



**HAL**  
open science

# Effect of a correlated competing risk on marginal survival estimation in an accelerated failure time model

Valérie Gares, Malcolm Hudson, Maurizio Manuguerra, Val GebSKI

## ► To cite this version:

Valérie Gares, Malcolm Hudson, Maurizio Manuguerra, Val GebSKI. Effect of a correlated competing risk on marginal survival estimation in an accelerated failure time model. *Communications in Statistics - Simulation and Computation*, 2024, pp.1-23. 10.1080/03610918.2024.2380005 . hal-04850098

**HAL Id: hal-04850098**

**<https://hal.science/hal-04850098v1>**

Submitted on 19 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Effect of a correlated competing risk on marginal survival estimation in an accelerated failure time model.

Valerie Gares<sup>a,c</sup> and Malcolm Hudson<sup>b,c</sup> and Maurizio Manuguerra<sup>b</sup> and Val Gebski<sup>c</sup>

<sup>a</sup> Univ. Rennes, INSA, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France;

<sup>b</sup> School of Mathematical and Physical Sciences, FSE, Macquarie University, Sydney, Australia;

<sup>c</sup> NHMRC Clinical Trials Centre, University of Sydney, Australia

## ARTICLE HISTORY

Compiled December 19, 2024

## ABSTRACT

In problems with a time to event outcome, subjects may experience competing events, which censor the outcome of interest. Cox's partial likelihood estimator treating competing events as independent censoring is commonly used to examine group differences in clinical trials but fails to adjust for omitted covariates and can bias the assessment of marginal benefit. A bivariate normal linear model generating latent data with dependent censoring is used to assess this bias. Our R-package *bnc* provides maximum penalized likelihood (MPL) parameter estimation using a novel EM algorithm. Using *bnc*, we study the properties of such MPL estimation. Simulation results for two-sample survival comparisons of time to an event of interest, with independent censoring accompanied by censoring from a correlated competing risk, are presented. Key parameters – means, hazard ratios, and correlation – are estimated. These results demonstrated that, despite ill-conditioning in models generating correlated competing risks, estimates of marginal effects are reliable. Bivariate normal models were fitted in a trial of head and neck cancer. Model fits help with clinical interpretation while also supplementing other standard methods for follow-up that are terminated by intervening risks.

## KEYWORDS

Time-to-event analysis Accelerated failure time; bivariate normal; Cox model; competing risks; dependent censoring; EM algorithm; ill-conditioning

## 1. Introduction<sup>1</sup>

Competing risks data often arise in medical follow-up studies or industrial life tests when several different types of events can end the follow-up period of a subject.

The role of covariates in competing risks is often studied modeling the cause-specific hazard function (Cox, 1972) or the subdistribution hazard function (Fine and Gray, 1999). Both these approaches assume proportionality of the hazards.

A competing risk occurring independently of the event of interest provides independent censoring. In this case a coefficient estimated in Cox regression provides a hazard ratio (HR) for the marginal survival of interest. However, a correlated competing risk introduces dependent censoring on the event of interest.

---

CONTACT Valerie Gares. Email: valerie.gares@insa-rennes.fr

Email: malcolm.hudson@mq.edu.au

<sup>1</sup>Supplementary Data to be annexed to the electronic mns is here included as Appendix C. It comprizes Supplementary Tables and Figures.

This work is motivated by a study in which 143 patients have received surgery for metastatic melanoma. The trial compares two highly correlated events, the local relapse (event of interest) and the regional relapse (competing event) in a group of patients receiving adjuvant radiotherapy and in a control group (O'Brien et al., 1997). We compare findings of the standard survival methods for competing risks with those of the methodology introduced in this paper. As shown by Keiding et al. (1997), when the correlation of competing risks is not effectively controlled by model variables, neglected covariates can induce unobservable correlation, toward which the Cox regression model is not robust (Hougaard, 1999). The effect is that the estimated cause-specific HR may differ from the marginal HR (Kalbfleisch and Prentice, 2002; Emura et al., 2020).

Emura and Chen (2016) analyze this difference under copula-based dependent censoring. They show that if the censoring probability is high, the difference is significant. Furthermore, the difference inflates as the dependence (copula) parameter deviates from zero. The paper of Emura et al. (2020) and book of Emura and Chen (2018) may also be consulted for a discussion on this and other relevant approaches to dependent censoring using techniques involving copulas. Lu et al. (2016) confirm in simulations that, in the presence of correlation, the estimated hazard ratios for an event of interest, estimated by using the Cox model and treating the competing risk events as independent censoring, can differ substantially from the *marginal* hazard ratio of the event of interest. Thus it is important with correlated competing risks to supplement cause-specific HR estimates by corresponding marginal estimates, which require a joint distribution of the bivariate outcomes.

In survival analysis, a useful alternative to proportional hazards (PH) models is the AFT model (Wei, 1992). Hougaard (1999), in a general review of survival methods, notes that parametric AFT models are robust in conditions where PH models are not. Klein et al. (1999) and Lambert et al. (2004) argue for parametric accelerated failure time models involving frailty-like terms as an effective alternative to parametric PH models. Weibull AFT models share the PH properties assumed by the semi-parametric Cox regression model, facilitating trial power and sample-size calculations under scenarios varying treatment effects. In the alternative lognormal AFT model the regression coefficient for the treatment indicator represents the treatment effect and its exponential provides the *median ratio* (MR, the ratio of median survival with and without treatment). Hence AFT models with lognormal marginals offer readily interpretable regression models for cause-specific median times to failure. In AFT models, bias in estimating marginal effects from regression coefficients is again anticipated when a correlated competing event censors follow-up.

Recently Deresa and Van Keilegom (2020b) review parametric and semi-parametric approaches to regression modelling for competing risks and introduce a multivariate normal regression model for dependent censoring. Their model allows for different forms of censoring (including loss to follow up or termination of the study) and for an initial parameterized power transformation of survival time. Assumptions for model identifiability, including those required by the parameterized class of power transformations, are provided. The multivariate normal linear model of Deresa and Van Keilegom (2020a) (the DVK model) is a more general formulation of the bivariate normal (BVN) censored linear model of Hudson et al. (2018). This earlier model was applied in the simpler context of bivariate outcomes, identity power transformation and covariates common to both causes. The implementation of the simpler model in our R package `bnc` allows for the efficient estimation of the parameters and the estimation of marginal hazards in the case of dependent and independent censoring.

Other recent examples of studies on correlated competing risks inducing dependent censoring include Tai et al. (2008), Dignam et al. (2012), Korn et al. (2015). The BVN model allows us to broaden the scope of these studies.

In many applications involving competing risks data, individual events within a cluster

may be correlated due to unobserved shared factors across individuals. Many authors are then concerned with non-parametric statistical tests for correlated competing risks data (see for example Chen et al. (2008); Ha et al. (2016)). Following Cox (interviewed in Reid, 1994), we argue that non-parametric and semi-parametric modeling in competing risks can lead to overlooking the biases implicit in proportional hazards and independent censoring assumptions, while parametric models can add biological insight. The gap between common methodology and more highly specialized models for correlated competing risks is one we seek to bridge.

In this work, we provide the estimator of the treatment effect in two-sample time-to-event studies and assess the effects of correlation between two competing risks on the estimation and robustness of the regression coefficients of the BVN model and the Cox model when data are generated under a non-PH model.

For the estimation of model parameters, the BVN model lends itself to the implementation of a novel Expectation Maximisation (EM) algorithm generalizing a similar algorithm for univariate survival proposed by Aitkin. Its objective can be maximum likelihood (ML) or maximum penalized likelihood (MPL) estimation. Our approach is related to one described for a bivariate normal stochastic censoring model example of Van Dyk et al. (1995, Section 4.1).

The paper is organized as follows. In section 2, we introduce the BVN linear model and the penalized likelihood for two competing risks. In section 3 we describe simulation studies conducted to: (i) assess the effect of correlation on the estimation of relative increase in survival in two-sample comparisons; and (ii) compare estimates provided by the BVN and the Cox PH models. In section 4 the BVN model is applied to a trial on the effects of adjuvant radiotherapy, as mentioned above. A summary and research questions for future work conclude the paper in section 5.

## 2. Linear survival models with dependent censoring

In this section, we establish notation and introduce the multivariate normal regression model of Deresa and Van Keilegom (the "DVK model") in the context of bivariate outcomes.

It is convenient in simulating an AFT model to measure times on a log scale, and to restrict attention to the occurrence of an event of interest (event 1) or the occurrence of a competing event (event 2). While more than two events might be considered, interest in many medical trials is focused on a single primary outcome, with any competing events distracting from the measurement of treatment effect, with few observations to individually model some causes. Thus pooling any competing risks together is common. We restrict discussion to two event causes.

Since a competing risk is an event whose occurrence precludes the occurrence of the primary event of interest, only the first-occurring event is observed; the observed time to an event is the minimum of two correlated times. Follow-up of events of types 1 and 2 are both subject to independent censoring at log-time  $C$ ; both event types are subject to the same censoring. This censoring is assumed to be non-informative.

Therefore, we observe  $(Y_j^o, D_j)$  for  $j = 1 \dots n$ , with  $Y_j^o = \min(Y_{j1}, Y_{j2}, C_j)$ , and  $D_j$  indicating event type or censoring. Here

$$D_j = \begin{cases} 1 & : \text{ if the event of interest is observed,} \\ 2 & : \text{ if an event of a competing risk is observed,} \\ 0 & : \text{ if no event is observed during follow-up.} \end{cases}$$

## 2.1. The DVK model for competing risks

Study of the effect of correlation between competing events presupposes the existence of a joint distribution of latent variables. These are latent failure times of each cause. However, this joint density is non-identifiable so estimation of parameters from competing events data is not possible in a fully nonparametric context (see Tsiatis, 1975). While identifiability is recovered by appropriate parametric assumptions, it remains important to evaluate the parametric model fit to observations.

First, consider the case of two competing events. Assume latent event times are pairs  $(Y_1, Y_2)$  with a bivariate Normal distribution, with a  $2 \times 2$  covariance matrix  $\Sigma$  common to all subjects. Further assume each pair has expected values specified in a linear model from covariates  $X_1, \dots, X_p$ . We refer to this as a BVN regression model for competing risks.

Specifically, assume that observations are an i.i.d. sample of size  $n$  from random vector  $(Y_1, Y_2, X_1, \dots, X_p, C)$ , with the conditional distribution of  $(Y_1, Y_2)$ , given  $X_1 = x_1, \dots, X_p = x_p$ , being bivariate normal with mean vector  $\boldsymbol{\mu} = (\boldsymbol{x}'\boldsymbol{\beta}_1, \boldsymbol{x}'\boldsymbol{\beta}_2)$  and covariance matrix  $\Sigma$ , of dimension  $2 \times 2$ . Here  $Y_1$  denotes the log-time to the event of interest ( $k = 1$ ) and  $Y_2$  to an event of (any) other cause ( $k = 2$ ), with  $\boldsymbol{x} = (x_1, \dots, x_p)'$ . Regression coefficients  $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$  form the columns of a  $p \times 2$  matrix  $\boldsymbol{B}$ . Last, random variable  $C$  denotes the log-time to independent censoring. Assume hereafter that

- (1)  $(Y_1, Y_2)'$  and  $C$  are conditionally independent given  $X = (X_1, \dots, X_p)'$ , and that
- (2)  $Y_1 - \mu_1, Y_2 - \mu_2$ , and  $C$  are independent of  $X$ .

The DVK model extends the bivariate normal model to the multivariate case and applies a parameterized transformation model to the log survival times of each cause. In it, covariates are specified to be cause-specific. Thus the DVK model is

$$\Lambda_\alpha(Y_k) = \boldsymbol{x}'_k \boldsymbol{\beta}_k + \epsilon_k, \quad (1)$$

for causes  $k = 1, \dots, m$ , where  $\Lambda_\alpha(\cdot)$ , is a generic parametric ( $\alpha$ ) class of monotone increasing transformations. The error vector  $(\epsilon_1, \epsilon_2)$  is bivariate normal, mean vector  $\mathbf{0}$ , covariance matrix  $\Sigma$ .

In order to simplify notation we confine discussion in following sections to the BVN model ( $m = 2$ ) with all covariates for each cause in common and simple log transformation of survival times ( $\Lambda : y \rightarrow y$  above). We use the more restrictive assumptions of the bivariate normal model to simplify notation in later sections; this is particularly appropriate when developing the EM algorithm. We shall rely in data analysis on an R package `bnc` prepared for bivariate normal censored data (Hudson et al., 2019).

## 2.2. Cause-specific hazards of competing events

With data on the log scale, the cause-specific hazard function to a first-occurring event of cause 1 induced by the BVN joint distribution of  $Y_1, Y_2$  is

$$\begin{aligned} \lambda_1(y; \boldsymbol{B}, \Sigma) &= \lim_{dy \rightarrow 0} \mathbb{P}[Y_1 \in (y, y + dy) | Y_1 > y, Y_2 > y] / dy \\ &= \lim_{dy \rightarrow 0} \mathbb{P}[Y_1 \in (y, y + dy), Y_2 > y] / [\mathbb{P}[Y_1 > y, Y_2 > y] dy]. \end{aligned}$$

Let  $\Phi(z)$  denote the survival function of the univariate standard normal distribution and  $\phi(z)$  the standard normal density (see Appendix A). Let  $\sigma_1, \sigma_2$  denote standard deviations of  $Y_1, Y_2$ . Define standardized values  $a, b$  of  $y$  under the two marginals by  $a = (y - \mu_1)/\sigma_1$ ,  $b =$

$(y - \mu_2)/\sigma_2$ , where  $\mu_1 = \mathbf{x}'\beta_1$  is the expectation of  $Y_1$  and  $\mu_2 = \mathbf{x}'\beta_2$  is the expectation of  $Y_2$ . Then the numerator of the above expression for the hazard function becomes

$$\begin{aligned} \lim_{dy \rightarrow 0} \mathbb{P}[Y_1 \in (y, y + dy), Y_2 > y] / dy &= f_{Y_1}(y) \mathbb{P}[Y_2 > y | Y_1 = y] \\ &= \sigma_1^{-1} \phi(a) \Phi\left(\frac{b - \rho a}{\sqrt{1 - \rho^2}}\right) \end{aligned}$$

Here  $f_{Y_1}(y) = \sigma_1^{-1} \phi(a)$  is the univariate normal density of  $Y_1$  and  $\rho$  is the correlation between  $Y_1$  and  $Y_2$ . The final equality for the numerator follows because the conditional distribution of  $Y_2$  given  $Y_1$  is univariate normal:

$$[Y_2 | Y_1 = y] \sim \mathcal{N}\left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1} (y - \mu_1), \sigma_2^2 (1 - \rho^2)\right).$$

Thus, defining the joint *bivariate* normal survival function as  $S_{12}(y_1, y_2) = \mathbb{P}[Y_1 > y_1, Y_2 > y_2]$ , we have

$$\lambda_1(y; \mathbf{B}, \boldsymbol{\Sigma}) = \sigma_1^{-1} \phi(a) \Phi\left(\frac{b - \rho a}{\sqrt{1 - \rho^2}}\right) / S_{12}(y, y). \quad (2)$$

The corresponding hazard  $\lambda_2(y; \mathbf{B}, \boldsymbol{\Sigma})$  for time to observing the competing risk is readily obtained by exchanging  $a$  and  $b$ . The sum of these hazards equals the hazard of time to the first event, their ratio determines the conditional probability of event type for first events at log time  $y$ .

### 2.3. Likelihood function and inference

In this section, we set notation for censoring outcomes. We then define a likelihood function for bivariate normal observations censored by a competing risk.

In competing risks data all observations are subject to censoring, not only by end of follow-up but also by competing events. When  $D_j = 1$ , time to event 2 is censored by an observed event of cause 1; i.e.  $Y_{j1} = y_j^o$  is observed and  $Y_{j2} > y_j^o$  is censored. Similarly, when  $D_j = 2$ ,  $Y_{j2} = y_j^o$  is observed and  $Y_{j1} > y_j^o$ . Finally,  $D_j = 0$  when times to both events exceed the period of follow-up:  $Y_{j1} > y_j^o$ ,  $Y_{j2} > y_j^o$  for observed end time of follow-up  $Y_j^o = y_j^o$ . The distribution of  $C_1, \dots, C_n$  need not be included in Likelihood calculations when censoring is independent and noninformative.

Hence the *Likelihood function* for competing risks observations ( $Y_j^o = y_j^o, D_j = d_j; j = 1, \dots, n$ ) is defined as:

$$\begin{aligned} L(\mathbf{B}, \boldsymbol{\Sigma}; \mathbf{y}^o, \mathbf{d}) &= \prod_{j:d_j=1} f_{Y_1}(y_j^o) \mathbb{P}[Y_2 > y_j^o | Y_1 = y_j^o] \times \prod_{j:d_j=2} f_{Y_2}(y_j^o) \mathbb{P}[Y_1 > y_j^o | Y_2 = y_j^o] \\ &\quad \times \prod_{j:d_j=0} S_{12}(y_j^o, y_j^o), \\ &= \prod_{j:d_j=1} f_{Y_1}(y_j^o) S_{2|1}(y_j^o | y_j^o) \prod_{j:d_j=2} f_{Y_2}(y_j^o) S_{1|2}(y_j^o | y_j^o) \prod_{j:d_j=0} S_{12}(y_j^o, y_j^o), \end{aligned} \quad (3)$$

where  $S_{1|2}(a|b) = \mathbb{P}[Y_1 > a | Y_2 = b]$ ,  $S_{2|1}(a|b) = \mathbb{P}[Y_2 > a | Y_1 = b]$ , and, as before,  $S_{12}(a, b) = \mathbb{P}[Y_1 > a, Y_2 > b]$ .

These probabilities depend on the design matrix  $\mathbf{X}$ , coefficient matrix  $\mathbf{B}$  and covariance matrix  $\Sigma$  of the bivariate normal distribution. In the linear model each observation has its own covariate values, a row vector of the covariate matrix  $\mathbf{X}$ . Let  $\mathbf{M} = \mathbf{X}\mathbf{B}$  be the matrix of expected values of  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$  where  $\mathbf{Y}_k = (Y_{1k}, \dots, Y_{nk})'$  for  $k = 1, 2$ . Then  $\mathbf{m}_k = \mathbf{X}\beta_k$  is the vector containing expected times to event  $k$ . Define  $\mathbf{z}_k = (\mathbf{y}^o - \mathbf{m}_k)/\sigma_k$ , with  $\sigma_k = \sqrt{\sigma_{kk}}$ , for cause  $k = 1, 2$ . Then  $(\mathbf{Y} - \mathbf{M})\mathbf{W} \sim \text{BVN}(\mathbf{0}, \mathbf{R})$  where the  $2 \times 2$  weight matrix  $\mathbf{W}$  is diagonal with entries  $\sigma_1^{-1}, \sigma_2^{-1}$ , and  $\mathbf{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ . Thus, given the covariate matrix  $\mathbf{X}$ , all likelihood terms may be expressed in terms of the standard bivariate normal distribution; the likelihood is a function of  $\rho$  and values  $\mathbf{z}_1, \mathbf{z}_2, \mathbf{d}$  dependent on the observed random sample.  $\mathbf{z}_1, \mathbf{z}_2$  are themselves functions of  $\mathbf{y}^o$  and parameters  $\beta_1, \beta_2, \sigma_1, \sigma_2$ .

The well-known separability of the likelihood by competing risk (Kalbfleisch and Prentice, 2002, Ch.8) in AFT models permits columns of  $\mathbf{B}$  to be estimated individually after censoring events of other causes. However, this separability depends on parameters not being common to components of different causes. For the likelihood of the BVN model, some parameters, specifically those of  $\Sigma$ , are involved in factors for each cause. Only in the case  $\rho = 0$  will the optimization simplify. As noted above, a correlated BVN distribution assumption can be replaced by the lesser assumption of cause-specific hazards  $\lambda_1(\cdot), \lambda_2(\cdot)$  defined in equation (2) in order to estimate  $\mathbf{B}$  by Maximum Likelihood. However, in the correlated BVN distribution, these cause-specific hazards will be functions of  $\Sigma$ , which we treat as unknown. Therefore, the optimization available for  $\rho = 0$  is not readily available with correlated competing risks (even when the correlation is known).

#### 2.4. Expectation Maximization algorithm

An EM algorithm may be employed to fit the BVN linear model parameters  $\mathbf{B}, \Sigma$ . EM algorithms provide a steady assured convergence to ML or MPL solutions. In each iteration, an EM algorithm *imputes* sufficient statistics of missing data (latent times of each unobserved event), substituting expected values of sufficient statistics conditional on observed data and current parameter estimates (E step). New parameter estimates are then obtained from these imputed sufficient statistics as though they originated from a complete sample (the M-step, Dempster et al. (1977)).

An EM algorithm for the likelihood function of survival  $Y_1$ , *not* subject to censoring by a competing risk, is described by Aitkin *op.cit.* For the likelihood function (3) for *bivariate* observations  $(\mathbf{Y}^o, \mathbf{D})$ , we introduce a new EM algorithm summarised below as Algorithm 1:

Here, current estimates of the model parameters in iteration  $i$  are  $\theta^{(i)} = (\mathbf{B}^{(i)}, \Sigma^{(i)})$ . The new estimate  $\mathbf{B}^{(i+1)}$  of  $\mathbf{B}$  is obtained as  $\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$  with imputed  $\mathbf{Y}$ . For this estimation, when  $D_j = 2$ , so that  $Y_{j1}$  is censored by the event of cause 2 at  $y_j^o$ , the censored observation  $Y_{j1}$  is imputed by  $\mathbb{E}[Y_1 | Y_1 > y_j^o, Y_2 = y_j^o]$ , where  $\mathbf{M}^{(i)} = \mathbf{X}\mathbf{B}^{(i)}$ . When  $D_j = 0$ ,  $Y_{j1}$  is imputed by  $\mathbb{E}[Y_1 | Y_1 > y_j^o, Y_2 > y_j^o; \theta^{(i)}]$ , with a similar expression for the imputation of  $Y_{j2}$ . The expected values are calculated in each case by using the current iteration's matrix of means  $\mathbf{M}^{(i)}$  and covariance matrix  $\Sigma^{(i)}$ , and the conditional expectations of Appendix A.

A complete data sufficient statistic for the covariance matrix  $\Sigma$  is  $\mathbf{V} = (\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}) = \mathbf{Y}'\mathbf{Q}\mathbf{Y}$  for known projection matrix  $\mathbf{Q} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ . The EM update to  $\Sigma$  therefore includes imputation of quadratic terms (squares and cross-products) in  $(Y_1, Y_2)$ . The new estimate is obtained using the censored observation  $y_j^o$  by replacing linear terms as above and quadratic terms using appropriate conditional distributions. For  $j' \neq j$  the statisti-

---

**Algorithm 1:** EM algorithm for BVN model

---

```
1 Initialisation:  $\theta^0 = (\mathbf{B}^0, \Sigma^0)$ . We used initial values:  $\rho^{(0)} = 0$ ;  $\beta_k^{(0)} = 0$ ,  $\sigma_k^{(0)} = 1$ , for
    $k = 1, 2$ . Let  $\mathbf{Q} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ .
2 for  $i = 0, 1, 2, \dots$  do
3   E-step : replace the unobserved variables by their conditional expectations given
    $\mathbf{Y}^o, \mathbf{D}, \mathbf{X}, \mathbf{C}$  at  $\theta^{(i)}$ :
4   if  $D_j = 2$ ,  $Y_{j1} \leftarrow \mathbb{E}[Y_{j1} | Y_{j1} > y_j^o, Y_2 = y_j^o; \theta^{(i)}]$ ,
    $Y_{j1}^2 \leftarrow \mathbb{E}[Y_{j1}^2 | Y_{j1} > y_j^o, Y_{j2} = y_j^o; \theta^{(i)}]$ ,
5   if  $D_j = 1$ ,  $Y_{j2} \leftarrow \mathbb{E}[Y_2 | Y_1 = y_j^o, Y_2 > y_j^o; \theta^{(i)}]$ ,
    $Y_{j2}^2 \leftarrow \mathbb{E}[Y_{j2}^2 | Y_{j1} = y_j^o, Y_{j2} > y_j^o; \theta^{(i)}]$ ,
6   if  $D_j = 0$ ,  $Y_{j1} \leftarrow \mathbb{E}[Y_1 | Y_1 > y_j^o, Y_2 > y_j^o; \theta^{(i)}]$ ,
    $Y_{j1}^2 \leftarrow \mathbb{E}[Y_1^2 | Y_1 > y_j^o, Y_2 > y_j^o; \theta^{(i)}]$ ,  $Y_{j2}^2 \leftarrow \mathbb{E}[Y_2^2 | Y_1 > y_j^o, Y_2 > y_j^o; \theta^{(i)}]$ ,
    $Y_{j1}Y_{j2} \leftarrow \mathbb{E}[Y_{j1}Y_{j2} | Y_{j1} > y_j^o, Y_{j2} > y_j^o; \theta^{(i)}]$ 
7   M-step : Maximize the likelihood of sufficient statistics  $\mathbf{X}'\mathbf{Y}$  and  $\mathbf{Y}'\mathbf{Y}$  of
   equations (B2), (B3) of Appendix B. Calculate  $\mathbf{B}^{(i+1)}, \Sigma^{(i+1)}$  using equations
   (B1), (B4), (B5) to define  $\mathbf{Y}^{(i)}$ , then equations (B6) and (B7):
8    $\mathbf{Y}^{(i)} \leftarrow \mathbb{E}[\mathbf{Y} | \mathbf{Y}^o, \mathbf{D}, \mathbf{X}, \mathbf{C}; \theta^{(i)}]$ 
9    $\mathbf{B}^{(i+1)} \leftarrow (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}^{(i)}$ 
10   $n\Sigma^{(i+1)} \leftarrow \mathbb{E}[\mathbf{Y}'\mathbf{Y} | \mathbf{Y}^o, \mathbf{D}, \mathbf{X}, \mathbf{C}; \theta^{(i)}] - \mathbf{Y}^{(i)'}\mathbf{Y}^{(i)} + \mathbf{Y}^{(i)'}\mathbf{Q}\mathbf{Y}^{(i)}$ 
```

---

cal independence of observations reduces calculations to imputation of linear statistics, but for  $j' = j$  more complex conditional expectations must be evaluated. For example, when  $D_j = 1$ , the quadratic term  $Y_{j1}^2$  is a known quantity. But, when  $D_j = 2$ , a quadratic term  $Y_{j1}^2$  is imputed using  $\mathbb{E}[Y_{j1}^2 | Y_{j1} > y_j^o, Y_{j2} = y_j^o]$ . Similarly, when  $D_j = 0$ , the same term is imputed as  $\mathbb{E}[Y_{j1}^2 | Y_{j1} > y_j^o, Y_{j2} > y_j^o]$ . When  $D_j = 0$ , we further require  $\mathbb{E}[Y_{j1}Y_{j2} | Y_{j1} > y_j^o, Y_{j2} > y_j^o]$ .

Again these conditional expectations are calculated assuming current parameter values  $\mathbf{B}^{(i)}, \Sigma^{(i)}$ , which provide the mean vector  $(m_{j1}^{(i)}, m_{j2}^{(i)})'$ , and then applying results of Appendix A. We provide all required moment results for the EM algorithm in a bivariate normal censored linear model in this Appendix, and include the corresponding R-code in the package `bnc`. Further details of the EM algorithm are given in Appendix B. Both the EM algorithm and these results for the bivariate normal distribution appear to be new and may prove useful in other contexts.

Standard errors of the EM algorithm solution are available using the numerical differentiation of Fisher scores (NDS) method of Jamshidian and Jennrich (2000) and implemented in the R package `turboEM` (Bobb and Varadhan, 2018). The score statistic is derived from evaluations of the score function  $Q(\theta, \tilde{\theta})$  which is accessible for computation using our probability results for the BVN complete data.

### 2.5. Mildly Penalize Likelihood

Convergence to a boundary point of the parameter space can introduce issues for optimization methods. Optimizers such as Newton's method break down near boundaries when iterations take them outside a constrained domain (such as  $|\rho| \leq 1$ ). Specialized methods are available for constrained domains, but may affect speed of convergence of the algorithm. In particular, the EM algorithm's rate of convergence is reduced when  $\theta$  lies on or near a boundary (Ng and McLachlan, 2003). This is problematic when conducting large numbers of simulations.



To accelerate EM’s convergence Bobb and Varadhan introduced the `squareEM` option in R package `turboEM`.

Rate of convergence is particularly affected in regions of flat likelihood. It is common to observe improvement after introducing a penalty term in the likelihood (Green, 1990; Cole et al., 2013).

Therefore, we regularize by penalizing log-Likelihood  $\log L = \log L(\mathbf{B}, \mathbf{\Sigma}; y^o, \mathbf{d})$  as:

$$\log L \leftarrow \log L - (\nu + 3) \times \log(\det(\mathbf{\Sigma})/2 - \kappa \operatorname{tr}(\mathbf{\Sigma}^{-1})/2)$$

where

$\nu$  is a positive integer, the degrees of freedom parameter of the inverse-Wishart prior,  $\mathbf{\Sigma} \sim \mathcal{IW}(\nu, \kappa \mathbf{I})$ , whose posterior distribution provides the penalty term (Zhang, 2021). Higher degrees of freedom increasingly encourage a solution consistent with the prior mean, for which we specify a multiple (default  $\kappa = 1$ ) of the identity matrix.

We then utilize the `turboEM` algorithm for this MPL, using an accelerated EM form (method `squareEM`). The resulting algorithm for bivariate normal maximum penalized likelihood (MPL) estimation is implemented in our R fit to censored data using function `bnc::bnc()`.

### 3. Simulation studies

In this section, we study the parametric estimation of first-event data from correlated bivariate-normal competing risks. Goals of the simulation are as described in the final paragraph of the Introduction. Performance criteria are sampling distribution (density plots and boxplots displaying quantiles) of marginal treatment effect for cause 1. In two-sample data sets of first-event times, we evaluate the estimation of the treatment effect, the mean difference in log time to event 1 between treated and control subjects.

#### 3.1. Simulation design

We generated censored data using the latent variable approach for a BVN linear model. The latent log-survival-times  $(Y_1, Y_2)$  were a random sample of  $n$  bivariate normally distributed observations with means  $\mathbf{M} = \mathbf{X}\mathbf{B}$  determined by the two-sample design, and fixed covariance matrix  $\mathbf{\Sigma}$  specified by choices  $\sigma_1, \sigma_2$ , and  $\rho$ .

The parameter  $\tau$  denotes the maximum length of follow-up and is chosen to fix the censoring proportion  $S_1(\tau)$  of events of interest not occurring before end of follow-up, or equivalently the marginal cumulative incidence  $P_1(\tau) = 1 - S_1(\tau)$  of events of interest. This proportion does not include the censoring of the event of interest by the competing event. Marginal cumulative incidence matches the observed CIF only when the competing event is absent or very late occurring.

In two-sample simulations with design covariates  $X_1 = 1$  (constant) and  $X_2$  with independent Bernoulli distributed values 0 and 1 (indicating treatment), the randomized trial treatment allocation produces mean vectors that differ only between treatment and control arms. The linear model constant coefficient sets baseline survivals while  $\beta_{k2}$  sets treatment effects. Then  $\boldsymbol{\beta}_k = (\beta_{1k}, \beta_{2k})'$ , so  $\mathbf{M} = (m_{jk})$  is an  $n \times 2$  matrix, with  $m_{jk} \in \{\mu_k^C, \mu_k^T\}$  for  $k = 1, 2$ . Group means  $\mu_k^C = \beta_{1k}$  and  $\mu_k^T = \beta_{1k} + \beta_{2k}$  were specified, for  $k = 1, 2$ , as the mean log times to event  $k$  applying to all observations in Control and Treatment groups, respectively. We fixed the mean in the Control group to be  $\mu_1^C = 0$ , and  $\sigma_1 = \sigma_2 = 1$ . *Treatment effect* is defined as the parameter  $\Delta = \mu_1^T - \mu_1^C = \beta_{21}$ , the expected difference between Treated and Controls in

log-time to the event of interest. Then  $\exp(\Delta)$  is the median ratio (MR) on the original time scale. By setting  $\beta_{22} = 0$ , we specified that the competing event (event 2) has the same risk of occurrence in both arms:  $\mu_2^T = \mu_2^C = \mu_2$ .

### 3.2. Hazard ratios of the BVN linear model

Including a treatment indicator in a two-sample BVN model provides an estimator of MR, the ratio of median survival time in treatment versus control populations. Comparisons are made between the parametric method and standard semi-parametric methods – Cox’s proportional cause-specific hazard models.

As is well known, proportional hazard models differ from AFT models except with Weibull distributional assumptions. Consequently, lognormal modelling implies a time-dependent HR, violating the proportional hazards assumption of cause-specific hazard models. The hazard ratio of Treatment versus Control for time to event 1 in the case  $\rho = 0$  is

$$HR_1(t) = \frac{\lambda_1^T(t)}{\lambda_1^C(t)} = \frac{\lambda_1(t; \Delta, \sigma_1)}{\lambda_1(t; 0, \sigma_1)}$$

for hazard function

$$\lambda_1(t; \Delta, \sigma_1) = \frac{1}{t\sigma_1} \Psi\left(\frac{\log t - \Delta}{\sigma_1}\right)$$

where  $\Psi$  is the hazard function of the (univariate) standard normal distribution (see Appendix A). So, for independent competing causes,

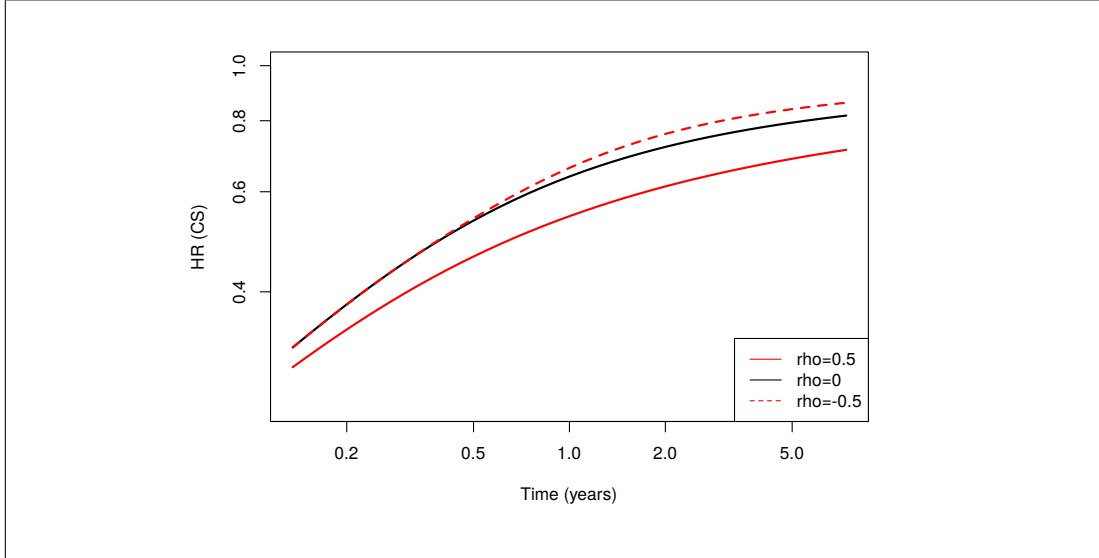
$$HR_1(t) = \frac{\Psi((\log t - \Delta)/\sigma_1)}{\Psi(\log t/\sigma_1)}$$

Note this HR is independent of  $\beta_{12}$  only when  $\rho = 0$ .

For dependent competing risks, the same HR depends also on parameters  $\rho$ ,  $\mu_2$  and  $\sigma_2$ . It is computed from the cause-specific hazard function for the event of interest in the bivariate normal model, as the ratio of its evaluation, at a given time, in case and control subjects. This cause-specific hazard function (referred to as a sub-density) is explicit in the Supplement to Deresa and Van Keilegom (2020a) and as R function `bnc::bnc_lambda()`. Figure 1 displays the hazard ratio of the lognormal distribution and its dependence on  $\rho$  in a two-sample comparison with two competing causes. The mean log times to the event of interest and competing event in the Control group are  $\beta_{11} = \beta_{12} = 0$ ; the mean log time to the event of interest on treatment is  $\Delta = \beta_{21} = 0.5$ . The mean log time to the competing event is unaffected by treatment ( $\beta_{22} = 0$ ) and  $\sigma_1 = \sigma_2 = 1$ . In this case (a strongly competitive competing risk unaffected by treatment) the hazard ratio is time-dependent, and smaller for  $\rho = 0.5$  than for the other two choices, meaning that treatment benefit is greater.

### 3.3. Estimation of marginal treatment effect.

In a two-sample comparison, where subjects are randomized equally to treatment and control arms, we focus on estimating the treatment effect. Varying the correlation allows us to assess its influence on the estimated treatment-benefit using the BVN linear model or using a standard approach (Cox proportional hazard model).



**Figure 1.:** Hazard ratio of two-sample model with log normal error distribution varying  $\rho \in \{-0.5, 0, 0.5\}$ . Hazard ratio is for treatment versus control group with treatment benefit  $\beta_{21} = \Delta = 0.5$ , with  $\sigma_{11} = \sigma_{22} = 1$ . Other parameters are  $\beta_{11} = \beta_{21} = \beta_{22} = 0$ .

We simulated data from a clinical trial with two arms as above, imposing a balanced design with equal numbers  $n/2$  of patients in treatment and control groups. The generated data follows an AFT model with expected log-survival-time  $m_{jk} = \beta_{1k} + \beta_{2k} X_j$ , for event cause  $k = 1, 2$ , subject  $j = 1, \dots, n$ , with covariate  $X_j$  the indicator of the randomized allocation of subject  $j$  to a treatment group. A lognormal distribution has expected value  $\exp(\mu + \sigma^2/2)$ , so the AFT assumption implies the ratio of median survivals for cause 1,  $MR_1 = \mathbb{E}[\exp(Y_{j1}|X_j = 1)] / \mathbb{E}[\exp(Y_{j1}|X_j = 0)] = \exp(\beta_{21})$  when variances  $\sigma_k^2$  are independent of treatment allocation.

The estimation methods considered were:

1. HR estimates from Cox's partial likelihood estimator ("Cox model") treating competing events as independent censoring. We fitted hazard ratio estimates of event 1 with the function `survival::coxph()`.
2. Regression coefficient estimates in the Generalized Estimating Equations (GEEs) fixing correlation: the Likelihood scores fix correlation  $\rho = 0$ . We assume the competing event is an (additional) independent censoring. We fitted regression coefficients of event 1 with the function `survival::survreg()` which implements MPL solution with  $\rho$  fixed, in the particular case  $\rho = 0$ . To fix  $\rho$  to other values, refer to Anderson and Olkin (1985).
3. Regression coefficient estimates in the BVN linear model using a `squareEM` algorithm. Estimation is implemented within our `bnc` package.

The Cox model estimates a time-averaged cause-specific hazard ratio (Kalbfleisch and Prentice, 1981). For graphical displays and comparisons, we estimated parameters of the proportional hazards models related to treatment effect  $\Delta$ . If the hazard functions of time to event 1 in Treatment and Control arms are each constant,  $\lambda_1^T$  and  $\lambda_1^C$  respectively, then the median ratio is  $MR = (1/\lambda_1^T)/(1/\lambda_1^C)$ , the reciprocal of the HR for such data. As  $1/MR = \exp(-\Delta)$ , the corresponding estimate of  $\Delta$  from the Cox model is the negative log of the HR. For comparability of HR and median ratio estimates, we therefore display the estimates of the negative logarithm of the hazard ratio (for Cox models) and of the treatment

effect  $\Delta = \mathbb{E}[Y_{jk}|X_j = 1] - \mathbb{E}[Y_{jk}|X_j = 0] = \beta_{21}$  (for BVN models and `survfit` models fixing  $\rho = 0$ ).

We varied parameters  $n = 100$ ,  $P_1(\tau) \in \{1.0, 0.8, 0.6\}$ ,  $\Delta, \beta_{12} \in \{0, 0.5, 1\}$  and  $\rho \in \{-0.5, -0.25, 0, 0.25, 0.5\}$ . Convergence of EM slows down considerably as  $\rho \rightarrow 1$ . This was a consideration when running large numbers of repeat simulations (remembering the need to vary treatment effects, differences in mean times to events of two causes and censoring fractions as well).

Simulation estimates from fitting the Cox model provide Figure 2. The Figure displays Cox model estimates of negative  $\log(\text{HR})$ , a measure of treatment benefit, which increasingly depart from the true difference  $\Delta$  in median survival (of log times) as the treatment benefit grows, or censoring increases.

Figure 3 shows that estimates of  $\Delta$  obtained under a parametric univariate AFT model – under marginal normality assumption fixing  $\rho = 0$  – from use of R package `survreg` are not robust for other  $\rho$ . The panels, for sample size  $n = 100$ , show the increasing bias of this GEE-like estimator as  $|\rho|$  grows.

Supplementary Table 2 shows bivariate normal MPL estimates  $\hat{\Delta}$  of  $\log \text{MR}$  for samples of size  $n = 1000$ . Medians of these treatment effect estimates deviate little from the target parameter  $\Delta$ . The BVN model improves estimation of treatment benefit (Figure 4) over that of a particular GEE solution fixing correlation.

The matrix plot (mayplot) Figure 4 is a graphical display containing boxplots of MPL estimates of  $\Delta$  for simulated data sets generated with parameters as shown. From it, we can assess variability in individual simulated data sets, in this case for the smaller sample size  $n = 100$ . Again observe that medians of treatment effect estimates  $\hat{\Delta}$  differ little from true values  $\Delta$ . The variability (measured by the inter-quartile range) is consistently small.

Alternative displays by density plots of corresponding estimates, Supplementary Figure 2, for  $n = 1000$ , confirm this finding and suggest that the sampling distributions of  $\hat{\Delta}$  are approaching normality, particularly as  $P_1(\tau)$  increases, and as  $\beta_{12}$  decreases.

These simulations demonstrate the influence of a correlated competing risk on the estimation of the treatment effect in BVN data. We found superior robustness of BVN estimation of marginal treatment effect to correlation when compared with fitting Cox PH models or a GEE approach using a marginal AFT after prespecifying  $\rho$ .

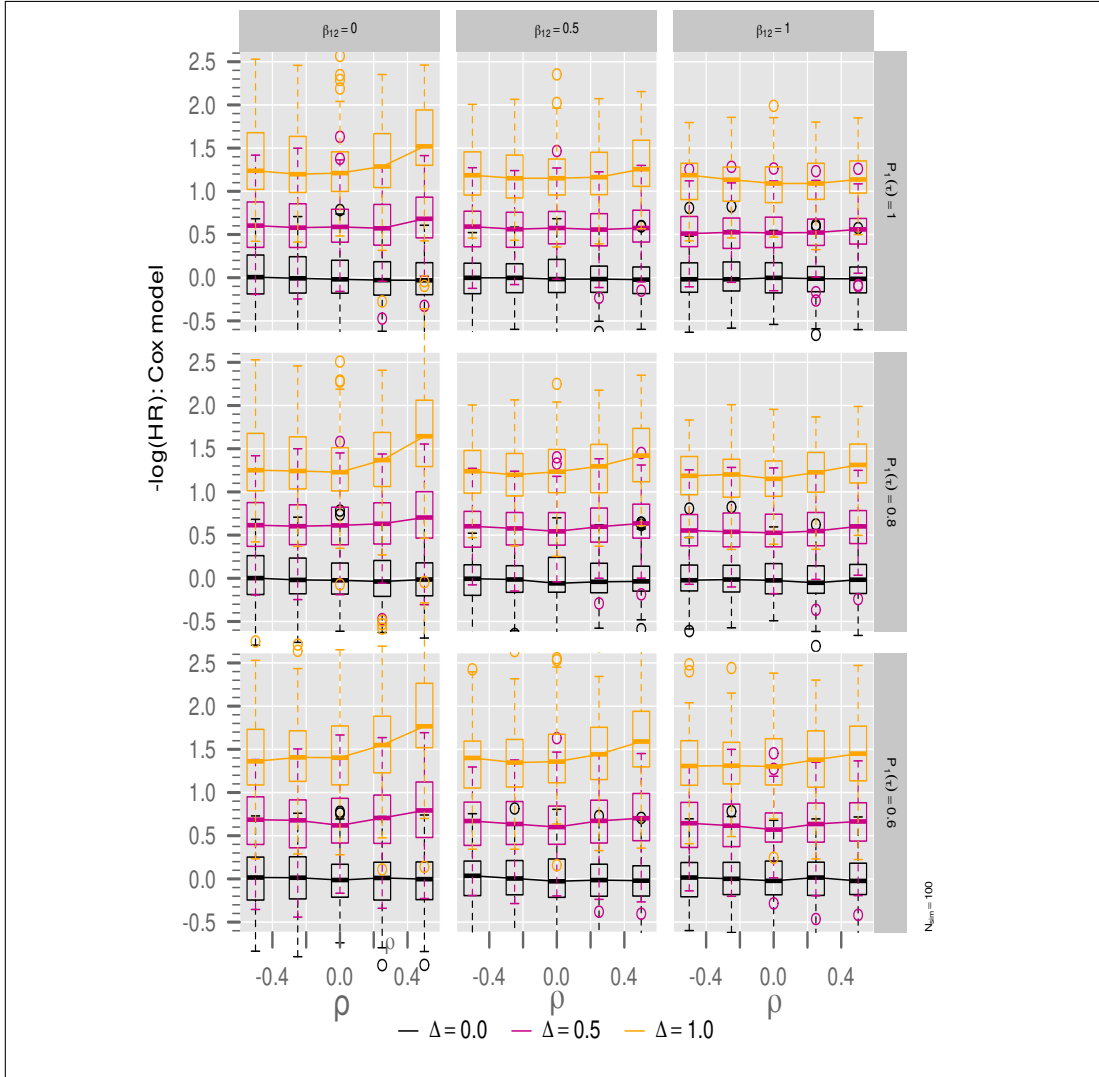
### 3.4. Stability of BVN estimation of correlation

While the BVN estimation of treatment effects is reliable and consistent, we considerable instability in MPL estimation of correlation  $\rho$  occurred. Supplementary Figure 1 shows density plots of  $\hat{\rho}$  in samples of size  $n = 1000$ . For all cases examined there is large variability in estimates; more so when there is no correlation or small correlation between competing events. More consistent estimation of  $\rho$  is not achieved even in samples of size  $n = 5000$ .

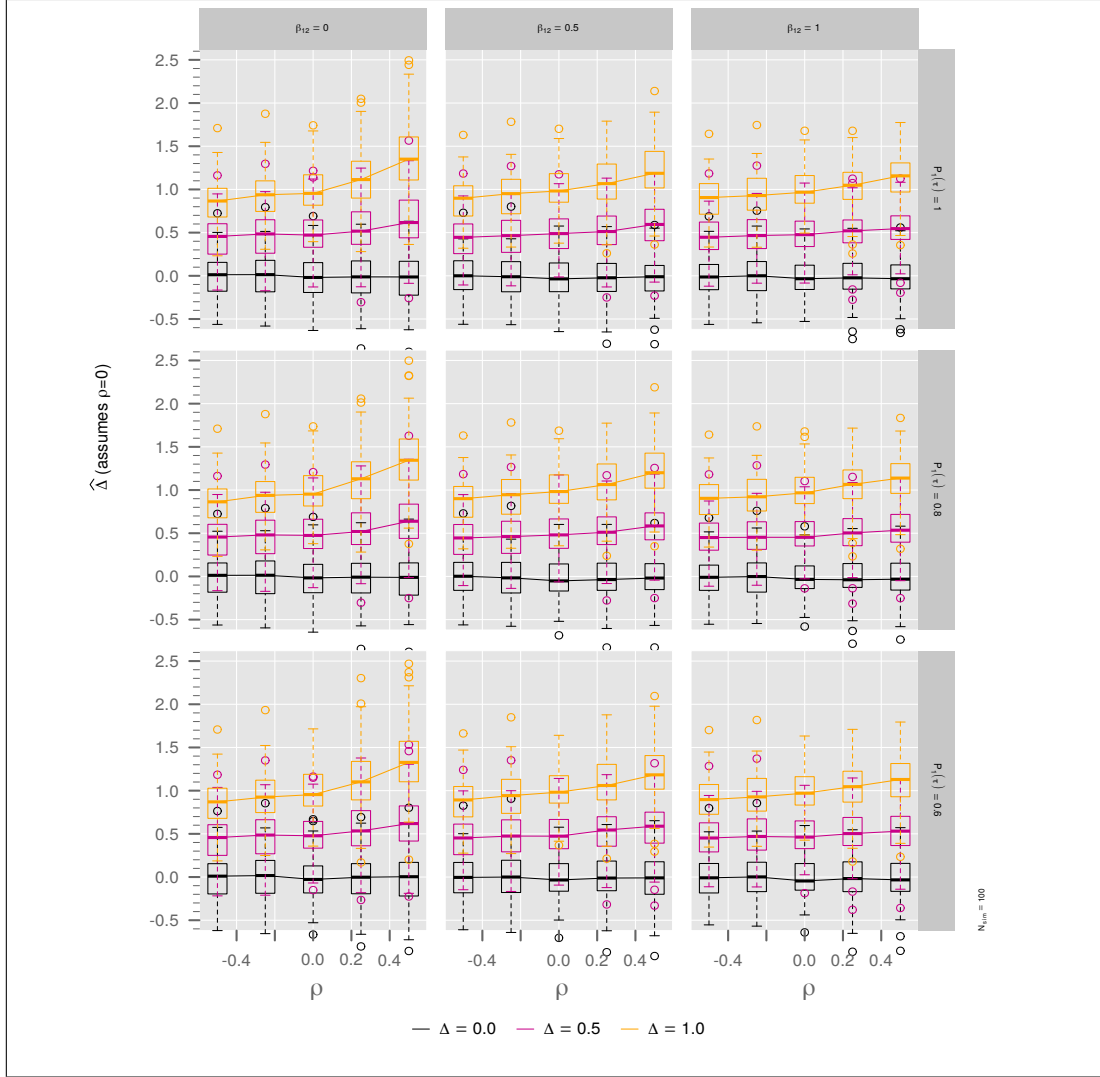
### 3.5. Robustness to bivariate non-normality

We simulated two-sample estimates of treatment effect with Frank-copula-generated data. Here  $(Y_1, Y_2)$  has a *non-normal* bivariate distribution but normal marginals. The level of association is controlled by the Frank parameter ( $\theta_F$ ) which is linked to Kendall and Pearson correlation coefficients Escarela and Carriere (2003). We varied parameters  $n = 100$ ,  $P_1(\tau) \in \{1.0, 0.8, 0.6\}$ ,  $\Delta, \beta_{12} \in \{0, 0.5, 1\}$  and  $\theta_F \in \{-3.31, -1.48, 0, 1.48, 3.31\}$  (linked respectively to  $\rho \in \{-0.5, -0.25, 0, 0.25, 0.5\}$ ).

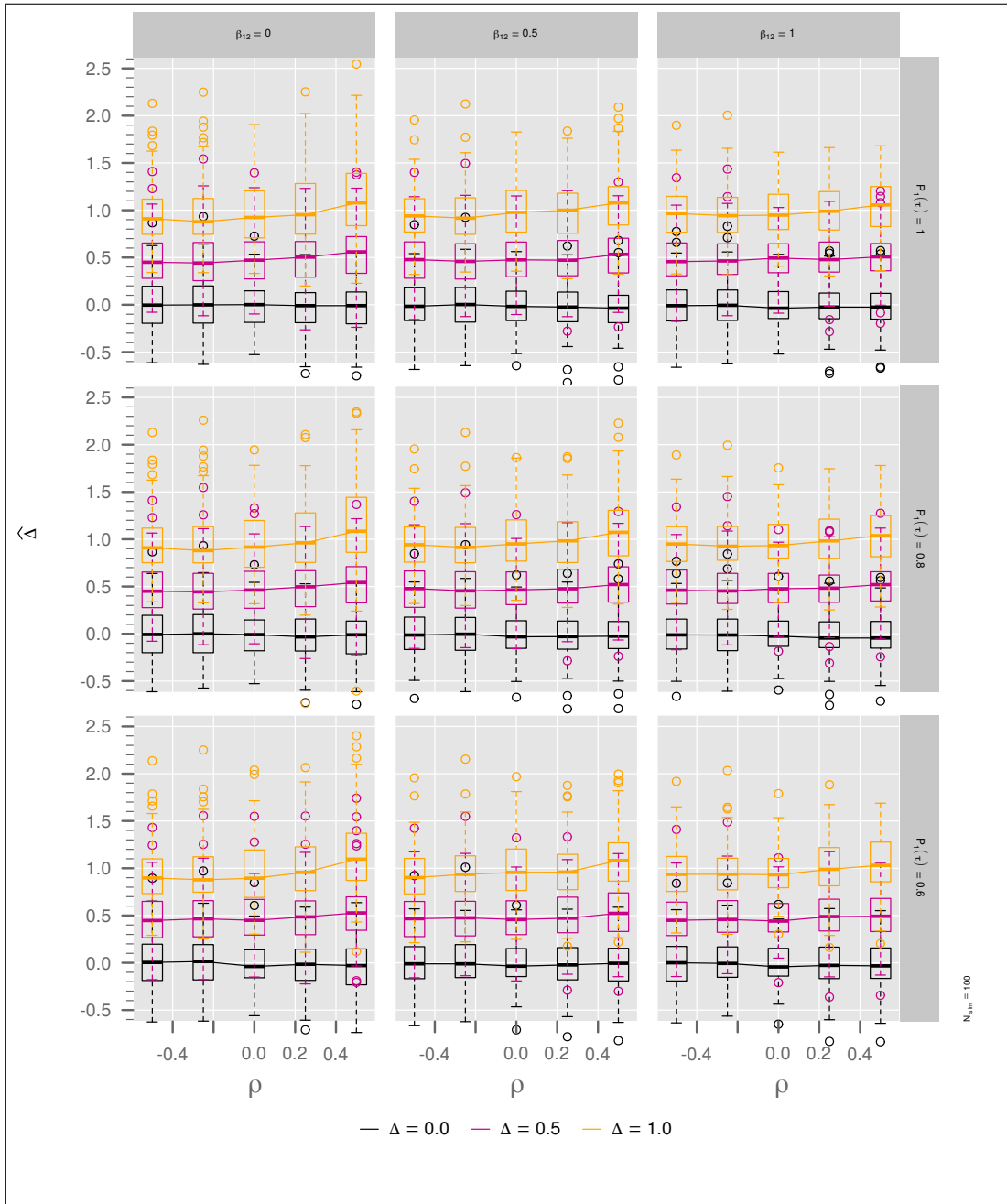
Supplementary Figure 5 displays the boxplot of simulations estimates of  $\Delta$  for those dif-



**Figure 2.: Cox model.** Boxplots of simulation estimates of the negative log hazard ratio ( $-\log \text{HR}$ ) of the Cox model,  $n = 100$ , for different value of  $\Delta \in \{0, 0.5, 1\}$ . Parameters in BVN data generation are :  $P_1(\tau)$  (rows), mean difference  $\mu_2$  (columns), correlation  $\rho$  (x-axis), with  $\mu_1^C = 0$ ,  $\beta_{22} = 0$  and  $\sigma_1 = \sigma_2 = 1$  fixed. The simulation comprized  $N_{\text{sim}} = 100$  samples.



**Figure 3.: Restricted MPL.** Boxplots of simulation restricted MPL estimates  $\widehat{\Delta} = \widehat{\beta}_{21}$  of treatment benefit for different expected benefits  $\Delta \in \{0, 0.5, 1\}$  and sample size  $n = 100$ . Estimation for fixed  $\rho = 0$  using R package `survreg`. BVN model parameters are :  $P_1(\tau)$  (rows), mean difference  $\mu_2$  (columns), correlation  $\rho$  (x-axis), with  $\mu_1^C = 0$ ,  $\beta_{22} = 0$  and  $\sigma_1 = \sigma_2 = 1$  fixed.  $N_{\text{sim}} = 100$ .



**Figure 4.: MPL.** Boxplot of simulation MPL estimates  $\hat{\Delta}$  of treatment benefit,  $n = 100$ . Parameters in BVN data generation varied in the graphic are:  $P_1(\tau)$  (rows), mean delay  $\beta_{12}$  (columns), correlation  $\rho$  (x-axis). Fixed parameters are  $\mu_1^C = 0$ ,  $\beta_{22} = 0$  and  $\sigma_1 = \sigma_2 = 1$ .  $N_{\text{sim}} = 100$ .

ferent parameters. For two-sample comparisons, with the small sample size  $n = 100$  and non-normal bivariate distribution, BVN regression estimates of the treatment effect  $\Delta$  show little bias. However, in some cases, convergence was not established even with 5000 iterations, and resets of initial values were necessary. As expected, accuracy in estimating treatment effect is less than for data generated as bivariate normal, particularly with strong positive association ( $\theta_F = 3.3$ , corresponding to  $\rho = 0.5$ ).

For copula generated data, GEE methods (Supplementary Figure 4) have worse performance than unrestricted MPL estimates, while Cox regression methods are biased (Supplementary Figures 3). All methods exhibit considerable variability in estimates.

#### 4. Example

O'Brien et al. (1997) studied the improvement in regional control after neck dissection following adjuvant radiotherapy in 143 patients receiving surgery at several sites for histologically positive metastatic malignant melanoma. This nonrandomized study documented prospectively a case series of cancer patients with histologically positive nodes receiving surgery. The trial compared a group of patients receiving adjuvant radiotherapy with the remaining patients who were not assigned adjuvant radiotherapy. The two competing risks appropriate to the study objective are local or regional recurrence. It is very plausible to believe that correlation exists between times to these two outcome types, perhaps due to common unmeasured covariates (age, ...) affecting both local recurrence and regional recurrence hazards.

Some patients underwent lymphadenectomies; recurrence data were obtained for all dissections ( $n=153$ ). We compare  $n = 46$  dissections with adjuvant radiotherapy (treatment group) with the remaining  $n = 107$  dissections (no radiotherapy, the control group). For the outcome, time from surgery to the first event, the event of interest was a local regional relapse, relapses at other sites or intercurrent death comprising competing risks. The events were recorded as "local regional relapse only" (Event 1, local relapse) or "distant relapse or death" (Event 2, distant relapse). Patients were censored at loss to follow-up or end-of-study.

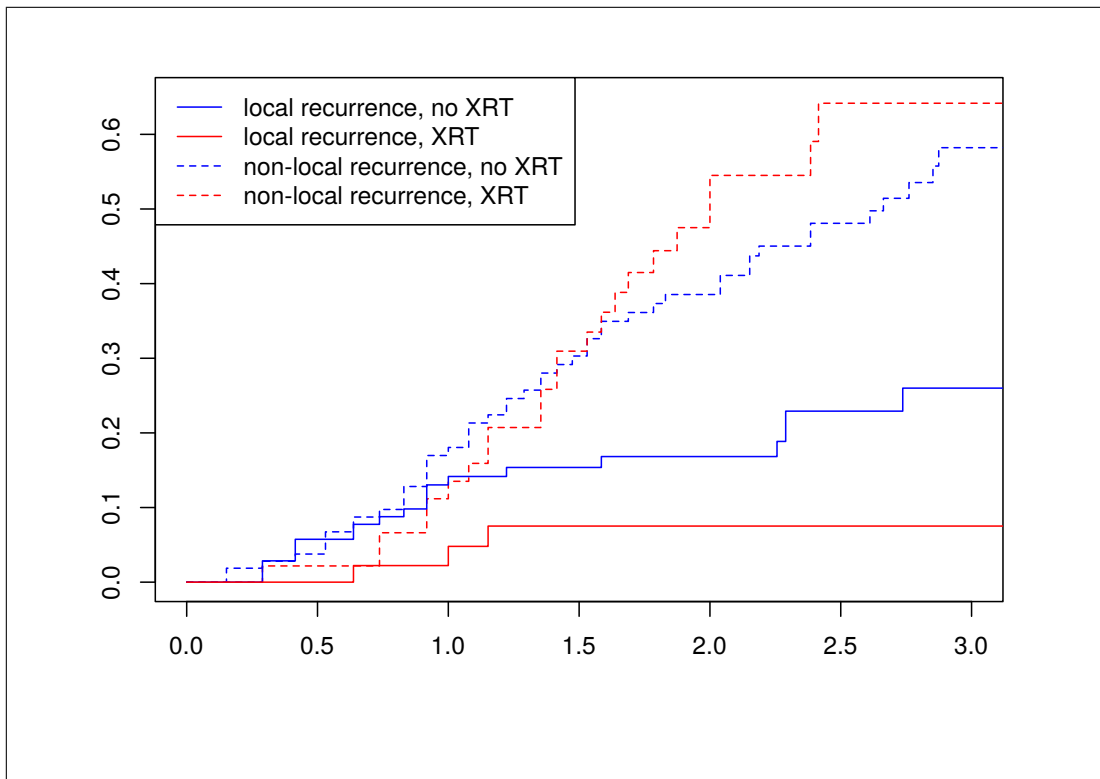
We used the transformation  $y : t \rightarrow \log(1 + t/0.75)$  for the relapse times (in years) to reduce skewness in the distribution of the observations and allow for deaths at the time of treatment (Supplementary Figure 6). Log transformations are commonly used to reduce skewness in distribution of non-negative data (Mosteller and Tukey, 1977); an offset is required for zeros. Observe that 5 years follow up corresponds to just under 3 transformed time units. This limit ( $y=3$ ) is also close to the 90-th percentile (3.02) of recorded survivals on the transformed time scale;  $y=3$  is the upper limit for our x-axis. The lower quartile (LQ) of time to event is 0.75 years ( $y=1$ ). We use offset 0.75 for the log transformation; the effect is initially (up to this LQ) linear but increasingly logarithmic as  $t$  increases above 0.75 years. We conducted a sensitivity analysis for the choice of the offset in Figures 7-9 in supplementary Appendix C. We noted little effect on fitted cumulative incidence, when referred to the original time scale.

Cumulative incidence, shown in Figure 5, records the number of events of each type recorded as time (from randomization) progresses. In each arm, local recurrences occur earlier than non-local recurrences. Local relapses occurred less frequently following adjuvant radiotherapy. The plots suggest longer time to local recurrence in the irradiated group but little difference between treatment groups in non-local recurrence.

Table 1 provides HR, 95% confidence interval (CI) and p-values of the Cox model. The 95% CI for the HR of interest is [0.1057, 1.21]. The treatment difference is not significant but approaching the level ( $\alpha = 0.05$ ) of statistical significance ( $p = 0.098$ ). This suggests local benefit of irradiation.

The BVN model was applied. Algorithm `squareEM` converged after 25 iterations. The es-





**Figure 5.:** Cumulative incidence of local and distant recurrence by treatment group (XRT=adjuvant radiotherapy). Nelson-Aalen estimates after  $y : t \rightarrow \log_2(1 + t/0.75)$  transformation of original time scale. See vignette **HNexample.Rmd** for code.

Method	Description	Factor	coef	exp(coef)	se(coef)	z	p-value
A.	Cox	adjuvant RT	-1.03	0.36	0.62	-1.65	0.10
B.	Fine-Gray	adjuvant RT	-1.05	0.35	0.61	-1.72	0.09

**Table 1.** Proportional hazard model results for local recurrence: A. Cox model; B. Fine and Gray model

Parameter	Estimate	Std.Error
$\beta_{11}$	2.633	0.123
$\Delta$	0.370	0.285
$\beta_{12}$	2.170	0.077
$\beta_{22}$	-0.026	0.212
$\sigma_1$	1.437	0.104
$\sigma_2$	1.198	0.045
$\rho$	0.897	-

**Table 2.** Estimates of BNC linear model parameters in the Head and Neck data using `squareEM` algorithm. The standard error of  $\hat{\rho}$  is not available, because of loss of precision. With default settings `squareEM` converges after 25 iterations. See vignette `HNexample.Rmd` for code.

timate of correlation in Table 2 was very high  $\hat{\rho} = 0.897$  and the median ratio was the factor  $\exp(0.370) = 1.45$ . This estimate of treatment effect compares median times to local recurrence of patients undergoing adjuvant radiotherapy with those without adjuvant radiotherapy. With adjuvant radiotherapy, the median time before recurrence is longer. However the parameter is not statistically significant ( $SE=0.285$ ,  $z=1.298$ ,  $p=0.19$ ), the difference remaining within the play of chance.

Confidence intervals for regression parameters can be formed using the asymptotic normality of maximum likelihood estimators – though this would not be appropriate here for correlation, noting the large uncertainty in estimation of  $\rho$  and the estimate’s closeness to the upper boundary  $\rho = 1$ . This uncertainty can be better evaluated using bootstrap replications of `bnc` estimation to form percentile confidence intervals. The 95% confidence interval for  $\rho$  using 400 bootstrap replications was obtained as (0.86,0.93).

Point estimates of the BVN hazard ratio can also be calculated. The BVN point estimate  $HR=0.30$  at time  $t=0$  gives an estimate of the hazard ratio of local recurrence with adjuvant radiotherapy versus local recurrence without radiotherapy immediately at the time radiotherapy treatment was assigned. The calculations involved plugging in the fitted BVN parameter estimates to the hazard ratio function of Section 3.2.

The estimated HR of the bivariate normal distribution,  $HR_1(t)$ , by its definition and as discussed in Section 3.2, is in general time-dependent, but in this case fairly constant when computed using the BVN model with MPL parameter estimates.

Point estimates and one-sided confidence bounds for these HRs obtained using 400 bootstrap resamples are given in Table 3. Point estimates reduce gradually from 0.30, at  $y=0$  ( $t=0$ ), to 0.27, at  $y=3$  ( $t=5$  years, the 90th percentile of survival times). Thus the Cox model PH assumption appears reasonably consistent with the BVN model fitted; parametric analysis using bivariate normal assumptions can meaningfully be compared with semi-parametric Cox regression findings in this study.

The BVN model HRs both at start of the study ( $t=0$ ) and later ( $t=0.75$ ,  $t=5$  years) remain slightly less (estimating greater radiotherapy benefit) than the estimate of the Cox model ( $HR=0.35$ , 95% confidence interval (0.11,1.21)). These intervals are similar to that of the Cox estimate provided above.

In checking model fit, fitted cumulative incidence functions may be compared with empirical estimates. Figure 6, and Figure 7 in supplementary Appendix C, compare the fitted CIFs

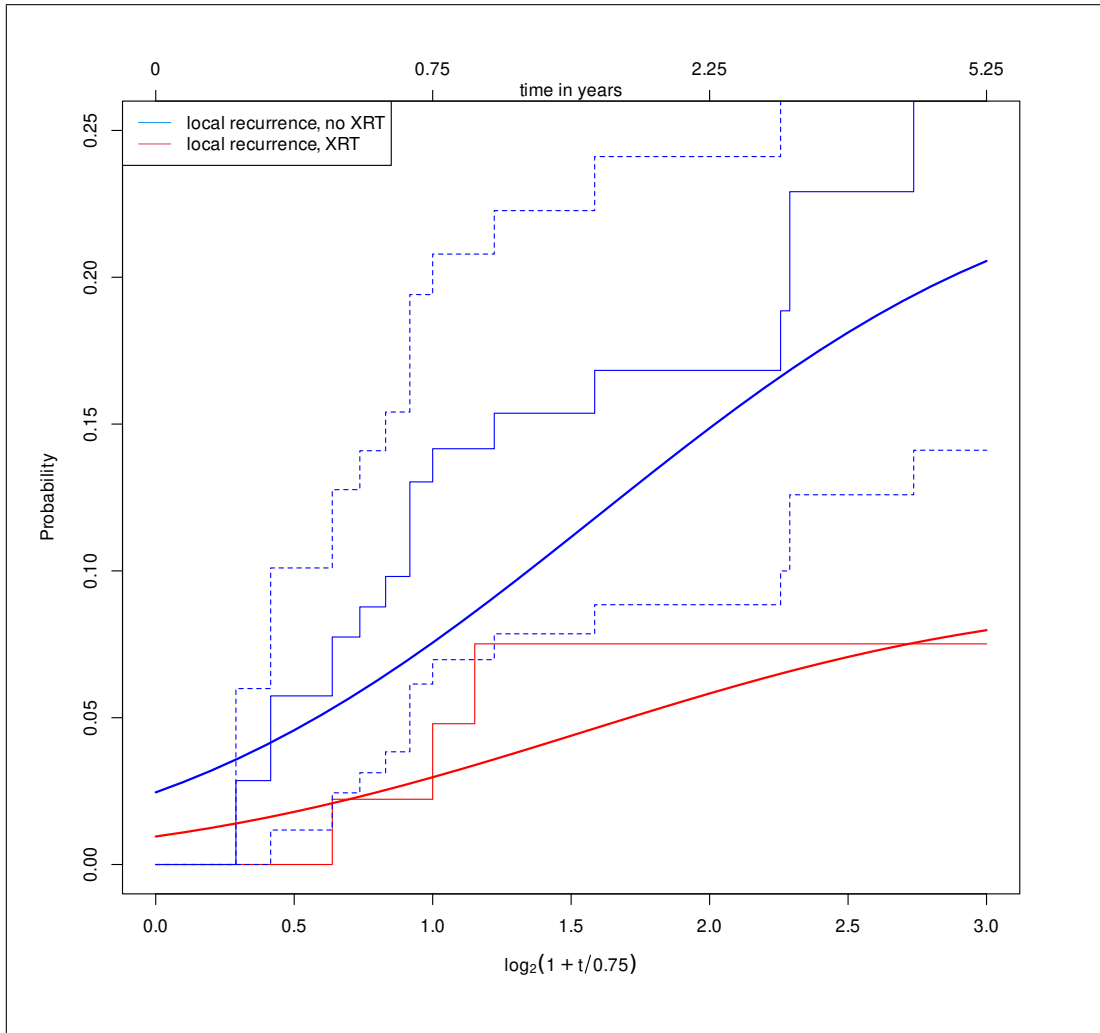
t	0.00	0.75	5.0
y	0.00	1.00	3.0
HR(t) estimate	0.30	0.29	0.27
lower confidence limit	0.01	0.00	0.00
upper confidence limit	1.06	1.00	0.97

**Table 3.** Point estimates and upper 95% two-sided confidence limits of the BVN HR at specified survival times (t=time in years, y=transformed by  $\log(1 + t/0.75)$ ). Each limit is a percentile of 400 bootstrap replications.

from the MPL estimates of the bivariate normal model with non-parametric Nelson-Aalen estimates. Fitted curves for patients receiving radiotherapy are shown in red, with solid lines indicating the study outcome, local recurrence. In Figure 6, corresponding CIFs shown in blue and the dotted blue confidence bounds of the non-parametric CIF are for study subjects not receiving radiotherapy.

Fitted CIFs in Figure 6 include both CIFs of the BVN model (smooth curves) and non-parametric Nelson-Aalen estimates (step functions). The agreement between the BVN model CIFs and empirical CIFs appears adequate. CIFs for both radiotherapy groups fall well within the confidence bands (for clarity, only displayed for the control group) for the empirical functions in this Figure.

In summary, the bivariate normal analysis supports the proportional hazards assumption of the Cox model, while offering lesser evidence of an apparent reduction in hazard or lesser incidence of local relapse in patients receiving adjuvant radiotherapy compared with patients receiving surgery alone. The BVN model estimates a very high level of correlation between log times to local and non-local recurrence.



**Figure 6.:** Cumulative incidence of local recurrence by treatment group (XRT=adjuvant radiotherapy). BNC parametric CIFs (smooth lines) superposed on Nelson-Aalen estimates (stepped lines) after  $y : t \rightarrow \log_2(1 + t/0.75)$  transformation of time scale. Dotted blue lines show the 95% confidence bounds for Nelson-Aalen estimates of the CIF in controls not receiving radiotherapy.

## 5. Conclusion

Multivariate survival outcomes are of importance in randomized clinical trials (Prentice and Zhao, 2019). In this paper, we addressed the importance of the correlation between two competing risks and its impact on the estimation of the ratio of median survival times to an event of interest in a non-PH, AFT model. The BVN linear model is a parametric AFT for simulation and estimation with multivariate and competing risks. The parametric specification as a bivariate normal distribution together with likelihood penalization better regularizes an otherwise ill-posed problem, as demonstrated in simulations. This stabilization of parameter estimation, together with the model’s ability to estimate marginal hazards and directly introduce correlation, are strengths of the strong parametric assumption.

Employing our package `bnc`, we fit a BVN competing risks model using an EM algorithm and a mildly penalized likelihood. In the presence of competing events, estimation of the marginal hazard of a specified event is feasible. For example, the novel application of the EM algorithm provides the reliable estimation of the MR in two-sample comparisons. The method is valid whenever cause-specific hazards of the event of interest and competing event conform to those of a BVN distribution. The existence of latent variables is not required for parameter estimation; the correct likelihood function of observable multistate data bears only a formal relationship to the latent (complete) data model. In particular, the EM algorithm does *not* require an assumption of bivariate normality of (sometimes uninterpretable) latent variables. The complete data construction is purely a device simplifying algorithmic maximization of Prentice’s likelihood.

Furthermore, while not studied here, the EM algorithm generalizes immediately to *semi*-competing risks, where the time to the competing cause (often death) is observed (e.g. Lee et al., 2017). As these authors noted in their Section 4.1, while semi-parametric models are flexible for large sample size, in small-sample settings a parametric specification might be more appealing due to its parsimony and computational simplicity (see also Cole et al., 2021). In semi-competing risks the bivariate normal correlation is fully identifiable, and the EM algorithm immediately applies to provide a useful estimation procedure.

Current limitations of our approach include bivariate outcomes, identity transformation, covariates common to both causes, and the use of an AFT model. The formulation of an EM algorithm for covariates that differ between causes is the subject of current work.

As an alternative to the identity transformation, the EM algorithm also makes available a stable estimation procedure for MPL estimation in the competing risks model of Deresa and Van Keilegom. Given a choice of transformation  $\Lambda_\alpha(Y)$  in that model, our EM algorithm is directly available. Different choices of the parameter  $\alpha$  determining the transformation each provide a marginal Likelihood indicating quality of fit of each choice. Furthermore, log-Likelihood comparisons provide statistical Likelihood Ratio assessment of improved fit of transformed versus untransformed BVN models. An important question for later study is the effect of the chosen power transform on estimates.

In two-sample AFT comparisons, BVN regression coefficients provide estimates of the marginal treatment-benefit of interest,  $\Delta$ . We found that this benefit is estimated accurately, even in small sample sizes ( $n = 100$ ), avoiding the bias apparent in other approaches.

Alternative Cox models are based on proportional hazards assumptions, which differ from those of AFT models. Moreover, with a positive correlation between competing risks our results on estimating marginal treatment-benefit show that Cox models are not robust to departures from  $\rho = 0$ . Thus the BVN model is a useful development for application when strong correlation of the survival outcome with a competing risk is suspected, perhaps as a consequence of unmeasured covariates.

The BVN linear model assumes that the time to event distribution is lognormal. While the

parametric form of hazards is difficult to identify from data, simulations with non-normal data suggest the model is robust. Our simulation findings concerning lack of bias in regression parameter estimation are in accord with those of a different approach, which induces correlation between latent event times through a copula model (Chen, 2010). However, when the degree of association is mis-specified in the copula, regression parameters estimates are severely biased. This indicates an advantage of the use of the BVN model, which provided reliable estimates by adapting to the level of association in the data.

Our simulations show the estimates of regression coefficients are reliable. These regression coefficients provide means and mean differences (relative differences when survival times are log-transformed). While the correlation estimation is unstable (even in sample size  $n = 1000$ ) it exhibits little bias; it may find use in validating external information on  $\rho$ .

The parametric model can be readily used as a sensitivity analysis for assessing the effects of correlation induced by neglected covariates. This method can introduce external information on the association of competing risks to assess its influence on other parameter estimates. Because of the uncertain ability to estimate  $\rho$ , our simulations considered an alternative to MPL estimates of treatment effect in the BVN model, fixing  $\rho$  (e.g. assuming  $\rho = 0$ ) to provide restricted estimates of other BVN model parameters. Asymptotic properties of this GEE-like procedure warrant investigation.

In O'Brien's study, the fitted cumulative incidences of the BVN model were similar to the Nelson-Aalen non-parametric estimates. This supports the use of a BVN model as consistent with the observable data. The evidence of a benefit of adjuvant radiotherapy treatment in survival to local-recurrence was judged a little weaker by BVN fits than by Cox models, though point estimates of the HR were nearly constant and similar to those of the Cox model.

Regression estimates of treatment effect may differ from corresponding Cox coefficients due to different models (accelerated failure time versus proportional hazard). When dependent censoring is present, and fitted parametric cumulative incidence functions agree with semi-parametric estimates, the BVN model can prove useful in estimating treatment effect for comparison with findings of alternative PH model fits.

**Acknowledgement:** The authors wish to thank the reviewers for their careful reading and constructive comments.

**Competing interests:** The authors declare they have no conflict of interest.

## References

- M. Aitkin. A note on the regression analysis of censored data. *Technometrics*, 23(2):161–163, 1981. ISSN 00401706. URL <http://www.jstor.org/stable/1268032>.
- J. F. Bobb and R. Varadhan. *turboEM: A Suite of Convergence Acceleration Schemes for EM, MM and Other Fixed-Point Algorithms*, 2018. URL <https://CRAN.R-project.org/package=turboEM>. R package version 2018.1.
- B. E. Chen, J. L. Kramer, M. H. Greene, and P. S. Rosenberg. Competing risks analysis of correlated failure time data. *Biometrics*, 64(1):172–179, 2008.
- Y.-H. Chen. Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *Journal of the Royal Statistical Society, Series B (Methodological)*, 72:235–251, 2010. ISSN 1369-7412.
- S. R. Cole, H. Chu, and S. Greenland. Maximum Likelihood, Profile Likelihood, and Penalized Likelihood: A Primer. *American Journal of Epidemiology*, 179(2):252–260, 10 2013. ISSN 0002-9262. . URL <https://doi.org/10.1093/aje/kwt245>.
- Stephen R Cole, Jessie K Edwards, Alexander Breskin, and Michael G Hudgens. Comparing Parametric, Nonparametric, and Semiparametric Estimators: The Weibull Trials. *American Journal of Epidemiology*, 190(8):1643–1651, 02 2021. ISSN 0002-9262. . URL <https://doi.org/10.1093/aje/kwab024>.
- D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, 34: 187–220, 1972. ISSN 0035-9246.
- A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977. ISSN 00359246. .
- N. Deresa and I. Van Keilegom. A multivariate normal regression model for survival data subject to different types of dependent censoring. *Computational Statistics & Data Analysis*, 144, 2020a. ISSN 0167-9473.
- N. W. Deresa and I. Van Keilegom. Flexible parametric model for survival data subject to dependent censoring. *Biometrical Journal*, 62(1):136–156, 2020b. ISSN 0323-3847.
- J. J. Dignam, Q. Zhang, and M. N. Kocherginsky. The use and interpretation of competing risks regression models. *Clinical Cancer Research*, 18(8):2301–2308, 2012.
- T. Emura and Y. H. Chen. Gene selection for survival data under dependent censoring: a copula-based approach. *Statistical Methods in Medical Research*, 25(6):2840–2857, 2016. . URL <https://doi.org/10.1177/0962280219892295>.
- T. Emura and Y.-H. Chen. *Analysis of Survival Data with Dependent Censoring. Copula-Based Approaches*. Springer Briefs in Statistics. Springer, 1 edition, 2018. .
- T. Emura, J. H. Shih, I. D. Ha, and R. A. Wilke. Comparison of the marginal hazard model and the sub-distribution hazard model for competing risks under an assumed copula. *Statistical Methods in Medical Research*, 29(8):2307–2327, 2020. . URL <https://doi.org/10.1177/0962280219892295>. PMID: 31868107.
- G. Escarela and J. F. Carriere. Fitting competing risks with an assumed copula. *Statistical Methods in Medical Research*, 12(4):333–349, 2003.
- J. P. Fine and R. J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.
- P. J. Green. On use of the EM algorithm for penalized likelihood estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):443–452, 1990. ISSN 00359246. URL JSTOR, [www.jstor.org/stable/2345668](http://www.jstor.org/stable/2345668).
- I. D. Ha, N. J. Christian, J.-H. Jeong, J. Park, and Y. Lee. Analysis of clustered competing risks data using subdistribution hazard models with multivariate frailties. *Statistical Methods in Medical Research*, 25:2488–2505, Dec 2016.
- P Hougaard. Fundamentals of survival data. *Biometrics*, 55(1):13, 1999. ISSN 0006-341X.
- H. M. Hudson, M. Manuguerra, and V. Gares. R package bnc: bivariate normal censored linear model for competing risks, 2019. URL [https://bitbucket.org/malcolm\\_hudson/bnc/src/master/](https://bitbucket.org/malcolm_hudson/bnc/src/master/). [Online; accessed 19-December-2019].

- H.M. Hudson, V. Gares, M. Manuguerra, and V. Gebski. Correlated bivariate normal competing risks – simulation findings in an ill-posed problem. ISCB ASC Melbourne 2018, presentation (slides) available on request, 2018.
- M. Jamshidian and R. I. Jennrich. Standard errors for EM estimation. *Journal of the Royal Statistical Society, Series B.*, 62:257–270, 2000.
- J.D. Kalbfleisch and R.L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley, New York, 2002.
- N. Keiding, P. K. Andersen, and J. P. Klein. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, 16(2):215–224, 1997. . URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0258%2819970130%2916%3A2%3C215%3A%3AAID-SIM481%3E3.0.CO%3B2-J>.
- J. P. Klein, C. Pelz, and M.-J. Zhang. Modeling random effects for censored data by a multivariate normal regression model. *Biometrics*, 55(2):497–506, 1999. ISSN 0006-341X.
- E. L. Korn, J. J. Dignam, and B. Freidlin. Assessing treatment benefit with competing risks not affected by the randomized treatment. *Statistics in Medicine*, 34(2):265–280, 2015. . URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.6353>.
- P. Lambert, D. Collett, A. Kimber, and R. Johnson. Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine*, 23(20): 3177–3192, 2004. ISSN 0277-6715.
- K. H. Lee, V. Rondeau, and S. Haneuse. Accelerated failure time models for semi-competing risks data in the presence of complex censoring. *Biometrics*, 73(4):1401–1412, 2017. . URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12696>.
- Z. J. I. Lu, J. Sykes, and M. Pintilie. Understanding dependent competing risks: A simulation study to illustrate the relationship between cause-specific hazard and marginal hazard. *Journal of Statistical Science and Application*, 4(03-04):96–107, April 2016. .
- F. Mosteller and J.W. Tukey. *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley, 1977.
- S. K. Ng and G. J. McLachlan. An em-based semi-parametric mixture model approach to the regression analysis of competing-risks data. *Statist. Med.*, 22(7):1097–1111, 2003. ISSN 1097-0258. .
- C. J. O’Brien, K. Petersen-Schaefer, et al. Adjuvant radiotherapy following neck dissection and parotidectomy for metastatic malignant melanoma. *Head and Neck*, 19(7):589–594, 1997.
- R. L. Prentice and S. Zhao. *The Statistical Analysis of Multivariate Failure Time Data: A Marginal Modeling Approach*. Number 163 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, June 2019. ISBN 978-1-482-25657-4.
- N. Reid. A conversation with Sir David Cox. *Statistical Science*, 9:439–455, 1994.
- B.-C. Tai, B. L. De Stavola, V. de Gruttola, V. Gebski, and D. Machin. First-event or marginal estimation of cause-specific hazards for analysing correlated multivariate failure-time data? *Statistics in Medicine*, 27(6):922–936, 2008. ISSN 1097-0258. .
- A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America*, 72(1):20, 1975. ISSN 0027-8424. URL <http://www.pnas.org/content/72/1/20.abstract>.
- D. A. Van Dyk, X.-L. Meng, and D. B. Rubin. Maximum likelihood estimation via the ECM algorithm: Computing the asymptotic variance. *Statistica Sinica*, 5(1):55–75, 1995. ISSN 10170405, 19968507. URL <http://www.jstor.org/stable/24305553>.
- Zhiyong Zhang. A note on Wishart and inverse Wishart priors for covariance matrix. *Journal of Behavioral Data Science*, 1(2):119–126, Aug. 2021. . URL <https://jbds.isdsa.org/jbds/article/view/19>.



## Appendix A. Conditional expectations of the standard BVN

The different conditional expectations used in the E-step of EM algorithm are given below.

Let  $\Phi(z)$  denote the survival function (i.e. *upper-tail* cumulative distribution) of the univariate standard normal distribution. Its density is given by  $\phi(z) = \exp(-z^2)/\sqrt{2\pi}$ , for  $-\infty < z < \infty$ , and its hazard function  $\Psi$  is  $\Psi(z) = \phi(z)/\Phi(z)$ .

Some properties of the bivariate Normal variates  $(Z_1, Z_2)$  with mean vector  $\mu = (0, 0)$ , variance matrix  $\Sigma$ , with  $\sigma_{11} = \sigma_{22} = 1$  and  $\sigma_{12} = \rho$  and density  $p_{12}(z_1, z_2)$  are given below. In particular, we provide calculations for conditional expectations used in the E-step of EM algorithms for bivariate normal linear models.

Denote by  $H$  the Heavyside function, the step function taking values 0 or 1 indicating whether its argument exceeds 0. Let  $P(a, b) = \mathbb{P}[Z_1 > a, Z_2 > b] = \mathbb{E}[H(Z_1 - a)H(Z_2 - b)]$ .

- (1) The conditional distribution of  $(Z_1|Z_2 = b)$  is  $\mathcal{N}(\rho b, 1 - \rho^2)$  with density

$$p_{1|2}(z|b) = \frac{1}{\sqrt{1 - \rho^2}} \phi\left(\frac{z - \rho b}{\sqrt{1 - \rho^2}}\right). \quad (\text{A1})$$

- (2) Conditional expectations: moments of survival time to event given observed time to a competing event

$$E_{01}^{10}(a, b) = \mathbb{E}[Z_1|Z_1 > a, Z_2 = b] \quad (\text{A2})$$

$$= \rho b + \sqrt{1 - \rho^2} \Psi\left(\frac{a - \rho b}{\sqrt{1 - \rho^2}}\right) \quad (\text{A3})$$

$$E_{01}^{20}(a, b) = \mathbb{E}[Z_1^2|Z_1 > a, Z_2 = b] \quad (\text{A4})$$

$$= 1 - \rho^2 + \sqrt{1 - \rho^2} a \Psi\left(\frac{a - \rho b}{\sqrt{1 - \rho^2}}\right) + \rho b E_{01}^{10} \quad (\text{A5})$$

or,

$$E_{01}^{20}(a, b) = \mathbb{E}[Z_1^2|Z_1 > a, Z_2 = b] \quad (\text{A6})$$

$$= 1 - \rho^2 + \sqrt{1 - \rho^2} (a + \rho b) \Psi\left(\frac{a - \rho b}{\sqrt{1 - \rho^2}}\right) + \rho^2 b^2$$

- (3) Conditional expectations: moments given lower bounds

$$\begin{aligned}
E_{00}^{11}(a, b) &= \mathbb{E}[Z_1 Z_2 | Z_1 > a, Z_2 > b] \\
&= \rho + (1 - \rho^2) \frac{p_{12}(a, b)}{P(a, b)} \\
&\quad + \rho \frac{a \phi(a) \Phi\left(\frac{b - \rho a}{\sqrt{1 - \rho^2}}\right) + b \phi(b) \Phi\left(\frac{a - \rho b}{\sqrt{1 - \rho^2}}\right)}{P(a, b)} \\
E_{00}^{10}(a, b) &= \mathbb{E}[Z_1 | Z_1 > a, Z_2 > b] \\
&= \frac{\phi(a) \Phi\left(\frac{b - \rho a}{\sqrt{1 - \rho^2}}\right) + \rho \phi(b) \Phi\left(\frac{a - \rho b}{\sqrt{1 - \rho^2}}\right)}{P(a, b)} \\
E_{00}^{20}(a, b) &= \mathbb{E}[Z_1^2 | Z_1 > a, Z_2 > b] \\
&= 1 + \frac{a \phi(a) \Phi\left(\frac{b - \rho a}{\sqrt{1 - \rho^2}}\right) + \rho^2 b \phi(b) \Phi\left(\frac{a - \rho b}{\sqrt{1 - \rho^2}}\right) + \rho(1 - \rho^2) p_{12}(a, b)}{P(a, b)}
\end{aligned} \tag{A7}$$

These identities were developed by use of Stein identities (Liu, 1994). See also Tallis (1961). Proofs are available on request.

## Appendix B. Imputing complete data sufficient statistics

Consider the E-step with bivariate normal correlated latent response random variables  $(Y_1, Y_2)$  providing  $n$  independent observations (row pairs) of the complete data forming the random matrix  $\mathbf{Y} = (y_{jk})$  of dimension  $n \times 2$  and covariate random variables  $\mathbf{x} = (X_1, \dots, X_p)'$  with corresponding data matrix  $\mathbf{X}$ ,  $n \times p$ , of observed values. The parameters  $\mathbf{B}$ ,  $\Sigma$  of the bivariate normal regression model are the matrix whose columns are vectors of regression coefficients of each cause,  $\beta_k$ ,  $k = 1, 2$ , and the covariance matrix common to the bivariate normal observations. Specify initial estimates  $\boldsymbol{\theta}^{(0)} = (\mathbf{B}^{(0)}, \Sigma^{(0)})$ .

Let  $\boldsymbol{\theta}^{(i)} = (\mathbf{B}^{(i)}, \Sigma^{(i)})$  be current estimates during iteration  $i$ , then  $\mathbf{M}^{(i)} = \mathbf{X}\mathbf{B}^{(i)}$  is the  $n \times 2$  matrix of corresponding expected values. The E-step imputes *complete data sufficient statistics*  $\mathbf{X}'\mathbf{Y}$ , of dimension  $p \times 2$ , and  $\mathbf{Y}'\mathbf{Y}$ ,  $2 \times 2$ . It evaluates expectations of  $\mathbf{X}'\mathbf{Y}$  and  $\mathbf{Y}'\mathbf{Y}$  at  $\boldsymbol{\theta}^{(i)}$ , conditional on observable (termed *incomplete*) data. Incomplete data here is defined as the observable random vectors  $\mathbf{Y}^o, \mathbf{D}, X_1, \dots, X_p$ , comprising a random sample of  $n$  bivariate observations  $\{Y_j^o, D_j, X_1, \dots, X_p, C_j\}$ , the component variables denoting survival time and event status, as well as covariates and follow-up time, with distribution  $[(Y^o, D) | X_1, \dots, X_p]$  depending on parameters  $\mathbf{B} = (\beta_1, \beta_2)$  and  $\Sigma$ . Follow-up time,  $C_j$ , is assumed independent of all other variables (independent censoring).

Observed data is then a function, by row, of complete data  $\mathbf{Y}$ , with  $\mathbf{Y}^o = \min(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{C})$ , so that, for observation  $j$ ,  $y_j^o = \min(y_{j1}, y_{j2}, c_j)$ , for  $j = 1, \dots, n$ . Here  $c_j$  is log-time to end of follow-up for subject  $j$ . The status  $\mathbf{D} = (D_1, \dots, D_n)$  for  $D_j \in \{0, 1, 2\}$  identifies the observed event (1 or 2), or, when  $y_j^o = c_j$ , assigns  $d_j = 0$ .

Note that, for the M-step optimisation by  $\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ , fitted values are  $\mathbf{X}\widehat{\mathbf{B}} = \mathbf{P}\mathbf{Y}$  and residuals  $\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}} = (\mathbf{I} - \mathbf{P})\mathbf{Y} = \mathbf{Q}\mathbf{Y}$ , where  $\mathbf{P}$  and  $\mathbf{Q}$  are projection matrices. Here  $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  and  $\mathbf{Q} = \mathbf{I} - \mathbf{P}$ .

In the E-step, conditional expectations are evaluated for specified parameter values  $\boldsymbol{\theta}^{(i)}$ .

To reduce complexity in equations, we will use the notation  $\mathbb{E}^{\text{obs},(i)}$  for expectations at  $\boldsymbol{\theta}^{(i)}$  conditional on  $\mathbf{Y}^o, \mathbf{D}, \mathbf{X}, \mathbf{C}$ . Again, to simplify presentation, while conditioning on  $\mathbf{Y}^o, \mathbf{D}, \boldsymbol{\theta}^{(i)}$  will be explicit, we then implicitly assume conditioning also on  $\mathbf{X}$  and  $\mathbf{C}$ , treating them as known. Current estimates  $\boldsymbol{\theta}^{(i)}$  are used to provide expected survival times  $\mathbf{M}^{(i)}$  and imputations  $\mathbf{P}\mathbf{Y}^{(i)}$  and  $\mathbb{E}^{\text{obs},(i)}[\mathbf{Y}'\mathbf{Y}]$  of complete data sufficient statistics  $\mathbf{X}'\mathbf{Y}$  and  $\mathbf{Y}'\mathbf{Y}$ . Here

$$\mathbf{Y}^{(i)} = \mathbb{E}[\mathbf{Y} | \mathbf{Y}^o, \mathbf{D}, \boldsymbol{\theta}^{(i)}]. \quad (\text{B1})$$

The imputed sufficient statistics for the E-step in iteration  $i$  are:

$$\mathbb{E}^{\text{obs},(i)}[\mathbf{X}'\mathbf{Y}] = \mathbf{X}' \mathbb{E}[\mathbf{Y} | \mathbf{Y}^o, \mathbf{D}, \boldsymbol{\theta}^{(i)}] = \mathbf{P}\mathbf{Y}^{(i)}, \quad (\text{B2})$$

and

$$\mathbb{E}^{\text{obs},(i)}[\mathbf{Y}'\mathbf{Y}] = \mathbb{E}[\mathbf{Y}'\mathbf{Y} | \mathbf{Y}^o, \mathbf{D}, \boldsymbol{\theta}^{(i)}]. \quad (\text{B3})$$

with the imputed complete data calculated from equation (B1). Computing  $\mathbf{Y}^{(i)}$  requires standardising the bivariate normal distribution of  $(Y_1, Y_2)$  in order to use results of A.

Define standardised variables  $(Z_1, Z_2)$  as  $Z_k = (Y_k - \mathbf{x}'\boldsymbol{\beta}_k^{(i)})/\sigma_k^{(i)}$ , for  $k = 1, 2$ . Then the pair  $(Z_1, Z_2)$  is standard bivariate normal, with means 0, standard deviations 1 and correlation  $\rho^{(i)}$  of  $\boldsymbol{\Sigma}^{(i)}$ . The evaluation of imputed complete data for parameters set for the current iteration  $(i)$ ,  $\mathbf{Y}^{(i)}$  utilises the formulas developed in Appendix A, noting that

$$\mathbb{E}[Y_{jk} | \mathbf{Y}^o = \mathbf{y}^o, \mathbf{D} = \mathbf{d}; \boldsymbol{\theta}^{(i)}] = m_{jk}^{(i)} + \sigma_k^{(i)} \mathbb{E}[Z_{jk} | Y_j^o = y_j^o, D_j = d_j] \quad (\text{B4})$$

where

$$\mathbb{E}[Z_{jk} | Y_j^o = y_j^o, D_j = d_j] = \begin{cases} \mathbb{E}[Z_{jk} | Z_{j1} = z_{j1}, Z_{j2} > z_{j2}] & \text{if } d_j = 1, \\ \mathbb{E}[Z_{jk} | Z_{j1} > z_{j1}, Z_{j2} = z_{j2}] & \text{if } d_j = 2, \\ \mathbb{E}[Z_{jk} | Z_{j1} > z_{j1}, Z_{j2} > z_{j2}] & \text{if } d_j = 0. \end{cases} \quad (\text{B5})$$

Here  $z_{jk} = (y_j^o - m_{jk})/\sigma_k$ , for  $k = 1, 2$ , where calculations in (B4) and (B5) utilise the expectations and covariance matrix  $(\mathbf{M}^{(i)}, \boldsymbol{\Sigma}^{(i)})$  of the current iteration.

In this case, the unconditional expectations of sufficient statistics  $\mathbf{X}'\mathbf{Y}$  and  $\mathbf{Y}'\mathbf{Y}$  are  $\mathbf{B}$  and  $\mathbf{M}'\mathbf{M} + n\boldsymbol{\Sigma}$ . Hence the M-step sets parameters  $\mathbf{B}$  and  $\boldsymbol{\Sigma}$  to the new E-step conditional expectations (*imputations*). The new estimates comprising  $\boldsymbol{\theta}^{(i+1)}$  are  $\mathbf{B}^{(i+1)}$  and  $\boldsymbol{\Sigma}^{(i+1)}$  where:

$$(\mathbf{X}'\mathbf{X})\mathbf{B}^{(i+1)} \leftarrow \mathbf{X}' \mathbb{E}^{\text{obs},(i)}[\mathbf{Y}] = \mathbf{X}'\mathbf{Y}^{(i)},$$

so that

$$\begin{aligned} \mathbf{B}^{(i+1)} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}^{(i)}, \\ \mathbf{M} &= \mathbf{M}^{(i+1)} \leftarrow \mathbf{X}\mathbf{B}^{(i+1)} = \mathbf{P}\mathbf{Y}^{(i)} \end{aligned} \quad (\text{B6})$$

and, for ML estimation,

$$\begin{aligned}
n\Sigma^{(i+1)} &= \mathbb{E}^{\text{obs},(i)} [\mathbf{Y}'\mathbf{Y}] - \mathbf{M}'\mathbf{M} \\
&= \mathbb{E}^{\text{obs},(i)} [\mathbf{Y}'\mathbf{Y}] - \mathbf{Y}^{(i)'} \mathbf{P} \mathbf{Y}^{(i)} \\
&= \mathbb{E}^{\text{obs},(i)} [\mathbf{Y}'\mathbf{Y}] - \mathbf{Y}^{(i)'} \mathbf{Y}^{(i)} + \mathbf{Y}^{(i)'} \mathbf{Q} \mathbf{Y}^{(i)}. \tag{B7}
\end{aligned}$$

For penalised likelihood, the mode of an inverse Wishart posterior for  $\Sigma$  provides adjusted degrees of freedom, substituting  $n + \nu + 3$  in place of  $n$  in equation (B7). The M-step then entails a shrinkage of the unbiased estimator using complete data of  $\Sigma$ .

In particular, from equation (B7) the M-step estimate of the covariance matrix  $\Sigma$  requires imputation of the sufficient statistic  $\mathbf{Y}'\mathbf{Y}$ . This requires, besides  $\mathbf{Y}^{(i)}$ , the evaluation of conditional expectations of additional quadratic and cross-product terms,  $\mathbb{E}^{\text{obs}}[Y_{jk} Y_{jk'}] = \mathbb{E}[Y_{jk} Y_{jk'} | y^o, D]$  for  $k, k' \in \{1, 2\}$ . These conditional expectations are again evaluated, after standardising for means specified in row  $j$  of  $\mathbf{M}^{(i)}$  and variance matrix  $\Sigma^{(i)}$ , using formulas from Appendix A. Let  $W^{(i)}$  be the diagonal matrix with elements  $\sigma_1^{(i)}, \sigma_2^{(i)}$ . Then

$$(\mathbf{Y} - \mathbf{M}^{(i)}) \leftarrow \mathbb{E}[\mathbf{Z} | \mathbf{Y}^o, \mathbf{D}, \boldsymbol{\theta}^{(i)}] W^{(i)} \tag{B8}$$

and

$$\mathbb{E}^{\text{obs},(i)} [\mathbf{Y}'\mathbf{Y}] \leftarrow W^{(i)} \mathbb{E}[\mathbf{Z}'\mathbf{Z} | \mathbf{Y}^o, \mathbf{D}, \boldsymbol{\theta}^{(i)}] W^{(i)}. \tag{B9}$$

The required imputations for row  $j$  are those of equations similar to those of equation (B4), evaluated using Appendix A.

## Appendix C. Supplementary data

This appendix provides a supplement to the paper submitted to Communications in Statistics - Simulation and Computation.

Supplementary material related to this article will be available online. The supplementary file contains Tables and Figures.

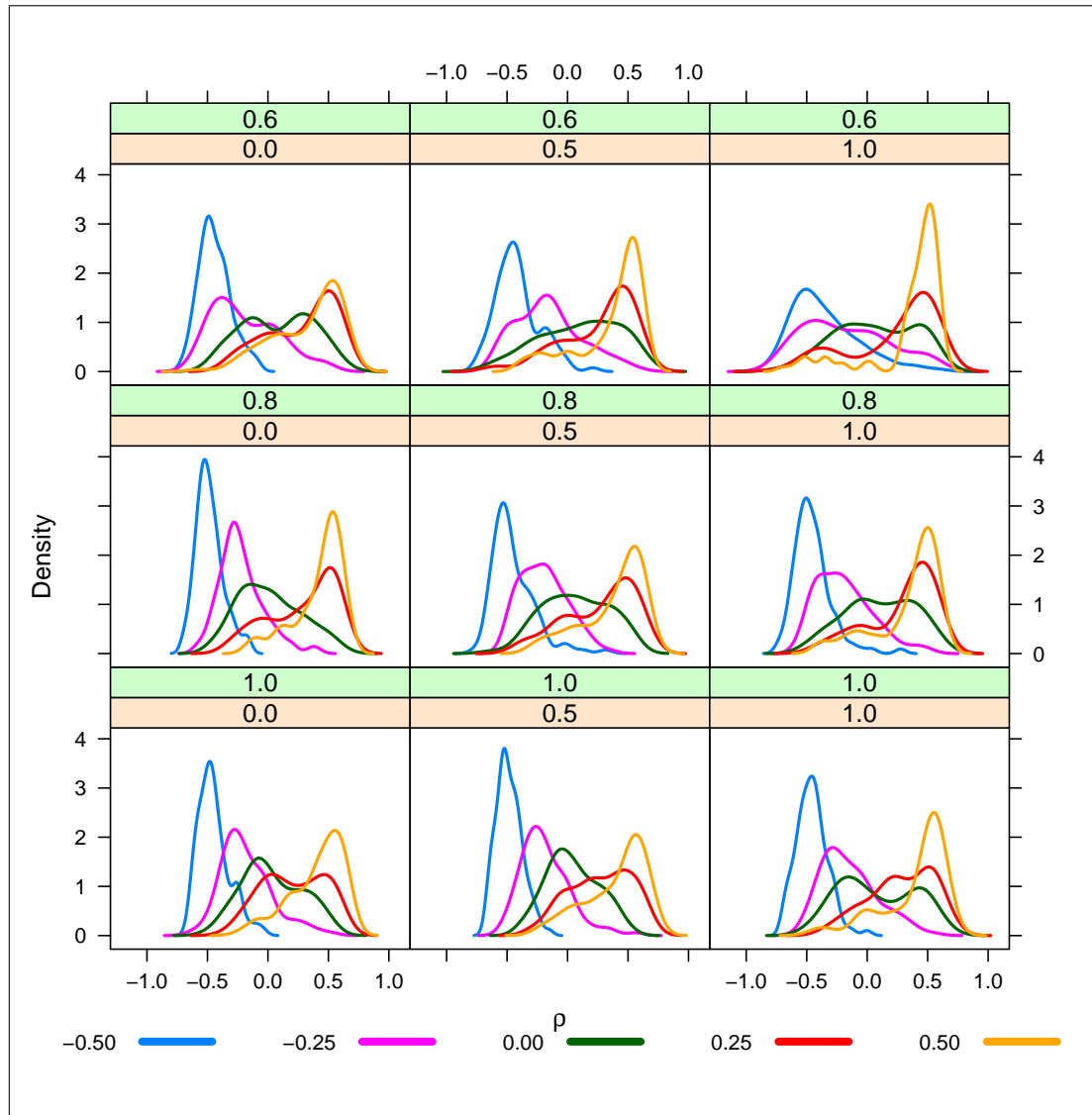
For the use of referees, it is reproduced below.

### C.1. Estimation of correlation in BVN data

For 100 simulated samples of size  $n = 1000$  from a BVN model Table 1 provides information on convergence of the EM algorithm and Figure 1 density plots of correlation estimates it provided.

$P_1(\tau)$	$\rho \mid \mu_2$	0.0	0.5	1.0
1.0	-0.50	32	29	44
	-0.25	62	104	106
	0.00	67	139	205
	0.25	69	152	145
	0.50	75	138	268
0.8	-0.50	30	61	65
	-0.25	60	104	103
	0.00	92	158	148
	0.25	90	139	237
	0.50	123	153	266
0.6	-0.50	36	104	130
	-0.25	96	127	162
	0.00	112	110	294
	0.25	131	175	426
	0.50	139	224	396

**Supplementary Table 1.** Maximum number of squareEM iterations to converge, for sample size  $n = 1000$ . Iterations were stopped after 500 iterations. The parameter  $\mu_k$  is the mean of  $Y_k$ , mean log-time to event  $k$ , for  $k = 1, 2$ . True parameters are:  $\mu_1 = 0, \sigma_1 = \sigma_2 = 1$ .  $P_1(\tau)$  denotes the pre-fixed (marginal) cumulative probability that event 1 occurs before end-of-follow up. This is the proportion of data free of independent censoring. The simulation comprised  $N_{\text{sim}} = 100$  repeat samples.



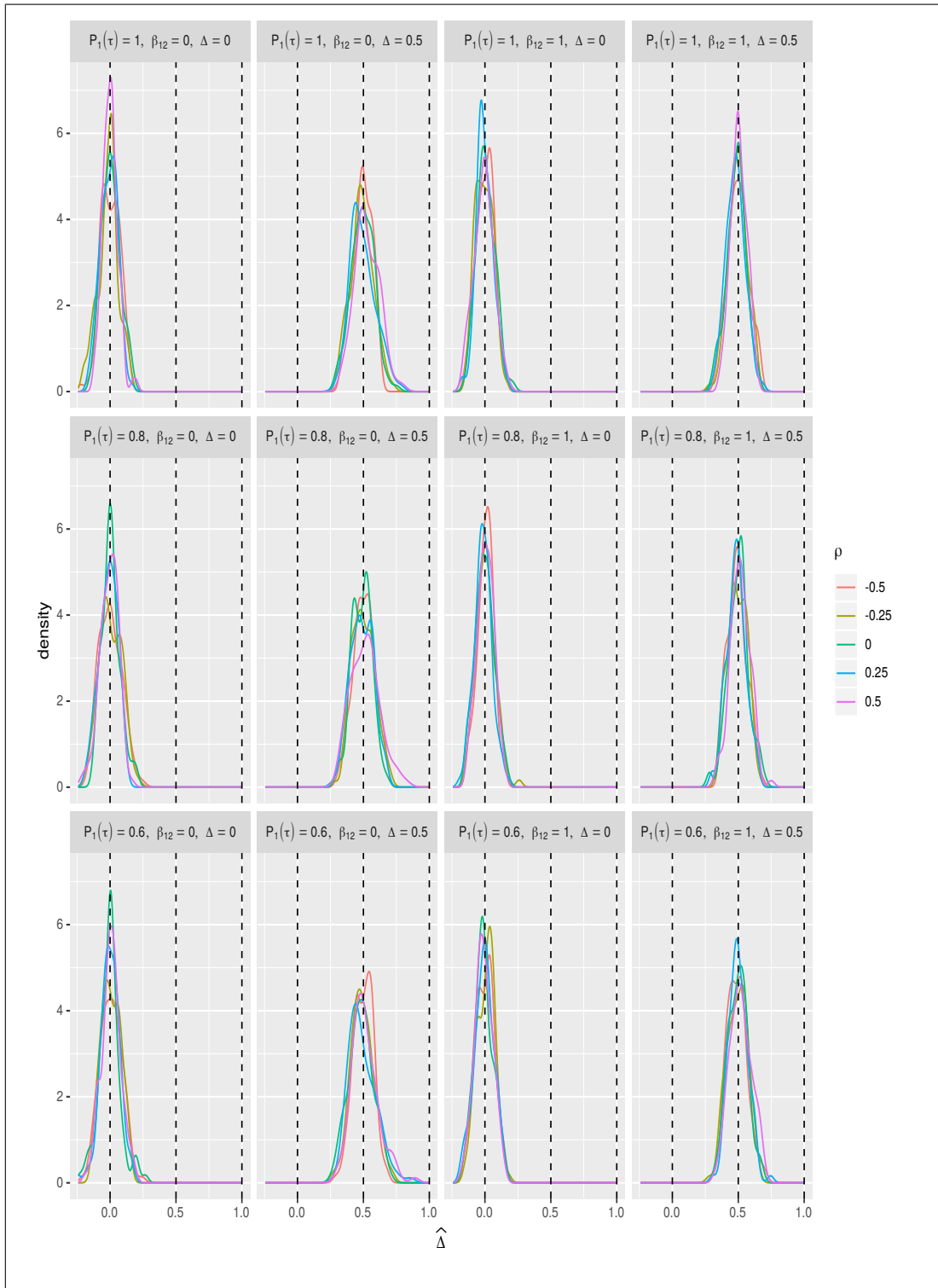
**Supplementary Figure 1.:** Density plots of correlation estimates  $\hat{\rho}$  in samples of size  $n = 1000$ . Varying BNC parameters:  $P_1(\tau) \in (0.6, 0.8, 1)$  (rows);  $\mu_2 \in (0, 0.5, 1)$  (cols). Fixed BNC parameters:  $\mu_1 = 0$ ,  $\sigma_1 = \sigma_2 = 1$ ,  $n = 1000$ . The simulation comprised  $N_{\text{sim}} = 100$  samples.

### C.2. Two-sample BVN data

We provide supplementary simulation information on bias in MPL estimates of two-sample comparison of treatment effect in Table 2, with density plots of estimates in Figure 2. Refer to Section 3.3 of the paper for further details.

$P_1(\tau)$	$\beta_{12}$ $\rho \Delta$	0		1	
		0	0.5	0	0.5
1	-0.5	-0.00	0.50	0.00	0.51
	-0.25	-0.00	0.48	-0.01	0.50
	0	0.00	0.46	0.00	0.50
	0.25	-0.01	0.49	0.00	0.49
	0.5	-0.00	0.52	-0.00	0.49
0.8	-0.5	-0.00	0.50	0.00	0.50
	-0.25	-0.00	0.48	0.01	0.49
	0	0.01	0.49	0.00	0.49
	0.25	-0.00	0.48	-0.01	0.48
	0.5	-0.01	0.51	0.01	0.51
0.6	-0.5	0.02	0.49	-0.01	0.48
	-0.25	-0.02	0.52	-0.00	0.48
	0	-0.01	0.49	0.01	0.48
	0.25	0.02	0.49	0.01	0.50
	0.5	0.00	0.51	-0.01	0.51

**Supplementary Table 2.** Medians of MPL estimates  $\widehat{\Delta}$  of treatment benefit in the BNC model fitted using the `squareEM` algorithm. True parameters are :  $\mu_1^C = 0$ , treatment benefit  $\Delta \in \{0, 0.5\}$ , event 2 delay  $\beta_{12} \in \{0, 1\}$ ,  $\sigma_1 = \sigma_2 = 1$ ,  $n = 1000$ . The simulation comprised  $N_{\text{sim}} = 100$  samples.



**Supplementary Figure 2.:** Density plots of the MPL estimator  $\hat{\Delta}$ ,  $n = 1000$ . BNC model parameters varying:  $P_1(\tau)$ , delay  $\beta_{12}$ ,  $\Delta$ , with  $\mu_1^C = 0$ ,  $\beta_{22} = 0$  and  $\sigma_1 = \sigma_2 = 1$ . The simulation comprised  $N_{\text{sim}} = 100$  samples.

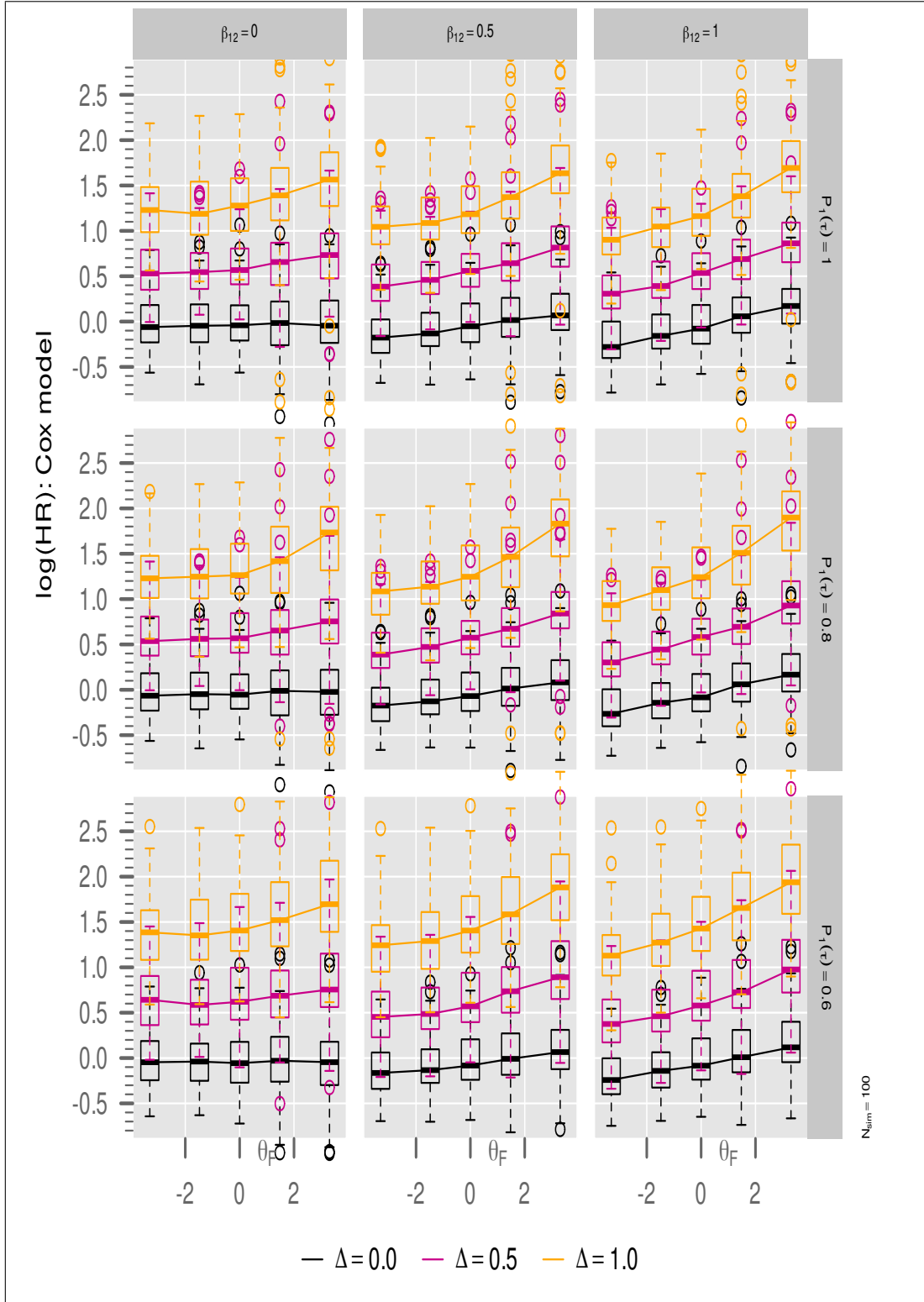


### ***C.3. 2-sample comparisons in non-BVN correlated data***

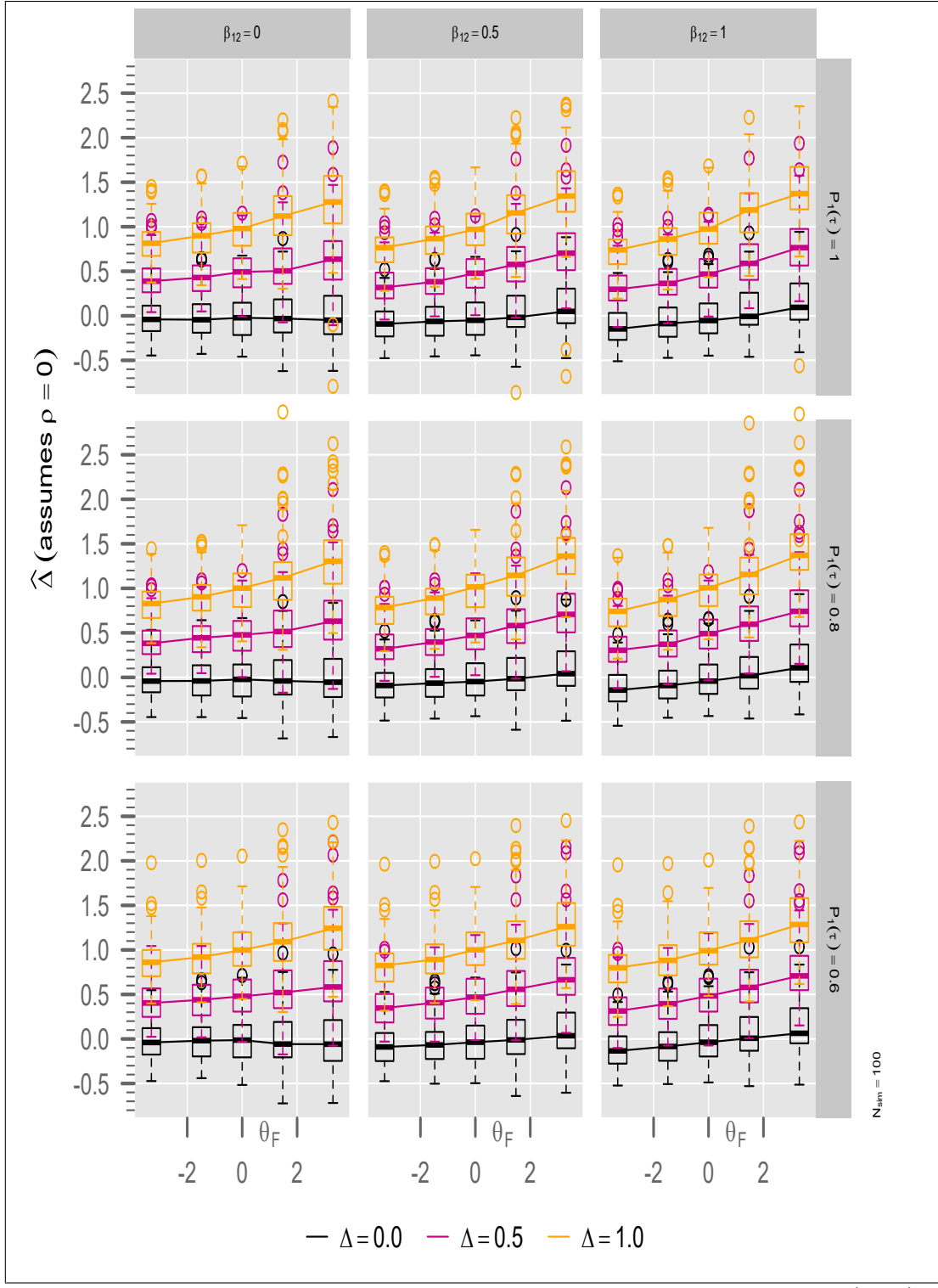
Robustness of model estimates of treatment benefits in two-sample randomised comparisons are assessed. Marginal distributions are normal, with a copula model producing correlation between causes. The three models are:

- Cox model (Figure 3);
- restricted estimation under the assumption  $\rho = 0$  (Figure 4);
- unconstrained MPL estimation (Figure 5).

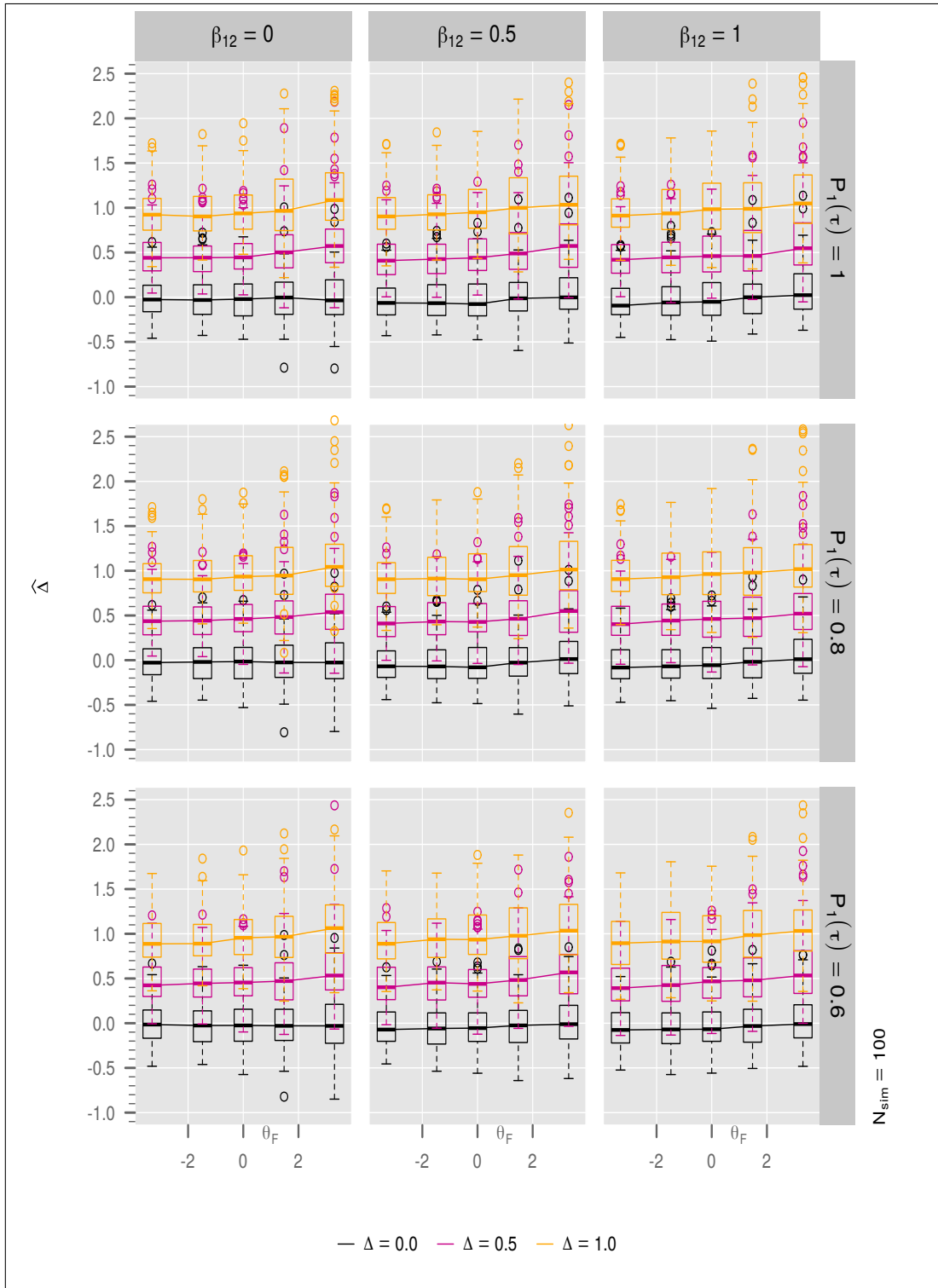
Refer to Section 3.5 of the paper for further details and findings.



**Supplementary Figure 3.: Cox model.** Boxplots of simulation estimates of the  $\beta$ -coefficient, i.e. negative log hazard ratio ( $-\log \text{HR}$ ) of the Cox model,  $n = 100$ , for different value of  $\Delta \in \{0, 0.5, 1\}$ . BNC model parameters are :  $P_1(\tau)$  (rows), mean difference  $\mu_2$  (columns), Frank's association  $\theta_F$  (x-axis),  $\mu_1 = 0, \sigma_1 = \sigma_2 = 1$ . The simulation comprised  $N_{\text{sim}} = 100$  samples.

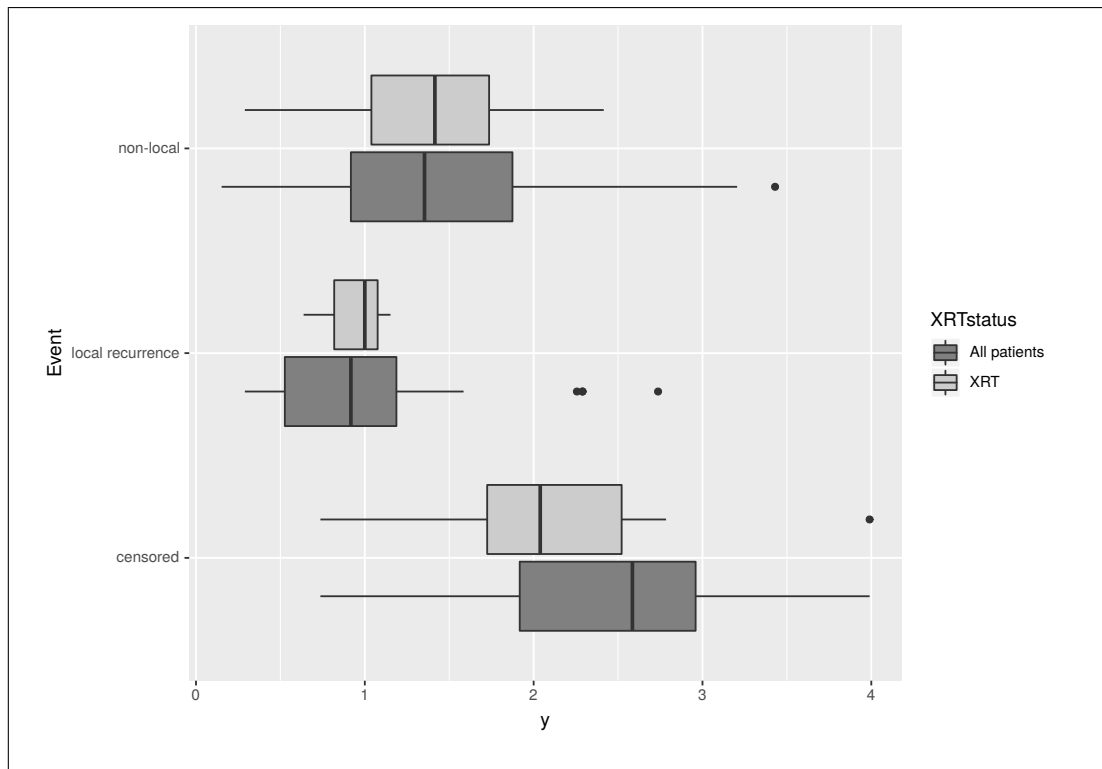


**Supplementary Figure 4.: Restricted MPL.** Boxplots of restricted MPL estimates  $\widehat{\Delta} = \widehat{\beta}_{21}$  for simulated non-Normal data and sample size  $n = 100$ . Estimation was for fixed  $\rho = 0$  using R package `survreg`. BNC model parameters are :  $P_1(\tau)$  (rows),  $\Delta \in \{0, 0.5, 1\}$ , mean difference  $\beta_{12}$  (columns), Frank's association  $\theta_F$  (x-axis),  $\mu_1^C = 0, \sigma_1 = \sigma_2 = 1$ . The simulation comprised  $N_{\text{sim}} = 100$  samples.



**Supplementary Figure 5.: Unrestricted MPL.** Boxplot of simulation estimates of Treatment benefit with non-Normal data (2-sample comparison of groups of size  $n = 100$ ). Varying  $P_1(\tau)$  (rows), mean difference  $\beta_{12}$  (columns),  $\Delta \in \{0, 0.5, 1\}$  and Frank's association  $\theta_F \in \{-3.31, -1.48, 0, 1.48, 3.31\}$  (x-axis). Each boxplot summarises  $N_{\text{sim}} = 100$  sample estimates.

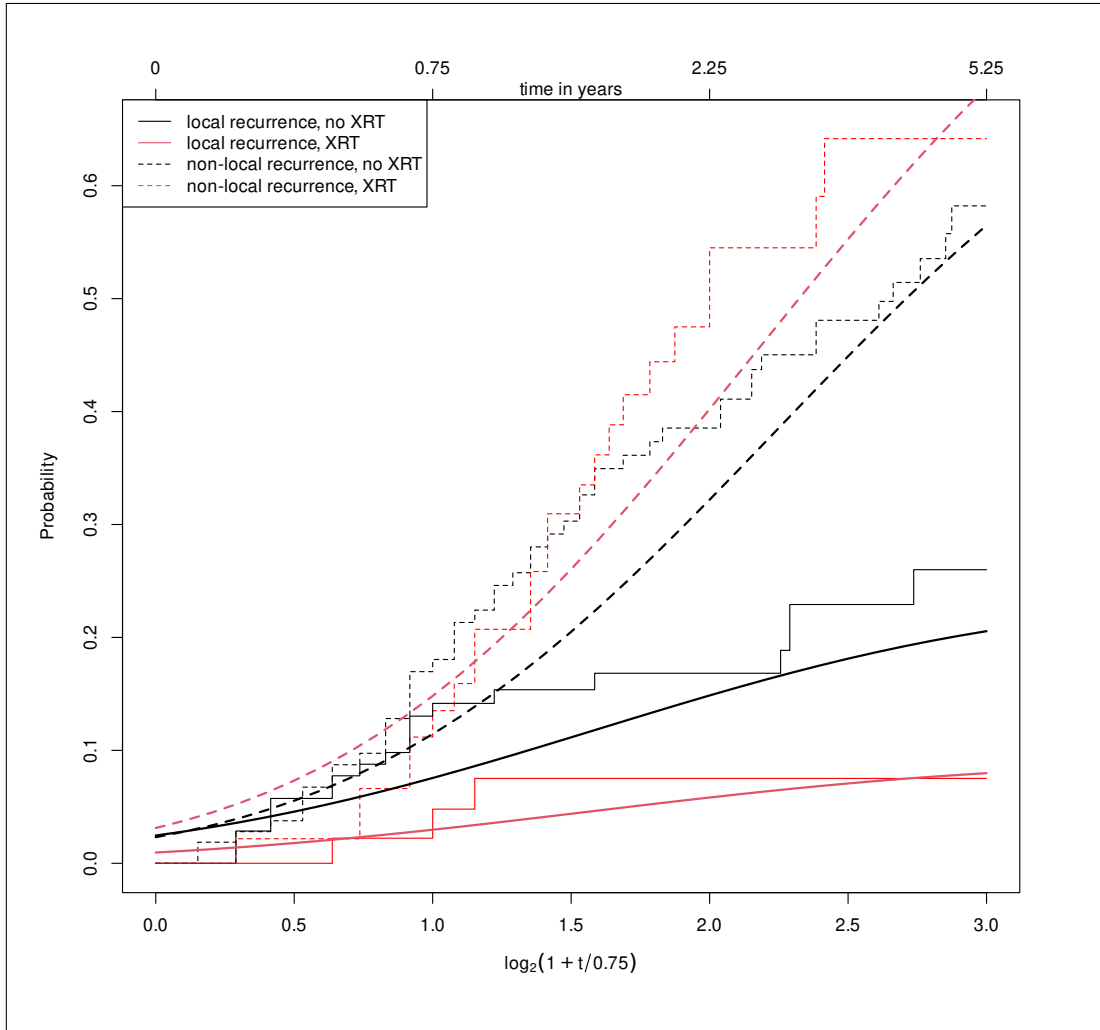
#### C.4. Head and Neck Cancer Example



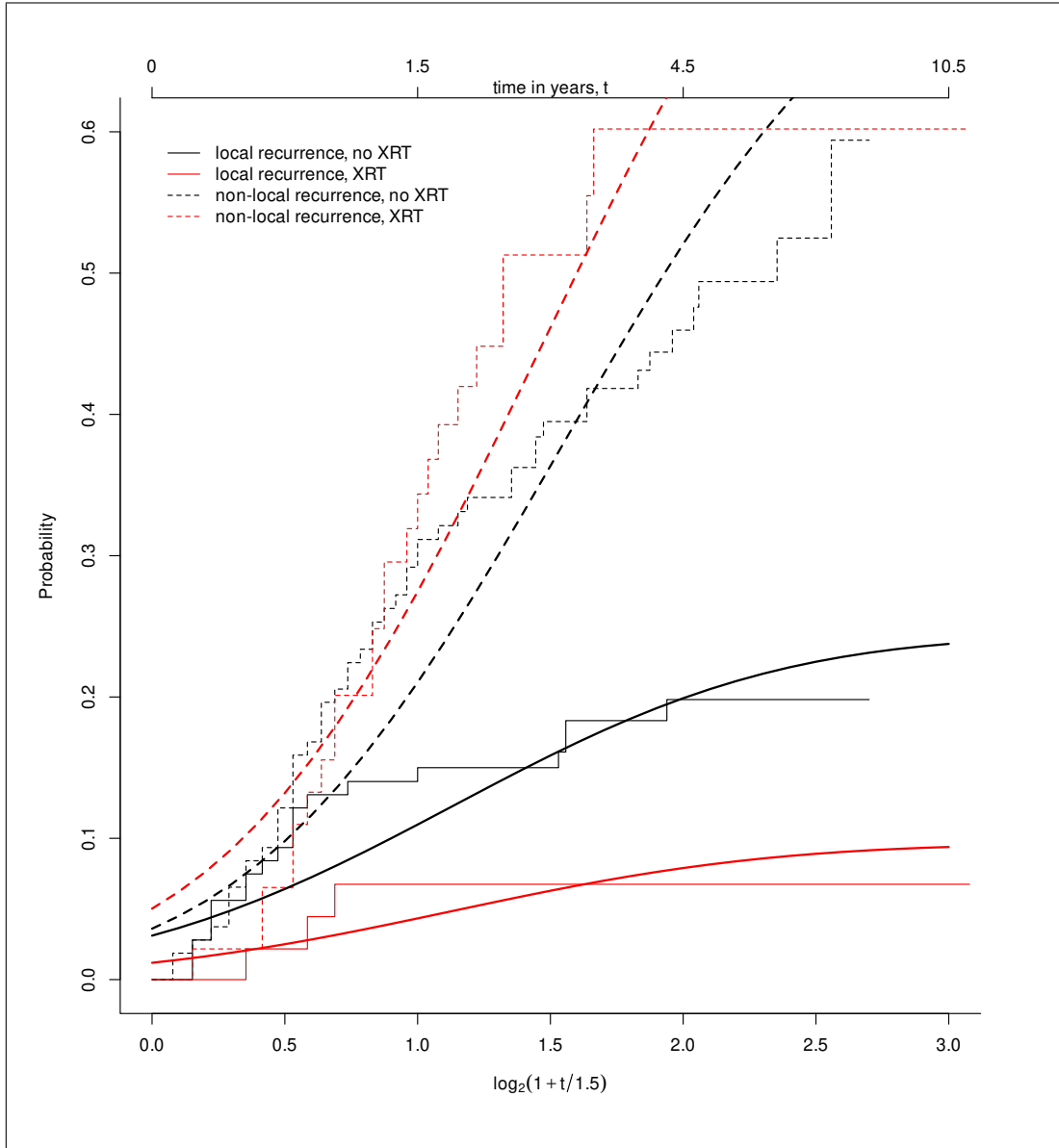
**Supplementary Figure 6.:** Data distributions of times to local and distant recurrence by treatment group and outcome status. Transformation  $y : t \rightarrow \log_2(1 + t/0.75)$  of original time scale. See vignette **HNexample.Rmd** for code.

Parameter	Estimates <sup>a</sup>	Estimates <sup>b</sup>	Estimates <sup>c</sup>
$\beta_{11}$	2.633	1.967	3.076
$\Delta$	0.370	0.322	0.401
$\beta_{12}$	2.170	1.531	2.600
$\beta_{22}$	-0.026	-0.062	0.001
$\sigma_1$	1.437	1.138	1.598
$\sigma_2$	1.198	0.926	1.346
$\rho$	0.897	0.834	0.917

**Supplementary Table 3.** Estimates of BNC linear model parameters in the Head and Neck data using `squareEM` algorithm. Transformations are  $y^a : t \rightarrow \log_2(1 + t/0.75)$ ,  $y^b : t \rightarrow \log_2(1 + t/1.5)$ , and  $y^c : t \rightarrow \log_2(1 + t/0.5)$  in this sensitivity analysis.

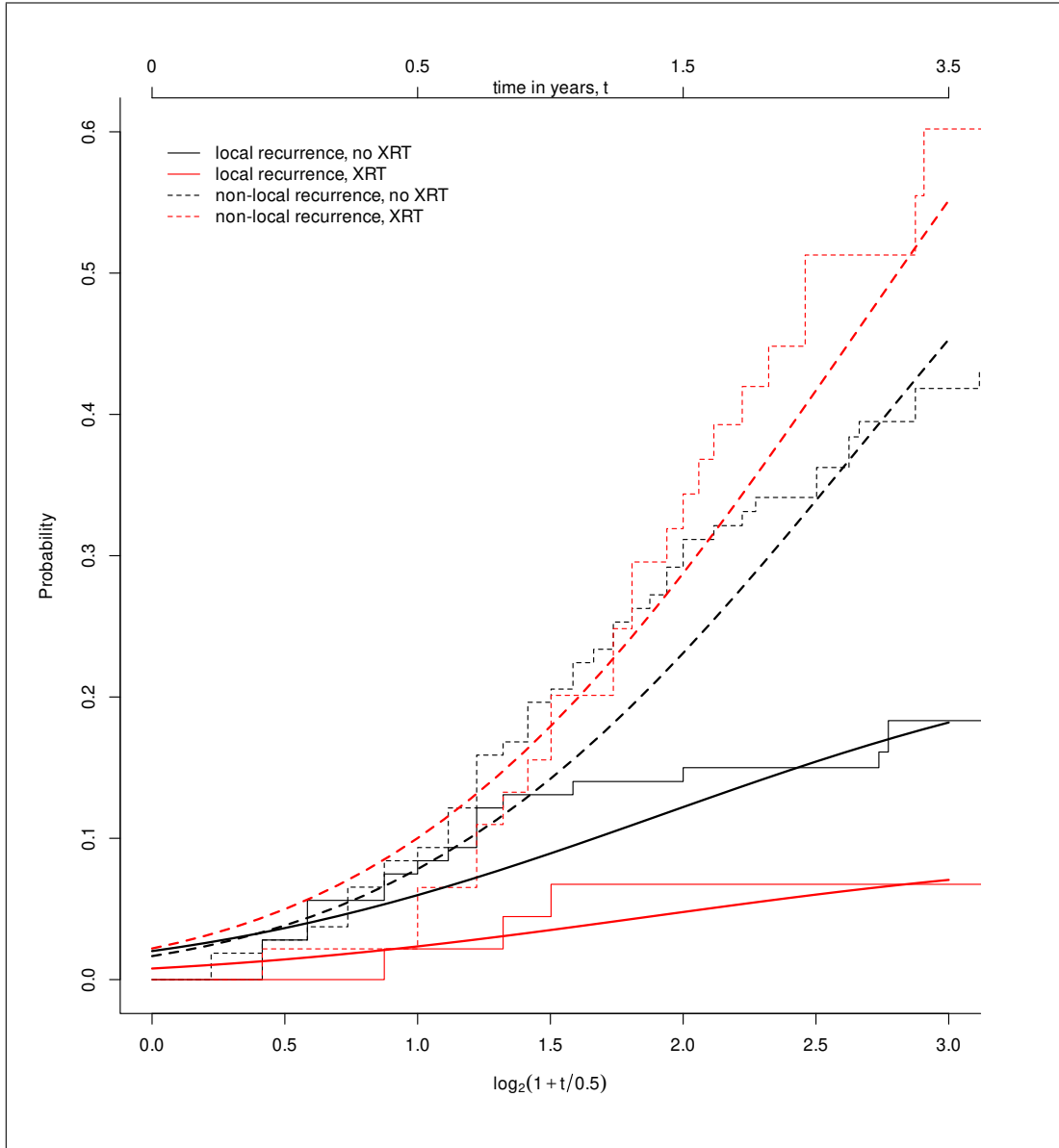


**Supplementary Figure 7.:** Cumulative incidence of local and distant recurrence by treatment group (XRT=adjuvant radiotherapy). Sensitivity analysis study of BNC parametric CIFs (smooth lines) superposed on Nelson-Aalen estimates (stepped lines) after  $y : t \rightarrow \log_2(1 + t/0.75)$  transformation of time scale.



**Supplementary Figure 8.:** Cumulative incidence of local and distant recurrence by treatment group (XRT=adjuvant radiotherapy). Sensitivity analysis study of BNC parametric CIFs superposed on Nelson-Aalen estimates after  $y : t \rightarrow \log_2(1 + t/1.5)$  transformation of time scale.





**Supplementary Figure 9.:** Cumulative incidence of local and distant recurrence by treatment group (XRT=adjuvant radiotherapy). Sensitivity analysis study of BNC parametric CIFs superposed on Nelson-Aalen estimates after  $y : t \rightarrow \log_2(1 + t/0.5)$  transformation of time scale.