, Peppercorn J, Sampson E, Knudsen K, Laufer M, Rosenberg R, Przypyszny M, Rein A, Winer EP. Web-based survey of fertility issues in young women with breast cancer. *Journal of Clinical Oncology*. 2004; 22:4174–4183. [PubMed: 15483028]
- Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics*. 2000; 56:237–243. [PubMed: 10783801]
- Peng Y, Dear KBG, Denham JW. A generalized F mixture model for cure rate estimation. *Statistics in Medicine*. 1998; 17:813–830. [PubMed: 9595613]
- Press, WH.; Teukolsky, SA.; Vetterling, WT.; Flannery, BP. *Numerical Recipes in C: the Art of Scientific Computing*. 2nd edn. New York: Cambridge University Press; 1992.
- Ritov Y. Estimation in a linear regression model with censored data. *The Annals of Statistics*. 1990; 18:303–328.
- Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics*. 2000; 56:227–236. [PubMed: 10783800]
- Szwarc SE, Bonetti M. Modeling menstrual status during and after adjuvant treatment for breast cancer. *Statistics in Medicine*. 2006; 25:3534–3547. [PubMed: 16345025]
- Taylor JMG. Semiparametric estimation in failure time mixture models. *Biometrics*. 1995; 51:899–907. [PubMed: 7548707]
- Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*. 1990; 18:354–372.
- Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*. 1992; 11:1871–1879. [PubMed: 1480879]
- Yu B, Tiwari RC, Cronin KA, Feuer EJ. Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine*. 2004; 23:1733–1747. [PubMed: 15160405]
- Zhang J, Peng Y. A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statistics in Medicine*. 2007; 26:3157–3171. [PubMed: 17094075]
- Zhang, JJ.; Wang, M. *Novel Methodologies for the Analysis of Complex Failure Time Data and Alternative Progression-Free Survival Estimators*. Ph.D. Thesis. Harvard University; 2008. Latent class joint model of ovarian function suppression and DFS for premenopausal breast cancer patients.

Figure 1.
Proportional hazards assumption check for goserelin arm of IBCSG Trial VIII.

Figure 2.
Probability of being uncured with respect to TIA for CMF and combination arms.

Figure 3.

Estimated survival curves with respect to TIA by treatment arm for ages 35, 45, and 55, where the conditional survival probability of TIA is $S_{T_1}(t_{1i}|\tau_i = 1)$ and the marginal survival probability of TIA is $S_{T_1}(t_{1i}) = \alpha_i S_{T_1}(t_{1i}|\tau_i = 1) + (1 - \alpha_i)$.

Figure 4.

Kaplan–Meier estimates of the estimated distribution evaluated at the observed data for the goserelin arm of IBCSG Trial VIII.

Table 1

Simulation results for 100 replications with sample size $n = 100$, where SE is the empirical standard error and the bootstrap SE is based on 100 replicates.

Simulation	Parameter	True value	Mean estimate (SE)	Bootstrap SE
A	p_0	-6	-7.954 (5.770)	6.057
	p_1	0.2	0.224 (0.166)	0.174
	β	-0.05	-0.051 (0.017)	0.018
B	p_0	-10	-11.984 (5.879)	6.142
	p_1	0.3	0.326 (0.168)	0.176
	β	0	-0.0004 (0.019)	0.020
C	β	-0.05	-0.050 (0.017)	0.018
D	β	0	0.0010 (0.018)	0.019

Table 2

Parameter estimates by treatment arm for IBCSG Trial VIII, where values in parentheses are bootstrap standard errors.

Parameter	Goserelin (SE)	CMF (SE)	CMF + goserelin (SE)
p_0	–	–7.0910 (2.153)	–2.3722 (2.626)
p_1	–	0.1863 (0.052)	0.1416 (0.074)
β	0.0000*	–0.0314 (0.006)	–0.0609 (0.005)

* SE < 0.00001.