Robustness of estimation methods in a survival cure model with mismeasured covariates

A. Bertrand[a,*], C. Legrand[a], D. Léonard[b], I. Van Keilegom[a]

[a] *Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Louvain-la-Neuve, Belgium*

[b] *Colorectal Surgery Unit, Department of Abdominal Surgery and Transplantation, Cliniques universitaires Saint-Luc, Brussels, Belgium*

Abstract

In medical applications, one frequently encounters time-to-event data. While classical survival methods are well known and broadly used to analyze such data, they do not allow one to take into account two phenomena which appear quite often in practice: individuals who will never experience the event of interest (they are cured from this event) and measurement error in the continuous covariates.

This paper deals with a model designed to take both features into account. Two approaches exist in the literature to estimate such a model. However, while they work well in many settings, they require information about the distribution of the measurement error which is rarely fully known in practice.

In this paper, we first justify the need to take the measurement error into account, via a theoretical study of the bias. We then present the results of an extensive simulation study investigating the robustness of both correction approaches with respect to their assumptions. The conclusions allow us to give some practical recommendations for similar situations. We conclude by analyzing the time until recurrence after surgery for rectal cancer patients, taking into account the advice from the simulation results.

Both correction methods have been implemented in the R package `miCoPTCM`.

Keywords: Bias correction, Cure fraction, Measurement error, Promotion time cure model, Semiparametric method.

1. Introduction

Time-to-event data occur frequently in medicine. While death is, of course, an example of such an event, other events might be of interest. For instance, in the study we consider here, the focus lies on the time until recurrence for patients suffering from rectal cancer and having been operated. When considering such an event, it is known that some patients will never experience this recurrence, even if they were followed for an infinite time. These nonsusceptible subjects are

*Corresponding author

Email addresses: `aurelie.bertrand@uclouvain.be` (A. Bertrand), `catherine.legrand@uclouvain.be` (C. Legrand), `daniel.leonard@uclouvain.be` (D. Léonard), `ingrid.vankeilegom@uclouvain.be` (I. Van Keilegom)

considered as cured, or immune, from the event of interest. This phenomenon appears regularly when studying non-lethal events in medicine, e.g. the contraction of the flu or the onset of age-related macular degeneration. For such data, classical survival analysis techniques are not suited, since they assume that, if the follow-up is sufficiently long, everyone will experience the event. This is why cure models appeared in the literature: they are specific models taking into account the presence of "cured" subjects in the population of interest.

In medical applications such as in the one we consider here, we are often interested in studying the impact of some patients or disease characteristics on the outcome. However, we often tend to ignore the fact that such biological variables may be measured with error. Such an error can appear when the device or the method used to measure the quantity of interest is not precise; an example is the measure of the maximal diameter of a tumor by an imaging technique. This feature is also present when what we want to measure fluctuates over time around its true value, as is the case, for example, with blood pressure. In the analysis of the time until recurrence in rectal cancer patients, one of the covariates of interest is the hemoglobin level in the blood, which is known to be measured with some error. The presence of mismeasured covariates in a statistical model can have several consequences (Carroll et al., 2006), among which the bias in the estimated effects of the covariates. It can lead to incorrectly concluding that a covariate has no significant effect on the response when it actually has (Cook and Stefanski, 1994).

In this paper, we address both features, cured individuals and mismeasured covariates, as they are often both present in medical problems, as illustrated by our example about rectal cancer, which we will study in this paper.

We assume that our data suffer from a problem which is very classical in survival analysis: right censoring. For some subjects, we are not able to observe the actual time point at which the event of interest occurs, since another event (called censoring) takes place before the event of interest. For such individuals, we only know that the actual event time is greater than the censoring time. Because of this right censoring, we are not able to distinguish between cured subjects (who are always censored) and non-cured subjects that are censored. This is why specific techniques are required to deal with such data.

More formally, with right-censored data, we observe $(Y_i, \Delta_i, \boldsymbol{X}_i)$ for $i = 1, \ldots, n$, where $Y_i = \min(T_i, C_i)$ with T_i the survival time and C_i the right-censoring time, $\Delta_i = I(T_i \leq C_i)$ is the censoring indicator (taking the value 1 for uncensored subjects, for whom T_i is observed, and the value 0 for censored subjects) and $\boldsymbol{X}_i$ is a P-dimensional vector of covariates. The vectors $(T_i, C_i, \boldsymbol{X}_i)$ are independent and identically distributed, with the same distribution as a generic vector $(T, C, \boldsymbol{X})$.

The promotion time cure model (Tsodikov, 1998) is one of the two survival cure models, which take into account the existence of cured subjects. Compared to the mixture cure model (Taylor, 1995), the promotion time cure model assumes a proportional structure for the hazard, as in the classical model of Cox (1972). The conditional survival function of the whole population, giving the probability of surviving up to time t, i.e. $S(t|\boldsymbol{x}) = P(T > t|\boldsymbol{X} = \boldsymbol{x})$, is modeled as

$$S(t|\boldsymbol{x}) = \exp\left\{-\theta(\boldsymbol{x})F(t)\right\}, \tag{1}$$

which is equivalent, for the conditional hazard function of T given $\boldsymbol{X} = \boldsymbol{x}$, to

$$h(t|\boldsymbol{x}) = \theta(\boldsymbol{x})F'(t), \tag{2}$$

where F is a proper baseline cumulative distribution function, θ is a known link function with an intercept, usually $\theta(\boldsymbol{x}) = \exp(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})$ for some P-dimensional vector of regression coefficients $\boldsymbol{\beta}$,

and $\boldsymbol{x}$ is the vector of covariates. We work with the semiparametric version of this model, in which no assumptions are made on the distribution of F.

Zeng et al. (2006) and Ma and Yin (2008) propose two different methods (based on respectively profiling and backfitting the likelihood function to be maximized) to estimate the model parameters of the promotion time cure model when there is no measurement error in the covariates. These methods provide estimates for the regression parameters $\boldsymbol{\beta}$, as well as for the baseline cumulative distribution function: it can be shown that its nonparametric maximum likelihood estimator is a step function which increases only at the observed event times. We denote by $\widehat{p}_i$ the estimated jump size of F at Y_i. For identifiability reasons, this estimated function is constrained to reach 1 at a predefined threshold (usually the largest observed event time), which amounts to considering as cured (for the estimation) the censored individuals with a censoring time larger than this threshold.

As motivated previously and studied in Bertrand et al. (2015), we consider that (some of) the continuous covariates are not correctly measured. Assuming the classical additive model for the error, we observe

$$\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}, \tag{3}$$

where $\boldsymbol{W}$ is the vector of observed covariates and $\boldsymbol{U}$ is the vector of measurement errors. We assume that $\boldsymbol{U}$ is independent of $\boldsymbol{X}$ and $\boldsymbol{U}$ follows a continuous distribution with mean zero and known covariance matrix $\boldsymbol{V}$. It is also assumed that (T, C) and $\boldsymbol{W}$ are independent given $\boldsymbol{X}$.

When (some of) the covariates are mismeasured, the technique of Zeng et al. (2006) and the non-corrected approach of Ma and Yin (2008) yield biased estimators. This is a well-known fact which holds for many statistical models; however, the form of this bias in the context of a promotion time cure model has, to the best of our knowledge, never been presented. This paper is the first one to study the form of the bias in this context.

Ma and Yin (2008) were the first authors to address the problem of measurement error in the covariates of the promotion time cure model. Their approach requires a Gaussian error with a known variance, and a specific form for θ, i.e. $\theta(\boldsymbol{x}) = \exp(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})$. They propose a corrected score strategy with a backfitting procedure which consists in solving (for $\boldsymbol{\beta}$, the p_i's and a Lagrange multiplier λ_{MY}) the score equations of the model in which the terms involving $\boldsymbol{X}$ have been replaced by terms involving $\boldsymbol{W}$ and $\boldsymbol{V}$:

$$\frac{1}{p_{(i)}} = \sum_{j=1}^{n} I\left(Y_{(i)} \leq Y_j < \infty\right) \exp(\boldsymbol{W}_j^T\boldsymbol{\beta} - \boldsymbol{\beta}^T\boldsymbol{V}\boldsymbol{\beta}/2) + n\lambda_{MY}, \qquad i = 1, \ldots, m$$

$$\sum_{i=1}^{m} p_{(i)} = 1$$

$$\sum_{i=1}^{n} \left\{\Delta_i I(Y_i < \infty)\boldsymbol{W}_i - F(Y_i)\exp(\boldsymbol{W}_j^T\boldsymbol{\beta} - \boldsymbol{\beta}^T\boldsymbol{V}\boldsymbol{\beta}/2)(\boldsymbol{W}_i - \boldsymbol{V}\boldsymbol{\beta})\right\} = \boldsymbol{0}$$

where $Y_{(i)}$ for $i = 1, \ldots, m$ denote the ordered distinct failure times, and $p_{(i)}$ the corresponding jumps of the baseline cumulative distribution function.

Bertrand et al. (2015) introduced an alternative correction method: they adapted the existing generic SIMEX algorithm (Cook and Stefanski, 1994) to the promotion time cure model. The SIMEX method has the advantage of being very intuitive and of allowing a graphical representation of the effect on the bias of both the measurement error and the correction. Moreover, it can be used with any continuous distribution for the measurement error, and with a flexible form

$\eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})$ of the link function. The scope of the correction can also be tuned through the choice of the extrapolation function. This approach consists of two steps. In the simulation step, a grid of positive values $\lambda = \lambda_1, \ldots, \lambda_K$ (e.g., $\{0, 0.5, 1, 1.5, 2\}$) is chosen. Then, for each of these values of λ, B (commonly, $B = 50$) new datasets are generated by adding artifical noise to the already mismeasured covariates: $\boldsymbol{W}_{i,\lambda,b} = \boldsymbol{W}_i + (\lambda \boldsymbol{V})^{1/2}\boldsymbol{Z}_{b,i}$ where $\boldsymbol{Z}_{b,i} \sim_{iid} N_P(\boldsymbol{0}, \boldsymbol{I}_P)$ and the $\boldsymbol{Z}_{b,i}$'s are independent of the data ($b = 1, \ldots, B$; $i = 1, \ldots, n$). The conditional variance of this contaminated vector is $\text{Var}\,(\boldsymbol{W}_{i,\lambda,b}|\boldsymbol{X}_i) = (1 + \lambda)\boldsymbol{V}$. The model parameters are then estimated using a method which does not take the measurement error into account. Here, as in Bertrand et al. (2015), the method introduced in Ma and Yin (2008) and described in some extent in Section 2 is used. For each λ and each $b \in \{1, 2, \ldots, B\}$, an estimate $\widehat{\boldsymbol{\beta}}_{\lambda,b}$ is obtained. For each λ, the average of the B estimates is then computed: $\widehat{\boldsymbol{\beta}}_\lambda = B^{-1}\sum_{b=1}^{B}\widehat{\boldsymbol{\beta}}_{\lambda,b}$. In the extrapolation step, an extrapolant (linear, quadratic, etc.) is chosen for each parameter, and the $\widehat{\boldsymbol{\beta}}_\lambda$ are modeled as a function of λ. The parameters of this model are estimated, and the SIMEX estimator is obtained by extrapolating it to $\lambda = -1$, the value at which $\text{Var}\,(\boldsymbol{W}_{i,\lambda,b}|\boldsymbol{X}_i)$ is the zero matrix.

Both the corrected score approach of Ma and Yin (2008) and the SIMEX method of Bertrand et al. (2015), compared in the latter paper via some simulations, proved to be useful for correcting for the bias due to mismeasured covariates in some simple settings. However, they are not always easy to use in concrete applications, since they require precise information about the distribution of the measurement error, which is not often available in practice. In particular, these methods rely on two main assumptions. The first one is the normality of the measurement error. It is required for the method of Ma and Yin (2008) and, although this is not the case for SIMEX, its practical performance in the non-Gaussian situation has never been verified in the promotion time cure model. The second strong assumption is the correct specification of the measurement error variance, which is necessary for both correction methods.

In practice, except in some specific contexts, the user rarely knows the exact form of the distribution of the error, or even its variance. In these cases, it is not obvious whether the correction methods really yield better results than the naive one, which does not take the measurement error into account. As a consequence, it is important to study the robustness of these methods with respect to their assumptions. We therefore performed an extensive simulation study aiming at investigating these questions and providing practical recommendations on whether it is useful to use a correction method, and which one is likely to provide the best results for various settings. Throughout this simulation study, we denote by "Naive" the naive method (i.e., not correcting for the measurement error) based on the backfitting approach, by "MY" the corrected score approach of Ma and Yin (2008) and by "SIMEX" the SIMEX algorithm designed for the promotion time cure model by Bertrand et al. (2015). Our R (R Core Team, 2015) implementation of the different estimation methods used in these simulations is provided in the R package `miCoPTCM`, available on the CRAN website (cran.r-project.org).

The structure of this paper is as follows. In the next section, we derive an expression allowing to assess the bias of the estimator of the regression parameters obtained with the naive method. Section 3 then presents the simulation results and the recommendations pertaining to the issues of non-Gaussian distribution and unknown variance of the measurement error. This allows us to elaborate a strategy to study the time until recurrence of cancer for patients suffering from rectal cancer; the obtained results are presented in Section 4. Finally, the conclusions and limitations of this study as well as some ideas for further research are discussed in Section 5. Appendix 1

contains the proof of the result of Section 2, while additional tables of simulation results pertaining to Section 3 can be found in Appendix 2.

2. The bias of the naive estimator

We first derive an expression for the asymptotic bias which appears in the naive estimator when we do not take the measurement error into account. Such an expression has indeed never been obtained in the context of the promotion time cure model with mismeasured covariates. Being able to assess this bias however helps justify the need for correction methods even in large samples.

We follow an approach similar to the ones used by Hughes (1993) and Li and Lin (2000), but here allowing for several covariates, in order to obtain an equation linking the large-sample estimator of $\boldsymbol{\beta}$ and the true parameter.

When we assume no measurement error, the promotion time cure model can be fitted using the backfitting approach. It can be shown (Ma and Yin, 2008) that the nonparametric maximum likelihood estimator of the baseline cumulative distribution function is a step function which increases only at the observed event times. We denote by $\widehat{p}_i$ the estimated jump size of F at Y_i.

The iterative backfitting procedure, as described in Ma and Yin (2008), alternates between two steps. We first solve the score equations for the p_i's and λ_{MY} (a Lagrange multiplier) by fixing $\boldsymbol{\beta}$, and obtain $\widehat{\lambda}_{MY,\beta}$ and the $\widehat{p}_{i,\beta}$'s. Then, we solve the score equations for $\boldsymbol{\beta}$ by fixing the p_i's and λ_{MY} at the values $\widehat{\lambda}_{MY,\beta}$ and $\widehat{p}_{i,\beta}$. The equation to be solved for $\boldsymbol{\beta}$ is

$$n^{-1}\sum_{i=1}^{n}\left\{\Delta_i - \widehat{F}_{\beta}(Y_i)e^{\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}}\right\}(1,\boldsymbol{X}_i^T)^T = \mathbf{0}, \tag{4}$$

where $\widehat{F}_{\beta}(Y_i) = \sum_{Y_j\leq Y_i,\Delta_j=1}\widehat{p}_{j,\beta}$.

When there is measurement error in the covariates, such that $\boldsymbol{W}$ is observed instead of $\boldsymbol{X}$, then the naive estimator $\widehat{\boldsymbol{\beta}}^*$ solves (in $\boldsymbol{\beta}^*$)

$$n^{-1}\sum_{i=1}^{n}\Delta_i(1,\boldsymbol{W}_i^T)^T - n^{-1}\sum_{i=1}^{n}\widehat{F}_{\beta^*}(Y_i)e^{\beta_0^*+\boldsymbol{W}_i^T\boldsymbol{\beta}^*}(1,\boldsymbol{W}_i^T)^T = \mathbf{0}. \tag{5}$$

Without loss of generality, we assume that $\boldsymbol{X}$ is standardized, so that $E(\boldsymbol{X}) = \mathbf{0}$, $V(\boldsymbol{X}) = \boldsymbol{I}_P$ (P being the dimension of $\boldsymbol{X}$) and $V(\boldsymbol{W}|\boldsymbol{X}) = V(\boldsymbol{X})^{-1/2}V(\boldsymbol{U})V(\boldsymbol{X})^{-1/2} = \boldsymbol{\Lambda}$ (assuming that we applied the same transformation to the error as to $\boldsymbol{X}$, i.e. $V(\boldsymbol{X})^{-1/2}(\cdot - E(\boldsymbol{X}))$).

To derive the expression of the bias, as in Hughes (1993) and Li and Lin (2000), we assume type I censoring, such that the observed time for individual i is $Y_i = \min(T_i, t_c)$ for a fixed t_c. This would be the case, for instance, when all patients in a study are followed for a fixed length of time.

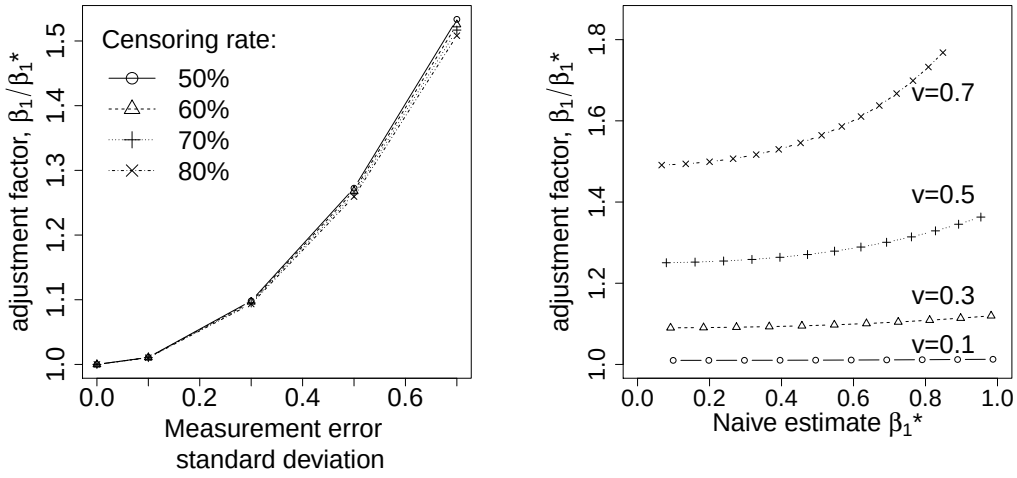
When n is large, (5) tends in probability to

$$
\begin{aligned}
& E_{\boldsymbol{W},T}(I(T \leq t_c)(1, \boldsymbol{W}^T)^T) - E_{\boldsymbol{W},Y}\left[F_{\beta^*}(Y)e^{\beta_0^* + \boldsymbol{W}^T\boldsymbol{\beta}^*}(1, \boldsymbol{W}^T)^T\right] \\
&= \begin{pmatrix} E_T(I(T \leq t_c)) \\ E_{\boldsymbol{W},T}(I(T \leq t_c)\boldsymbol{W}) \end{pmatrix} - \begin{pmatrix} E_{\boldsymbol{W},Y}\left[F_{\beta^*}(Y)e^{\beta_0^* + \boldsymbol{W}^T\boldsymbol{\beta}^*}\right] \\ E_{\boldsymbol{W},Y}\left[F_{\beta^*}(Y)e^{\beta_0^* + \boldsymbol{W}^T\boldsymbol{\beta}^*}\boldsymbol{W}\right] \end{pmatrix} \\
&= \begin{pmatrix} E_T(I(T \leq t_c)) \\ E_{\boldsymbol{W},T}(I(T \leq t_c)\boldsymbol{W}) \end{pmatrix} - \begin{pmatrix} E_{\boldsymbol{W},T}\left[I(T \leq t_c)F_{\beta^*}(T)e^{\beta_0^* + \boldsymbol{W}^T\boldsymbol{\beta}^*}\right] \\ E_{\boldsymbol{W},T}\left[I(T \leq t_c)F_{\beta^*}(T)e^{\beta_0^* + \boldsymbol{W}^T\boldsymbol{\beta}^*}\boldsymbol{W}\right] \end{pmatrix} \\
&\quad - \begin{pmatrix} E_{\boldsymbol{W},T}\left[I(T > t_c)F_{\beta^*}(t_c)e^{\beta_0^* + \boldsymbol{W}^T\boldsymbol{\beta}^*}\right] \\ E_{\boldsymbol{W},T}\left[I(T > t_c)F_{\beta^*}(t_c)e^{\beta_0^* + \boldsymbol{W}^T\boldsymbol{\beta}^*}\boldsymbol{W}\right] \end{pmatrix} \\
&= \begin{pmatrix} E_T(I(T \leq t_c)) \\ E_T\left[I(T \leq t_c)E_{\boldsymbol{W}|T}(\boldsymbol{W})\right] \end{pmatrix} - \begin{pmatrix} E_T\left[I(T \leq t_c)F_{\beta^*}(T)E_{\boldsymbol{W}|T}\left(e^{\beta_0^* + \boldsymbol{W}^T\boldsymbol{\beta}^*}\right)\right] \\ E_T\left[I(T \leq t_c)F_{\beta^*}(T)E_{\boldsymbol{W}|T}\left(e^{\beta_0^* + \boldsymbol{W}^T\boldsymbol{\beta}^*}\boldsymbol{W}\right)\right] \end{pmatrix} \\
&\quad - \begin{pmatrix} E_T\left[I(T > t_c)F_{\beta^*}(t_c)E_{\boldsymbol{W}|T}\left(e^{\beta_0^* + \boldsymbol{W}^T\boldsymbol{\beta}^*}\right)\right] \\ E_T\left[I(T > t_c)F_{\beta^*}(t_c)E_{\boldsymbol{W}|T}\left(e^{\beta_0^* + \boldsymbol{W}^T\boldsymbol{\beta}^*}\boldsymbol{W}\right)\right] \end{pmatrix} = \mathbf{0}. \qquad (6)
\end{aligned}
$$

Each of the three above expectations can be calculated assuming the promotion time cure model (1) and the additive noise model (3). Details are given in Appendix 1. If we assume that $(\boldsymbol{W}|\boldsymbol{X} = \boldsymbol{x}) \sim N(\boldsymbol{x}, \boldsymbol{\Lambda})$, we obtain the following formulation of (6):

$$
\begin{aligned}
& \begin{pmatrix} 1 - \int e^{-\eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \\ -\int \boldsymbol{x} f_{\boldsymbol{X}}(\boldsymbol{x})e^{-\eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}d\boldsymbol{x} \end{pmatrix} \\
& - \begin{pmatrix} e^{\beta_0^* + \boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}\int e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})\left\{\int_0^{t_c} F_{\beta^*}(t)\eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)}dt\right\} d\boldsymbol{x} \\ e^{\beta_0^* + \boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}\int(\boldsymbol{x} + \boldsymbol{\Lambda}\boldsymbol{\beta}^*)e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})\left\{\int_0^{t_c} F_{\beta^*}(t)\eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)}dt\right\} d\boldsymbol{x} \end{pmatrix} \\
& - \begin{pmatrix} e^{\beta_0^* + \boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}F_{\beta^*}(t_c)\int e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})e^{-\eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}\, d\boldsymbol{x} \\ e^{\beta_0^* + \boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}F_{\beta^*}(t_c)\int(\boldsymbol{x} + \boldsymbol{\Lambda}\boldsymbol{\beta}^*)e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})e^{-\eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}\, d\boldsymbol{x} \end{pmatrix} = \mathbf{0}.
\end{aligned}
$$

This equation can be solved numerically, for $\boldsymbol{\beta}^*$ given $\boldsymbol{\beta}$, in order to obtain the relationship between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$. The effect on the asymptotic bias of each parameter of a particular setting can then be investigated and represented on a graph. As an illustration, we consider a model with $\beta_0 = -0.1$, $\boldsymbol{\beta} = (0.5, -0.5)^T$, $V(X_1) = 1$, $V(X_2) = 1$, $Cov(X_1, X_2) = 0$. We assume that X_1 is measured with some error, but X_2 is not. As mentioned by Li and Lin (2000) in the context of frailty models, the asymptotic bias is not easily computed when the baseline function is left unspecified. We hence proceed similarly as in their paper, and assume (for this illustration) an exponential distribution of mean 8 for the baseline cumulative distribution function. We consider four different values for the censoring rate, 50%, 60%, 70% and 80%. Since the censoring rate equals $\int S(t_c|\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}$, this can be achieved for appropriate choices of t_c. Figure 1a represents the adjustment factor for the coefficient related to the mismeasured covariate, i.e. the ratio β_1/β_1^*, as a function of the measurement error variance. We see that, in this particular setting, the censoring rate does not influence the adjustment factor much. Moreover, as expected, the bias increases with the measurement error variance.



(a) Adjustment factor as a function of the standard deviation of the measurement error, for different values of the censoring rate.

(b) Adjustment factor as a function of the naive estimate, for different values of the standard deviation v of the measurement error.

Figure 1: Two examples of plots that can be obtained from Equation (6).

However, as explained by Hughes (1993), a graphical representation of greater practical interest is one such as in Figure 1b: it allows one to deduce the adjustment factor (and hence, the true value β_1) to be expected for a known value of the estimate β_1^* obtained by using the naive method. Here, models with values of β_1 between 0.1 and 1.5 with a censoring rate of 70% were estimated. The ratio β_1/β_1^* is plotted as a function of the asymptotic naive estimate, for different values of the measurement error standard deviation. Here again, it appears that the adjustment factor increases with the measurement error standard deviation. Moreover, for the largest values of the variance, the adjustment factor is not constant, but increases with the value of the naive estimate.

3. Simulation study

The previous section justifies the need for a correction, in order to avoid (or, at least, reduce) the bias in the estimated regression coefficients. However, as was discussed above, the use of both the corrected score approach (Ma and Yin, 2008) and the SIMEX algorithm (Bertrand et al., 2015) requires two types of information about the error that are not always known in practice. In this section, we present the results of an extensive simulation study investigating the robustness of both approaches (as well as the naive method) with respect to the assumptions of normality and known variance of the measurement error. These results allow us to provide practical recommendations on whether, in a given context, it is useful to correct and, if so, about which method should be preferred. These recommendations are based on two criteria, the mean squared error (MSE) and the bias. Techniques dealing with measurement error classically focus on the latter criterion; the former one allows one to consider a balance between the decrease in the bias and the increase in the variance.

The model used for simulating the data is

$$S(t|X_1, X_2) = \exp\left\{-\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2) F(t)\right\},$$

where the parameters were chosen to mimic a possible practical setting, i.e. $F(t)$ is the cumulative distribution function (cdf) of an exponential distribution with mean 6, truncated at $t = 20$, $X_2 \sim Bernoulli(0.5)$, $X_1 = (Y_1 - 0.5)(1/12)^{-1/2}$ where $Y_1 \sim Uni[0, 1]$, such that $E(X_1) = 0$ and $V(X_1) = 1$. X_2 can then correspond to a treatment indicator, while X_1 can represent a continuous biomarker measured with error. The censoring times are generated (independently of the covariates and of the survival time) from an exponential distribution with mean 5 truncated at $t = 30$.

For each setting, we performed 500 replications. We report the bias and MSE for both covariate coefficients, β_1 and β_2. Besides these coefficients, another quantity of interest in practical applications is the estimated cure rate at a given value of the covariates. We denote by CR_α the estimated cure rate at the quantile of order α of the continuous covariate X_1 (and at a fixed value of X_2). We computed the MSE and bias for $\widehat{CR}_{.25}$ and $\widehat{CR}_{.75}$ in all settings; we do not report extensively these results here, but we mention the most important conclusions.

3.1. Non-normality of the measurement error

In this first section, we investigate the robustness of each method with respect to the assumption of normality of the measurement error, an assumption which can rarely be assessed in practice. It is required for MY, while SIMEX can explicitly accommodate non-Gaussian distributions, but its practical performance has never been verified in this context.

For reasons of brevity, we restrict attention in this simulation study to three representative settings (summarized in Table 1), which cover a range of cure proportions and censoring proportions, and which include both continuous and binary covariates. The true cure rate reporter in the table is the proportion of observations that are generated as cured, while the observed cure rate is the proportion of censored observations with a censoring time larger than the largest observed event time (the cure threshold). Other settings have been simulated but their results are not reported here.

Table 1: The three settings considered in the simulation study

	Setting 1	Setting 2	Setting 3
β_0	1	0.2	-0.5
β_1	0.5	1	0.75
β_2	-0.5	-0.5	-0.5
censoring rate	42%	56%	69%
true cure rate	16%	39%	58%
observed cure rate	3%	6%	9%

We assume that X_1 is measured with error, so that $W = X_1 + U$ is observed. As far as the distribution of this error U is concerned, four different settings are used. In the first one, $U \sim N(0, \sigma_U^2)$ with variance $\sigma_U^2 \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$. This Gaussian case will be used as a standard of comparison, but it will also allow us to investigate the effect on the estimation of increasing the measurement error variance. The second distribution under study is the Student-t: $U = aT$ where $T \sim t_\nu$, such that $V(U) = a^2 \frac{\nu}{\nu-2}$. We take $a = 0.5$ and $\nu \in \{4, 10\}$, hence $V(U) \in \{0.5, 0.3125\}$. This distribution allows us to consider a case close to the Gaussian one, but with

heavier tails. In the third setting, we use an asymmetrical distribution: $U = b(T-k)$ where $T \sim \chi^2_k$, such that $V(U) = 2b^2k$. We take $b = \sqrt{0.1}$ and $k \in \{1, 2, 3\}$, hence $V(U) \in \{0.2, 0.4, 0.6\}$. Finally, we consider the case where $U = T - \mu$ where $T \sim \textit{Laplace}(\mu, b)$, with variance $2b^2$. This is another symmetrical distribution. We take $b \in \{\sqrt{0.05}, \sqrt{0.15}, \sqrt{0.25}\}$, so that $V(U) \in \{0.1, 0.3, 0.5\}$.

3.1.1. Results

The numerical results can be found in Tables 2 and 3 and in the Supplementary Materials for Settings 1, 2 and 3, respectively. The results pertaining to the Gaussian distribution with $\sigma^2_U \in \{0.2, 0.4, 0.6\}$ can also be found in the Supplementary Materials.

In general, as already observed in Bertrand et al. (2015), SIMEX is usually the best method for decreasing the MSE, while MY often outperforms SIMEX for the bias.

The sign of this bias clearly depends on the estimation method. The bias for β_1 is usually negative for Naive and SIMEX, and positive for MY (except with the Chi-squared distribution). This illustrates the fact that MY tends to overcorrect for the bias, a tendency that already appeared in the simulation results in Ma and Yin (2008) and in Bertrand et al. (2015). MY hence corrects more than SIMEX, which is known to be a more conservative method, particularly with the quadratic extrapolant (Cook and Stefanski, 1994).

Moreover, the effect of an increasing measurement error variance also depends on the method under consideration. With Naive, both the bias and the MSE for β_1 increase with the measurement error variance, while this is not necessarily the case for β_2. This increase in the MSE is due to the increase in the bias, since the variance tends to decrease (this decrease was already observed by Ma and Yin (2008), as a result of more variation in the covariate). The conclusion is the same with SIMEX, which could be expected, since this method heavily depends on the naive estimates. With MY, there is no clear trend for the bias, while the MSE of both β_1 and β_2 increases with the measurement error variance, driven by increases in the variance of the estimators.

The bias in the estimation of β_1 has consequences on the estimated cure rates. $\widehat{CR}_{.25}$ is an increasing function of $\widehat{\beta}_1$, while $\widehat{CR}_{.75}$ is a decreasing function of it. As a consequence, *for a fixed value of* $\widehat{\beta}_0$, an underestimation of β_1 will yield a decrease in the estimation of $CR_{.25}$. This will result in a decrease in the bias of $\widehat{CR}_{.25}$, and an increase in the bias of $\widehat{CR}_{.75}$. Furthermore, the conclusions for $\widehat{CR}_{.25}$ are less clear than those for $\widehat{CR}_{.75}$. When $\widehat{\beta}_1$ and X_1 do not have the same sign (in our settings, this is the case for $\widehat{CR}_{.25}$), $\widehat{\beta}_0 + \widehat{\beta}_1 X_1$ tends to be small, so that small changes in $\widehat{\beta}_1$ or in $\widehat{\beta}_0$ have only a small effect on $\exp(\widehat{\beta}_0 + \widehat{\beta}_1 X_1)$, and consequently on the estimated cure rate. On the contrary, when $\widehat{\beta}_1$ and X_1 have the same sign, the estimated cure rate is much more sensitive to variations in the estimated parameters.

Finally, the three settings under consideration can be compared. Settings 1 and 3 are the settings with the lowest true values for β_1 (0.5 and 0.75, instead of 1). This fact could explain some observations: a smaller impact of the correction (since this correction mainly impacts β_1) and of the increased measurement error variance, a better robustness to a misspecification of the error distribution in X_1. This could also justify the similarity in the MSE's for the cure rate across methods and measurement error variances in Settings 1 and 3. Moreover, in Setting 1 (with the smallest value of β_1), the MSE for β_2 is quite similar among methods. Finally, with Naive, the bias for β_2 is usually smaller in Setting 1 than in the other two settings (although the value of β_2 used in the simulations is the same, -0.5).

To assess the robustness of a method to a misspecification of the measurement error distribution, we compare, for a given measurement error variance, the value of one criterion (MSE or bias) in

the Gaussian case with the value of this criterion obtained with the other distributions.

We first consider the two possible estimation methods which can not explicitely accommodate non-Gaussian distributions: Naive and MY. For the MSE of β_2, Naive seems robust (the increases in the MSE are not large); for β_1, this method stays relatively robust, except when the distribution of the measurement error is asymmetrical. As far as the bias is concerned, for β_1, Naive seems quite robust, except when the measurement error distribution is asymmetrical (in that case, the increase in the bias can be very large). For β_2, Naive is quite robust for Settings 1 and 3, a bit less in Setting 2.

For the MSE in β_1, MY is not as robust as Naive: the MSE can stay similar, decrease, or increase (differences can be dramatic when the measurement error variance is large). For β_2, MY seems robust (the increases or decreases in the MSE are much smaller). When considering the bias, for β_2, there is a decrease (sometimes very large) in the bias with MY when switching from a Gaussian to a non-Gaussian distribution. This counterintuitive result could be explained by the already observed tendency of MY to correct more than necessary in the Gaussian case. For β_1, the conclusion is the same as for β_2, except when the error distribution is asymmetrical: in that case, the bias can either increase or decrease.

Since SIMEX can be run with a non-Gaussian distribution, two questions are of interest here: Is SIMEX robust with respect to the distribution of the measurement error? If so, is it worth using the true distribution when it is known? For β_1 and β_2, the increases in the MSE when the true distribution is not Gaussian (compared to when it is) are not large, except in Setting 2 (and, for β_1, when the measurement error distribution is asymmetrical). Taking the true distribution into account in SIMEX does not improve the MSE (except for β_1 in Setting 2, when the measurement error distribution is asymmetrical). The bias for β_1 does not increase a lot from the Gaussian to the non-Gaussian case, except for Setting 2, and when the measurement error distribution is asymmetrical. For β_2, there can be increases or decreases of relatively moderate size in the bias. Taking the true distribution into account in SIMEX improves the bias for β_1 when the measurement error distribution is asymmetrical; otherwise, it yields results very similar to those obtained without taking the true distribution into account.

Finally, the last information that can be retrieved from the simulation results is the minimum value of the measurement error variance for which it is useful to correct. In terms of MSE, for β_1, it is always useful to correct with SIMEX (i.e., the decrease in the MSE seems significant), except in some cases for the lowest value of the measurement error variance. The conclusions for MY are not so clear. For β_2, the coefficient associated with the variable measured without error, we never benefit from correcting. In terms of bias, for β_1, it is always useful to correct, even for a small measurement error variance. For β_2, there is no clear conclusion.

3.1.2. Recommendations

Based on the observations from the previous subsection, which are consistent with the results from other simulations not reported here, we can make the following recommendations.

When the main objective is to decrease the MSE, the best strategy depends on which covariate effect is of interest. For β_2, the coefficient of the covariate without measurement error, given the two correction methods that are at our disposal, it is not useful to correct. If we nevertheless want to use a correction method (because, for example, the focus of the study is another parameter for which it is better to correct), then it should be SIMEX, since it does not increase the MSE much. For β_1, the parameter related to the mismeasured covariate, we should always correct with SIMEX (when MY decreases the MSE, the decrease is smaller than with SIMEX; and in some cases, MY

Table 2: Results for the robustness regarding the assumption of normality of the measurement error in Setting 1.

			Naive		MY		SIMEX (correct dist.)		SIMEX (Gaussian dist.)	
Distribution	Var.		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Gaussian	0.1	Bias	-0.040	-0.014	0.015	-0.020	0.007	-0.018	0.007	-0.018
		Var	0.009	0.038	0.012	0.040	0.011	0.040	0.011	0.040
		MSE	0.011	0.039	0.012	0.040	0.011	0.040	0.011	0.040
	0.3	Bias	-0.114	-0.007	0.030	-0.024	-0.014	-0.017	-0.014	-0.017
		Var	0.007	0.038	0.018	0.042	0.013	0.041	0.013	0.041
		MSE	0.020	0.038	0.019	0.043	0.013	0.041	0.013	0.041
	0.5	Bias	-0.168	-0.002	0.048	-0.030	-0.048	-0.015	-0.048	-0.015
		Var	0.006	0.038	0.029	0.046	0.013	0.042	0.013	0.042
		MSE	0.034	0.038	0.031	0.047	0.015	0.042	0.015	0.042
	0.7	Bias	-0.212	0.004	0.061	-0.029	-0.087	-0.008	-0.087	-0.008
		Var	0.005	0.038	0.038	0.050	0.011	0.041	0.011	0.041
		MSE	0.050	0.038	0.042	0.051	0.019	0.042	0.019	0.042
Student-t	0.3125	Bias	-0.125	-0.002	0.017	-0.017	-0.025	-0.011	-0.026	-0.011
		Var	0.007	0.039	0.019	0.044	0.013	0.042	0.013	0.042
		MSE	0.023	0.039	0.020	0.044	0.014	0.042	0.014	0.042
	0.5	Bias	-0.170	0.000	0.043	-0.026	-0.051	-0.012	-0.050	-0.012
		Var	0.007	0.038	0.035	0.048	0.015	0.041	0.015	0.042
		MSE	0.036	0.038	0.037	0.049	0.018	0.041	0.018	0.042
Chi-square	0.2	Bias	-0.098	-0.006	-0.011	-0.016	-0.017	-0.014	-0.028	-0.013
		Var	0.008	0.040	0.015	0.044	0.014	0.043	0.013	0.043
		MSE	0.018	0.040	0.015	0.044	0.015	0.043	0.014	0.043
	0.4	Bias	-0.163	-0.001	-0.024	-0.018	-0.054	-0.012	-0.067	-0.010
		Var	0.007	0.039	0.021	0.046	0.016	0.043	0.014	0.042
		MSE	0.034	0.039	0.022	0.046	0.019	0.043	0.019	0.042
	0.6	Bias	-0.213	0.005	-0.044	-0.017	-0.101	-0.006	-0.110	-0.006
		Var	0.006	0.039	0.027	0.049	0.014	0.043	0.013	0.042
		MSE	0.051	0.039	0.029	0.049	0.025	0.043	0.025	0.042
Laplace	0.1	Bias	-0.043	-0.010	0.011	-0.015	0.004	-0.014	0.003	-0.014
		Var	0.009	0.039	0.012	0.041	0.012	0.041	0.012	0.040
		MSE	0.011	0.039	0.012	0.041	0.012	0.041	0.012	0.041
	0.3	Bias	-0.119	-0.001	0.020	-0.016	-0.019	-0.009	-0.021	-0.009
		Var	0.008	0.039	0.018	0.044	0.014	0.042	0.014	0.042
		MSE	0.022	0.039	0.019	0.044	0.014	0.042	0.014	0.042
	0.5	Bias	-0.172	0.004	0.033	-0.018	-0.053	-0.005	-0.055	-0.005
		Var	0.007	0.039	0.030	0.049	0.015	0.043	0.015	0.043
		MSE	0.037	0.039	0.031	0.049	0.018	0.043	0.018	0.043

Table 3: Results for the robustness regarding the assumption of normality of the measurement error in Setting 2.

			Naive		MY		SIMEX (correct dist.)		SIMEX (Gaussian dist.)	
Distribution	Var.		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Gaussian	0.1	Bias	-0.104	-0.006	0.041	-0.024	0.010	-0.018	0.010	-0.018
		Var	0.017	0.060	0.030	0.065	0.025	0.063	0.025	0.063
		MSE	0.028	0.060	0.032	0.066	0.025	0.064	0.025	0.064
	0.3	Bias	-0.269	0.011	0.107	-0.040	-0.058	-0.014	-0.058	-0.014
		Var	0.013	0.060	0.074	0.083	0.028	0.069	0.028	0.069
		MSE	0.086	0.060	0.086	0.085	0.031	0.069	0.031	0.069
	0.5	Bias	-0.408	0.021	0.060	-0.037	-0.189	-0.001	-0.189	-0.001
		Var	0.008	0.056	0.154	0.101	0.017	0.064	0.017	0.064
		MSE	0.174	0.057	0.157	0.103	0.053	0.064	0.053	0.064
	0.7	Bias	-0.506	0.034	-0.056	-0.045	-0.298	0.013	-0.298	0.013
		Var	0.006	0.057	0.291	0.129	0.013	0.065	0.013	0.065
		MSE	0.262	0.059	0.294	0.131	0.102	0.065	0.102	0.065
Student-t	0.3125	Bias	-0.293	0.022	0.066	-0.020	-0.086	0.001	-0.088	0.001
		Var	0.015	0.064	0.080	0.089	0.034	0.075	0.033	0.074
		MSE	0.101	0.064	0.084	0.090	0.041	0.075	0.041	0.074
	0.5	Bias	-0.419	0.042	-0.005	-0.004	-0.208	0.023	-0.213	0.022
		Var	0.012	0.057	0.182	0.099	0.029	0.068	0.028	0.069
		MSE	0.187	0.059	0.182	0.099	0.073	0.069	0.073	0.069
Chi-square	0.2	Bias	-0.266	0.014	-0.075	-0.012	-0.087	-0.006	-0.131	-0.001
		Var	0.021	0.069	0.063	0.088	0.050	0.083	0.040	0.080
		MSE	0.091	0.069	0.069	0.088	0.058	0.084	0.057	0.080
	0.4	Bias	-0.410	0.034	-0.163	0.001	-0.207	0.013	-0.247	0.017
		Var	0.015	0.066	0.064	0.103	0.042	0.085	0.034	0.081
		MSE	0.183	0.067	0.090	0.103	0.085	0.085	0.095	0.081
	0.6	Bias	-0.499	0.037	-0.231	-0.004	-0.297	0.013	-0.328	0.015
		Var	0.011	0.066	0.057	0.109	0.033	0.088	0.027	0.084
		MSE	0.260	0.067	0.111	0.109	0.121	0.088	0.135	0.084
Laplace	0.1	Bias	-0.113	0.001	0.027	-0.016	-0.001	-0.012	-0.002	-0.011
		Var	0.018	0.063	0.030	0.070	0.026	0.068	0.025	0.067
		MSE	0.030	0.063	0.031	0.070	0.026	0.068	0.025	0.068
	0.3	Bias	-0.284	0.022	0.055	-0.018	-0.081	0.000	-0.085	0.001
		Var	0.014	0.064	0.075	0.090	0.031	0.076	0.030	0.075
		MSE	0.095	0.064	0.078	0.090	0.037	0.076	0.037	0.075
	0.5	Bias	-0.417	0.042	-0.004	-0.009	-0.205	0.021	-0.209	0.021
		Var	0.009	0.062	0.074	0.104	0.021	0.075	0.021	0.075
		MSE	0.183	0.064	0.074	0.104	0.063	0.075	0.064	0.075

actually increases the MSE).

The strategy which should be used when the criterion to be minimized is the bias is slightly different. For β_2, there is now no clear conclusion about whether it is useful to correct, although the best correction method is SIMEX, since it nearly always decreases the bias. For β_1, using a correction method is always profitable; the best correction method is MY, except when the distribution is Gaussian and the variance is not too large (under 0.3 or 0.4): in that case, SIMEX must be preferred.

3.2. Misspecification of the measurement error variance

The second strong assumption which is often not met in practice is the knownledge of the measurement error variance. As a consequence, we would like to know the consequences of using each of the correction methods with a misspecified value for it. In this section, we summarize the simulation results about the robustness of the correction methods with respect to this assumption. The objective is not only to obtain practical recommendations about which estimation method to use in a given setting, but also to discover whether we could gain, in some cases, from deliberately over- or underspecifying this variance.

Here, we present the results for the same three settings as in Section 3.1, with the measurement error generated as $U \sim N(0, \sigma_U^2)$ in the simulations but assumed to be $N(0, \sigma_E^2)$ in the estimation procedure. We consider $\sigma_U^2 = 0.1$ associated with $\sigma_E^2 \in \{0.05, 0.10, 0.15\}$ and $\sigma_U^2 = 0.3$ associated with $\sigma_E^2 \in \{0.20, 0.25, 0.30, 0.35, 0.40\}$.

3.2.1. Results

The numerical results for Settings 1, 2 and 3 can be found in Tables 4, 5 and in the Supplementary Materials, respectively, and allow us to draw the following conclusions about the robustness of SIMEX and MY.

For the MSE for β_2, SIMEX is robust. For the MSE for β_2 and β_1, SIMEX seems more robust than MY to variations in the assumed measurement error variance: changing the variance used in the estimation procedure leads to larger changes (increases or decreases) in the MSE with MY than with SIMEX. The changes are also larger for β_1 than for β_2.

As far as the bias is concerned, for both methods, the results for β_1 depend a lot on the variance used in the estimation procedure. There is more change in the bias for β_1 than for β_2. SIMEX also seems a bit more robust than MY. SIMEX is quite robust for the bias in β_2.

Another interesting phenomenon appearing in the results is that MY works best (according to the bias or the MSE) when the measurement error variance is (to some extent) underspecified: this is because, as already noted in Section 3.1, MY tends to correct more than necessary. The opposite is true for β_1 with SIMEX: this is a conservative method, so we can gain from overspecifying the error variance.

3.2.2. Recommendations

Here are the practical pieces of advice that can be derived from the simulation results.

We first consider the case in which we assume the same (incorrect) value of the measurement error variance for both approaches. The lowest MSE for β_2, the parameter related to the covariate without measurement error, is obtained with Naive. If the analyst is led to use a correction method, it should be SIMEX, which yields lower MSE than MY. For β_1, the coefficient of the mismeasured covariate, the conclusions about the MSE are a bit more subtle. When the measurement error variance is small, especially if this small variance is overspecified, a correction is not needed. For

Table 4: Results for the case of misspecification of the measurement error variance in Setting 1.

			Naive		MY		SIMEX	
Variance in simulations	Variance in estimation		β_1	β_2	β_1	β_2	β_1	β_2
0.1	0.05	Bias	-0.040	-0.014	-0.014	-0.017	-0.017	-0.016
		Var	0.009	0.038	0.010	0.039	0.010	0.039
		MSE	0.011	0.039	0.010	0.039	0.010	0.039
	0.1	Bias	-0.040	-0.014	0.015	-0.020	0.007	-0.018
		Var	0.009	0.038	0.012	0.040	0.011	0.040
		MSE	0.011	0.039	0.012	0.040	0.011	0.040
	0.15	Bias	-0.040	-0.014	0.049	-0.024	0.032	-0.021
		Var	0.009	0.038	0.014	0.040	0.013	0.040
		MSE	0.011	0.039	0.017	0.041	0.014	0.041
	0.2	Bias	-0.040	-0.014	0.088	-0.028	0.056	-0.023
		Var	0.009	0.038	0.018	0.042	0.014	0.041
		MSE	0.011	0.039	0.025	0.042	0.017	0.041
0.3	0.2	Bias	-0.114	-0.007	-0.030	-0.017	-0.047	-0.014
		Var	0.007	0.038	0.013	0.040	0.011	0.040
		MSE	0.020	0.038	0.013	0.041	0.013	0.040
	0.25	Bias	-0.114	-0.007	-0.002	-0.020	-0.031	-0.016
		Var	0.007	0.038	0.015	0.041	0.012	0.040
		MSE	0.020	0.038	0.015	0.042	0.013	0.041
	0.3	Bias	-0.114	-0.007	0.030	-0.024	-0.014	-0.017
		Var	0.007	0.038	0.018	0.042	0.013	0.041
		MSE	0.020	0.038	0.019	0.043	0.013	0.041
	0.35	Bias	-0.114	-0.007	0.067	-0.029	0.002	-0.019
		Var	0.007	0.038	0.023	0.044	0.014	0.041
		MSE	0.020	0.038	0.027	0.045	0.014	0.042
	0.4	Bias	-0.114	-0.007	0.113	-0.035	0.017	-0.020
		Var	0.007	0.038	0.031	0.046	0.015	0.042
		MSE	0.020	0.038	0.044	0.047	0.015	0.042

Table 5: Results for the case of misspecification of the measurement error variance in Setting 2.

			Naive		MY		SIMEX	
Variance in simulations	Variance in estimation		β_1	β_2	β_1	β_2	β_1	β_2
0.1	0.05	Bias	-0.104	-0.006	-0.039	-0.014	-0.048	-0.012
		Var	0.017	0.060	0.022	0.062	0.021	0.061
		MSE	0.028	0.060	0.023	0.062	0.023	0.061
	0.1	Bias	-0.104	-0.006	0.041	-0.024	0.010	-0.018
		Var	0.017	0.060	0.030	0.065	0.025	0.063
		MSE	0.028	0.060	0.032	0.066	0.025	0.064
	0.15	Bias	-0.104	-0.006	0.145	-0.038	0.066	-0.024
		Var	0.017	0.060	0.046	0.071	0.029	0.066
		MSE	0.028	0.060	0.067	0.072	0.034	0.066
	0.2	Bias	-0.104	-0.006	0.284	-0.055	0.120	-0.030
		Var	0.017	0.060	0.080	0.078	0.034	0.069
		MSE	0.028	0.060	0.161	0.081	0.048	0.069
0.3	0.2	Bias	-0.269	0.011	-0.068	-0.016	-0.126	-0.006
		Var	0.013	0.060	0.034	0.070	0.023	0.065
		MSE	0.086	0.060	0.039	0.070	0.039	0.066
	0.25	Bias	-0.269	0.011	0.008	-0.027	-0.092	-0.010
		Var	0.013	0.060	0.047	0.075	0.026	0.067
		MSE	0.086	0.060	0.047	0.076	0.034	0.067
	0.3	Bias	-0.269	0.011	0.107	-0.040	-0.058	-0.014
		Var	0.013	0.060	0.074	0.083	0.028	0.069
		MSE	0.086	0.060	0.086	0.085	0.031	0.069
	0.35	Bias	-0.290	0.012	0.180	-0.052	-0.061	-0.013
		Var	0.010	0.059	0.085	0.096	0.022	0.070
		MSE	0.095	0.059	0.117	0.099	0.026	0.070
	0.4	Bias	-0.310	0.011	0.211	-0.053	-0.068	-0.014
		Var	0.009	0.057	0.191	0.104	0.018	0.066
		MSE	0.105	0.057	0.236	0.107	0.023	0.066

a larger variance, we should always correct with SIMEX. MY can also be useful (although not as much as SIMEX), but only if the variance is underspecified.

If we are mainly interested in decreasing the bias, then the recommendations are as follows. For β_2, it is better not to correct, but the correction method which decreases most the bias is SIMEX. For β_1, when the true variance is low and is overspecified, we should not correct with MY; with SIMEX, we should correct if the variance is not too overspecified. When the variance is low and underspecified, we should always use a correction method, the best one being MY. When the true variance is larger, we always benefit from correcting. The best correction method is SIMEX when the variance is overspecified, and MY when the variance is underspecified.

However, in practice, we do not necessarily know whether the measurement error variance is rather over- or underspecified. In order to take that fact into account, we can express the recommendations as follows. The simulation results showed that we can lower the MSE or the bias by deliberately assuming a "low" or "high" value for the measurement error variance (even if we knew its true value). We first focus on the MSE. For β_1 and β_2, with MY, we should underspecify the variance, even if we know its true value. With SIMEX, there is no clear conclusion for β_1, while, for β_2, we could underspecify the variance, but the effect is very small (since SIMEX is robust for the MSE in β_2).

Similar recommendations can be made for the bias. With MY, for β_1 and β_2, we should always underspecify the variance (but preferably slightly for β_1), even if we know its true value. This, again, can be linked to the tendency of this method to over-correct for the measurement error. With SIMEX, for β_1, we should correctly estimate the variance when it is low (if it is not well estimated, it should rather be underspecified); when the variance is large, we should generally overspecify the variance. For β_2 with SIMEX, we should underspecify the variance when it is low; when the variance is high, we should also underspecify it, since the effect of overspecifying it is not clear (but the bias does not change a lot with the assumed variance).

3.3. Implementation of the different methods in the R package miCoPTCM

We implemented both the corrected score approach of Ma and Yin (2008) and the SIMEX method adapted to the promotion time cure model in Bertrand et al. (2015) using the open-source statistical language R (R Core Team, 2015). These implementations can be found in the package `miCoPTCM` (for **mi**smeasured **Co**variates in the **P**romotion **T**ime **C**ure **M**odel), available on the CRAN website.

This package consists of two functions. The first one, `PTCMestimBF`, is the R implementation of the corrected score approach of Ma and Yin (2008). In addition to the data and the formula of the model, it only requires the variance-covariance matrix of the measurement error, and a set of initial values for the regression parameters $(\beta_0, \boldsymbol{\beta}^T)^T$. This function also allows one to estimate a model without measurement error (or without taking the measurement error into account), thanks to the backfitting procedure presented in Ma and Yin (2008), by simply setting all the elements of the variance-covariance matrix to 0.

The second function is `PTCMestimSIMEX`. It allows to estimate a promotion time cure model with mismeasured covariates by using the SIMEX approach (Bertrand et al., 2015). The user can tune the parameters related to the SIMEX algorithm: not only the variance-covariance matrix of the measurement error and initial values, but also the grid of levels for the variance of the artificial noise (λ), the number of replicates for each of these values (B) and the order of the extrapolation function. Moreover, the distribution of the error to be used can also be chosen, among the four distributions that are considered in this paper: Gaussian, Student-t, chi-square and Laplace.

Both functions return the estimated coefficients and their standard errors and the estimated baseline cumulative distribution function F.

4. Analysis of recurrence in rectal cancer patients

We would like to study the impact of several risk factors on the time until recurrence for patients suffering from rectal cancer and having been operated. The data we are going to analyze consists of 224 patients with complete data, without confirmed metastasis at diagnosis and without synchronous metastasis indicated at follow-up. They were followed between 1998 and 2014 at the Cliniques universitaires Saint-Luc (Brussels, Belgium). In such a context, we know that some patients will never experience a recurrence: they can be considered as cured. Moreover, one of the covariates is known to be measured with some error: the hemoglobin level. We want to take it into account, in case it would not be negligible. However, the distribution of this error is not known for our data. In addition to the hemoglobin level (median 13.45, range 5.5-17.7, mean 13.30, standard deviation 1.92), we want to include in the analysis three other characteristics of the patients: the ratio of infiltrated versus examined nodes (median 0, range $[0, 1]$), the presence of lymphatic permation (20% yes) and of peri-nervous permation (11% yes).

85% of the patients considered for the analysis are still alive at the end of the follow-up (median follow-up time: 3.9 years). We see a plateau in the Kaplan-Meier estimated survival curve represented on Figure 2, which confirms the medical assumption of the presence of cured patients.

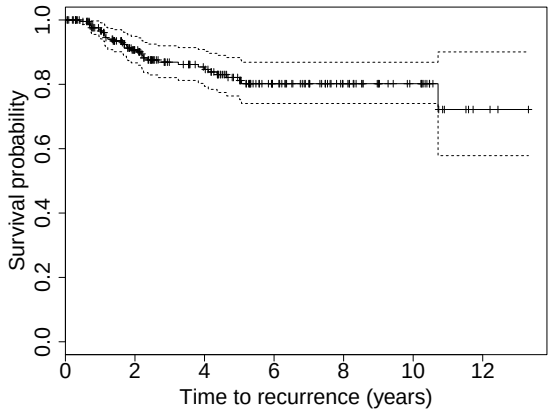


Figure 2: Kaplan-Meier estimate of the distribution of the time to recurrence for the patients from the rectal cancer database.

As we know neither the distribution nor the variance of the measurement error in the hemoglobin level, we estimate several models, assuming different values for the standard error v of the noise: 0.25, 0.5 and 1. Table 6 presents the results.

We observe that correcting for the measurement error yields larger estimated (negative) effects of the hemoglobin level. Moreover, the larger the assumed error variance, the larger (in absolute value) the estimated coefficient. MY and SIMEX yield very similar results for the two lowest values of the error variance (0.25^2 and 0.5^2); for the third one (the unit variance), the correction is larger with MY than with SIMEX.

Using SIMEX, if we assume Gaussian errors and a value of 0.25 for the measurement error standard deviation, we obtain the following results. The coefficient of the hemoglobin level is negative: a larger value of this covariate is associated with a better survival, and a larger probability of being cured. The situation is reversed for the ratio of infiltrated versus examined nodes: an increase in this variable yields worse survival and cure probability. However, the effects of the hemoglobin level and of the presence of lympathic permation are not significant (at a level 5%). For a patient with a median value of hemoglobin level and ratio of infiltrated versus examined nodes and with no lymphatic permation nor peri-nervous permation, the estimated probability of being cured is 83%. This probability drops to 23% when both permations are present.

Table 6: Regression coefficient estimates and estimated standard deviations in the case of Gaussian errors.

Estimate	Intercept	Hemoglobin	Nodes ratio	Lymphatic permation	Peri-nervous permation
Naive	0.25844	-0.14181	2.61933	0.73940	1.31332
(Estim. s.d.)	(1.18776)	(0.08741)	(0.61822)	(0.43467)	(0.47107)
Ma&Yin ($v = 1$)	0.90584	-0.18994	2.61309	0.73209	1.30976
(Estim. s.d.)	(1.47122)	(0.11054)	(0.61317)	(0.44011)	(0.47887)
Ma&Yin ($v = 0.5$)	0.39199	-0.15174	2.61770	0.73768	1.31228
(Estim. s.d.)	(1.24936)	(0.09247)	(0.61714)	(0.43566)	(0.47256)
Ma&Yin ($v = 0.25$)	0.29036	-0.14418	2.61893	0.73898	1.31305
(Estim. s.d.)	(1.20263)	(0.08860)	(0.61796)	(0.43490)	(0.47142)
SIMEX ($v = 1$)	0.64377	-0.17021	2.61246	0.73376	1.31348
(Estim. s.d.)	(1.27416)	(0.09529)	(0.61488)	(0.43481)	(0.47184)
SIMEX ($v = 0.5$)	0.40209	-0.15246	2.62173	0.73507	1.30969
(Estim. s.d.)	(1.19241)	(0.08820)	(0.61720)	(0.43549)	(0.47246)
SIMEX ($v = 0.25$)	0.29904	-0.14483	2.62136	0.73796	1.31197
(Estim. s.d.)	(1.18786)	(0.08752)	(0.61790)	(0.43497)	(0.47158)

However, the measurement error could be non-Gaussian, and its standard deviation could be different from 0.25. From our simulation results, we know that, if we are interested in the effect of hemoglobin and if we want to decrease the bias in its estimation, we should use a correction method. The estimate obtained with Naive, -0.1418, is probably somewhat too small (in absolute value). If we assume that the measurement error is Gaussian and that its variance is not too large (less than 0.6 or 0.8, according to the simulation results), then SIMEX is expected to yield the best results, even if the variance is (slightly) overspecified. In that case, the estimated effect of the hemoglobin level is -0.1448 or -0.1525. However, if we suspect that the variance could be underspecified, MY should be preferred. In this example, both strategies lead to very similar results. If the variance is larger (in this example, 1), then MY should decrease the bias the most, if we tend to underspecify the variance. The estimated effect of -0.1899 is hence expected to be larger (in absolute value) than the true value, unless the measurement error standard deviation is larger than 1.

On the other hand, if we prefer to decrease the MSE in the estimated coefficient of the level of hemoglobin, we should use SIMEX, whatever the error distribution, unless the error variance is really too low and overspecified (in which case, Naive should be used). For a Gaussian error, MY with an underspecified error variance should also yield useful results. If we suspect a measurement

error standard deviation around 0.5, then we would expect the effect to be between -0.1442 (MY with $v = 0.25$) and -0.1525 (SIMEX with $v = 0.5$).

In this particular study, it appears that the conclusions are very similar, whatever the method and parameters used for the estimation: the sign and significance of each coefficient are not affected.

5. Discussion

In medicine, as well as in other application areas, time-to-event data often exhibit features that prevent one from using classical statistical tools of survival analysis. The first of the two characteristics that we have considered here is the presence of cured subjects in the data, i.e. subjects who will never experience the event of interest. The other data property which is often ignored is the presence of measurement error in some of the continuous covariates. Our motivating example consists of the analysis of the impact of the hemoglobin level on the time until recurrence for rectal cancer patients having undergone surgery.

In this paper, we focused on one of the survival models taking cured subjects into account: the promotion time cure model. We first derived an expression of the asymptotic bias which is present in the traditional estimator when the measurement error is not taken into account. We illustrated the use of this expression by creating plots allowing to visually display and investigate the effect of the different parameters (among which the censoring rate and measurement error variance) on the bias. We then presented the results of an extensive simulation study investigating the performance of both existing estimation methods in real settings, when two of their underlying assumptions are not met: the correct specification of the measurement error variance, and the Gaussian distribution of this error. These results allowed us to provide practical recommendations which can help investigators to build a strategy to estimate a promotion time cure model. It appeared that the conclusions depend on the estimation method, on the criterion to be used (bias or MSE) and on the covariate of interest. For instance, the coefficient related to the covariate without measurement error is typically not impacted a lot by a deviation from the normality assumption with the naive method, so that a correction method is rarely needed when the estimation of this effect is the objective of the study. We also discovered that when using the corrected score method of Ma and Yin (2008), both the bias and the MSE can be decreased by assuming a low value of the measurement error variance (even if we actually know it). We also briefly described our R package `miCoPTCM`, which is freely available on the CRAN website, and which encompasses the implementation of all estimation methods used in this paper.

The recommendations allowed us to analyze the impact of the hemoglobin level on the time until recurrence of cancer in the rectal cancer database. However, in this application, it appeared that the conclusions do not change a lot with the method and the parameters used for estimating the coefficients: the estimated effect of the hemoglobin is always negative (a higher level of hemoglobin is associated with a longer time before recurrence, and with a higher cure probability), but never significant.

Of course, the conclusions and recommendations given in this paper should always be considered with care, as, for practical reasons, only a limited number of simulation settings were considered. The results may not always be generalized; however, an explanation was provided for some observed phenomena: they allow the reader to understand the mechanism at work, and, as a result, to infer the possible consequences in a different setting.

An interesting further step in the analysis of time-to-event data with a cured fraction and mismeasured covariates would be to implement a way of estimating the measurement error variance

directly in the model estimation procedure. This is currently not possible, except when we have repeated measurements (Carroll et al., 2006) of the covariates, in which case approaches such as joint modelling can be used. For example, Crowther et al. (2013) consider a longitudinal linear mixed effects submodel for the covariate with error (with its true value as explanatory variable), and a survival submodel depending on the true value of the covariate. However, in many studies, such as the one considered in this work, repeated measures are not available. The estimation of the measurement error variance would hence make the use of promotion time cure models with mismeasured covariates much easier in practice. This extension is under development (for some specific choices of the error distribution), but out of the scope of this paper.

Acknowledgements

The authors thank Professor Alex Kartheuser (Colorectal Surgery Unit, Department of Abdominal Surgery and Transplantation, St-Luc University Hospital, Université catholique de Louvain, Brussels, Belgium) and Aline Van Maanen (Institut Roi Albert II, St-Luc University Hospital, Brussels, Belgium) for providing the data.

Aurélie Bertrand, Catherine Legrand and Ingrid Van Keilegom were supported by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy) and by the contract "Projet d'Actions de Recherche Concertées" (ARC) 11/16-039 of the "Communauté française de Belgique" (granted by the "Académie universitaire Louvain").

Appendix

We proceed similarly to Hughes (1993). We first compute $f_T(t)$ and $f_{\boldsymbol{W}|T}(\boldsymbol{w}|t)$. We easily find

$$f_T(t) = \int f_{T,\boldsymbol{X}}(t,\boldsymbol{x})d\boldsymbol{x} = \int f_{T|\boldsymbol{X}}(t|\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}$$

and, consequently,

$$\begin{aligned} f_{\boldsymbol{W}|T}(\boldsymbol{w}|t) &= \int f_{\boldsymbol{W},\boldsymbol{X}|T}(\boldsymbol{w},\boldsymbol{x}|t)d\boldsymbol{x} \\ &= \int \frac{f_{T|\boldsymbol{W},\boldsymbol{X}}(t|\boldsymbol{w},\boldsymbol{x})f_{\boldsymbol{W},\boldsymbol{X}}(\boldsymbol{w},\boldsymbol{x})}{f_T(t)}d\boldsymbol{x} \\ &= \frac{\int f_{T|\boldsymbol{X}}(t|\boldsymbol{x})f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}|\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}{\int f_{T|\boldsymbol{X}}(t|\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}. \end{aligned}$$

In these expressions, $f_{\boldsymbol{X}}(\boldsymbol{x})$ has to be chosen according to the context, while $f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}|\boldsymbol{x})$ depends on the model assumed for the error. Here, we assume that $(\boldsymbol{W}|\boldsymbol{X}=\boldsymbol{x}) \sim N(\boldsymbol{x},\boldsymbol{\Lambda})$. $f_{T|\boldsymbol{X}}(t|\boldsymbol{x})$ is found using the promotion time cure model, i.e.

$$\begin{aligned} f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) &= -\frac{d}{dt}S(t|\boldsymbol{x}) = -\frac{d}{dt}\exp\left\{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)\right\} \\ &= \eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)\exp\left\{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)\right\}. \end{aligned}$$

We can now obtain the conditional expectations in (6). The first one is

$$\begin{aligned} E_{\boldsymbol{W}|T=t}(\boldsymbol{W}) &= \int \boldsymbol{w} f_{\boldsymbol{W}|T}(\boldsymbol{w}|t)d\boldsymbol{w} \\ &= \frac{\int \left(\int \boldsymbol{w} f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}|\boldsymbol{x})d\boldsymbol{w}\right) f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}{\int f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}} \\ &= \frac{\int \boldsymbol{x} f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}{\int f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}} \\ &= \frac{\int \boldsymbol{x}\ \eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta}) e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}{\int \eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta}) e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}, \end{aligned} \tag{7}$$

where the third equality is due to the fact that $E(\boldsymbol{W}|\boldsymbol{X}) = \boldsymbol{X}$. The second conditional expectation in (6) is

$$\begin{aligned} E_{\boldsymbol{W}|T=t}\left(e^{\beta_0^*+\boldsymbol{W}^T\boldsymbol{\beta}^*}\right) &= \int e^{\beta_0^*+\boldsymbol{w}^T\boldsymbol{\beta}^*} f_{\boldsymbol{W}|T}(\boldsymbol{w}|t)d\boldsymbol{w} \\ &= e^{\beta_0^*} \frac{\int \left(\int e^{\boldsymbol{w}^T\boldsymbol{\beta}^*} f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}|\boldsymbol{x})d\boldsymbol{w}\right) f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}{\int f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}} \\ &= e^{\beta_0^*} \frac{\int e^{\boldsymbol{x}^T\boldsymbol{\beta}^* + \boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2} f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}{\int f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}} \\ &= e^{\beta_0^* + \boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2} \frac{\int e^{\boldsymbol{x}^T\boldsymbol{\beta}^*} \eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta}) e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}{\int \eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta}) e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}, \end{aligned} \tag{8}$$

where the third equality follows directly from the definition of the moment generating function of $(\boldsymbol{W}|\boldsymbol{X} = \boldsymbol{x}) \sim N(\boldsymbol{x}, \boldsymbol{\Lambda})$.

The third conditional expectation in (6) is

$$\begin{aligned} E_{\boldsymbol{W}|T=t}\left(e^{\beta_0^*+\boldsymbol{W}^T\boldsymbol{\beta}^*}\boldsymbol{W}\right) &= \int e^{\beta_0^*+\boldsymbol{w}^T\boldsymbol{\beta}^*} \boldsymbol{w} f_{\boldsymbol{W}|T}(\boldsymbol{w}|t)d\boldsymbol{w} \\ &= e^{\beta_0^*} \frac{\int \left(\int e^{\boldsymbol{w}^T\boldsymbol{\beta}^*} \boldsymbol{w} f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}|\boldsymbol{x})d\boldsymbol{w}\right) f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}{\int f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}} \\ &= e^{\beta_0^*} \frac{\int (\boldsymbol{x} + \boldsymbol{\Lambda}\boldsymbol{\beta}^*) e^{\boldsymbol{x}^T\boldsymbol{\beta}^* + \boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2} f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}{\int f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}} \\ &= e^{\beta_0^* + \boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2} \frac{\int (\boldsymbol{x} + \boldsymbol{\Lambda}\boldsymbol{\beta}^*) e^{\boldsymbol{x}^T\boldsymbol{\beta}^*} \eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta}) e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}{\int \eta(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta}) e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}}, \end{aligned} \tag{9}$$

where the third equality can be found by using the moment generating function of $(\boldsymbol{W}|\boldsymbol{X} = \boldsymbol{x}) \sim N(\boldsymbol{x}, \boldsymbol{\Lambda})$, since

$$\begin{aligned} \int e^{\boldsymbol{w}^T\boldsymbol{\beta}^*} \boldsymbol{w} f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}|\boldsymbol{x})d\boldsymbol{w} &= \frac{\partial}{\partial \boldsymbol{\beta}^*} \int e^{\boldsymbol{w}^T\boldsymbol{\beta}^*} f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}|\boldsymbol{x})d\boldsymbol{w} = \frac{\partial}{\partial \boldsymbol{\beta}^*} m_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{\beta}^*) \\ &= \frac{\partial}{\partial \boldsymbol{\beta}^*} e^{\boldsymbol{x}^T\boldsymbol{\beta}^* + \boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2} = (\boldsymbol{x} + \boldsymbol{\Lambda}\boldsymbol{\beta}^*) e^{\boldsymbol{x}^T\boldsymbol{\beta}^* + \boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}. \end{aligned}$$

The first element of the first term of (6) can now be written as

$$E_T(I(T \leq t_c)) = \int I(t \leq t_c) f_T(t) dt = \int \left(\int_0^{t_c} f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) dt \right) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}$$
$$= \int (1 - S(t_c|\boldsymbol{x})) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} = 1 - \int e^{-\eta(\beta_0 + \boldsymbol{x}^T \boldsymbol{\beta}) F_\beta(t_c)} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}.$$

Using (7), the second element of the first term of (6) becomes

$$\begin{aligned}
& E_T \left[I(T \leq t_c) E_{\boldsymbol{W}|T}(\boldsymbol{W}) \right] \\
&= \int I(t \leq t_c) \frac{\int \boldsymbol{x} f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}}{\int f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}} f_T(t) dt \\
&= \int I(t \leq t_c) \left\{ \int \boldsymbol{x} f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} \right\} dt \\
&= \int \boldsymbol{x} f_{\boldsymbol{X}}(\boldsymbol{x}) \left\{ \int_0^{t_c} f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) dt \right\} d\boldsymbol{x} \\
&= \int \boldsymbol{x} f_{\boldsymbol{X}}(\boldsymbol{x}) \left\{ 1 - S(t_c|\boldsymbol{x}) \right\} d\boldsymbol{x} \\
&= E(\boldsymbol{X}) - \int \boldsymbol{x} f_{\boldsymbol{X}}(\boldsymbol{x}) e^{-\eta(\beta_0 + \boldsymbol{x}^T \boldsymbol{\beta}) F_\beta(t_c)} d\boldsymbol{x} \\
&= - \int \boldsymbol{x} f_{\boldsymbol{X}}(\boldsymbol{x}) e^{-\eta(\beta_0 + \boldsymbol{x}^T \boldsymbol{\beta}) F_\beta(t_c)} d\boldsymbol{x}.
\end{aligned}$$

Thanks to (8), the first element of the second term in (6) can be rewritten as

$$\begin{aligned}
& E_T \left[I(T \leq t_c) F_{\beta^*}(T) E_{\boldsymbol{W}|T} \left(e^{\beta_0^* + \boldsymbol{W}^T \boldsymbol{\beta}^*} \right) \right] \\
&= e^{\beta_0^* + \boldsymbol{\beta}^{*T} \boldsymbol{\Lambda} \boldsymbol{\beta}^*/2} E_T \left[I(T \leq t_c) F_{\beta^*}(T) \frac{\int e^{\boldsymbol{x}^T \boldsymbol{\beta}^*} f_{T|\boldsymbol{X}}(T|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}}{\int f_{T|\boldsymbol{X}}(T|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}} \right] \\
&= e^{\beta_0^* + \boldsymbol{\beta}^{*T} \boldsymbol{\Lambda} \boldsymbol{\beta}^*/2} \int_0^{t_c} F_{\beta^*}(t) \frac{\int e^{\boldsymbol{x}^T \boldsymbol{\beta}^*} f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}}{\int f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}} f_T(t) dt \\
&= e^{\beta_0^* + \boldsymbol{\beta}^{*T} \boldsymbol{\Lambda} \boldsymbol{\beta}^*/2} \int e^{\boldsymbol{x}^T \boldsymbol{\beta}^*} f_{\boldsymbol{X}}(\boldsymbol{x}) \left\{ \int_0^{t_c} F_{\beta^*}(t) f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) dt \right\} d\boldsymbol{x} \\
&= e^{\beta_0^* + \boldsymbol{\beta}^{*T} \boldsymbol{\Lambda} \boldsymbol{\beta}^*/2} \int e^{\boldsymbol{x}^T \boldsymbol{\beta}^*} f_{\boldsymbol{X}}(\boldsymbol{x}) \left\{ \int_0^{t_c} F_{\beta^*}(t) \eta(\beta_0 + \boldsymbol{x}^T \boldsymbol{\beta}) F'_\beta(t) e^{-\eta(\beta_0 + \boldsymbol{x}^T \boldsymbol{\beta}) F_\beta(t)} dt \right\} d\boldsymbol{x}.
\end{aligned}$$

Equation (9) allows us to express the second element of the second term of (6) as

$$\begin{aligned}
& E_T \left[I(T \leq t_c) F_{\beta^*}(T) E_{\boldsymbol{W}|T} \left(e^{\beta_0^* + \boldsymbol{W}^T \boldsymbol{\beta}^*} \boldsymbol{W} \right) \right] \\
&= e^{\beta_0^* + \boldsymbol{\beta}^{*T} \boldsymbol{\Lambda} \boldsymbol{\beta}^*/2} E_T \left[I(T \leq t_c) F_{\beta^*}(T) \frac{\int (\boldsymbol{x} + \boldsymbol{\Lambda} \boldsymbol{\beta}^*) e^{\boldsymbol{x}^T \boldsymbol{\beta}^*} f_{T|\boldsymbol{X}}(T|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}}{\int f_{T|\boldsymbol{X}}(T|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}} \right] \\
&= e^{\beta_0^* + \boldsymbol{\beta}^{*T} \boldsymbol{\Lambda} \boldsymbol{\beta}^*/2} \int (\boldsymbol{x} + \boldsymbol{\Lambda} \boldsymbol{\beta}^*) e^{\boldsymbol{x}^T \boldsymbol{\beta}^*} f_{\boldsymbol{X}}(\boldsymbol{x}) \left\{ \int_0^{t_c} F_{\beta^*}(t) f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) dt \right\} d\boldsymbol{x} \\
&= e^{\beta_0^* + \boldsymbol{\beta}^{*T} \boldsymbol{\Lambda} \boldsymbol{\beta}^*/2} \int (\boldsymbol{x} + \boldsymbol{\Lambda} \boldsymbol{\beta}^*) e^{\boldsymbol{x}^T \boldsymbol{\beta}^*} f_{\boldsymbol{X}}(\boldsymbol{x}) \left\{ \int_0^{t_c} F_{\beta^*}(t) \eta(\beta_0 + \boldsymbol{x}^T \boldsymbol{\beta}) F'_\beta(t) e^{-\eta(\beta_0 + \boldsymbol{x}^T \boldsymbol{\beta}) F_\beta(t)} dt \right\} d\boldsymbol{x}
\end{aligned}$$

and the elements of the third term of (6) as

$$\begin{aligned}
&E_T\left[I(T>t_c)F_{\beta^*}(t_c)E_{\boldsymbol{W}|T}\left(e^{\beta_0^*+\boldsymbol{W}^T\boldsymbol{\beta}^*}\right)\right]\\
&= e^{\beta_0^*+\boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}F_{\beta^*}(t_c)\int_{t_c}^{\infty}\int e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{T|\boldsymbol{X}}(t|\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}\ dt\\
&= e^{\beta_0^*+\boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}F_{\beta^*}(t_c)\int e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})S(t_c|\boldsymbol{x})\ d\boldsymbol{x}\\
&= e^{\beta_0^*+\boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}F_{\beta^*}(t_c)\int e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}\ d\boldsymbol{x}
\end{aligned}$$

and

$$\begin{aligned}
&E_T\left[I(T>t_c)F_{\beta^*}(t_c)E_{\boldsymbol{W}|T}\left(e^{\beta_0^*+\boldsymbol{W}^T\boldsymbol{\beta}^*}\boldsymbol{W}\right)\right]\\
&= e^{\beta_0^*+\boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}F_{\beta^*}(t_c)\int_{t_c}^{\infty}\int(\boldsymbol{x}+\boldsymbol{\Lambda}\boldsymbol{\beta}^*)e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{T|\boldsymbol{X}}(t|\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}\ dt\\
&= e^{\beta_0^*+\boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}F_{\beta^*}(t_c)\int(\boldsymbol{x}+\boldsymbol{\Lambda}\boldsymbol{\beta}^*)e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})S(t_c|\boldsymbol{x})\ d\boldsymbol{x}\\
&= e^{\beta_0^*+\boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}F_{\beta^*}(t_c)\int(\boldsymbol{x}+\boldsymbol{\Lambda}\boldsymbol{\beta}^*)e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}\ d\boldsymbol{x}.
\end{aligned}$$

The final expression of (6) is then

$$\begin{pmatrix} 1-\int f_{\boldsymbol{X}}(\boldsymbol{x})e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}d\boldsymbol{x} \\ -\int \boldsymbol{x}f_{\boldsymbol{X}}(\boldsymbol{x})e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}d\boldsymbol{x} \end{pmatrix}$$
$$-\begin{pmatrix} e^{\beta_0^*+\boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}\int e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})\left\{\int_0^{t_c}F_{\beta^*}(t)\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)}dt\right\}\ d\boldsymbol{x} \\ e^{\beta_0^*+\boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}\int(\boldsymbol{x}+\boldsymbol{\Lambda}\boldsymbol{\beta}^*)e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})\left\{\int_0^{t_c}F_{\beta^*}(t)\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)}dt\right\}\ d\boldsymbol{x} \end{pmatrix}$$
$$-\begin{pmatrix} e^{\beta_0^*+\boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}F_{\beta^*}(t_c)\int e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}\ d\boldsymbol{x} \\ e^{\beta_0^*+\boldsymbol{\beta}^{*T}\boldsymbol{\Lambda}\boldsymbol{\beta}^*/2}F_{\beta^*}(t_c)\int(\boldsymbol{x}+\boldsymbol{\Lambda}\boldsymbol{\beta}^*)e^{\boldsymbol{x}^T\boldsymbol{\beta}^*}f_{\boldsymbol{X}}(\boldsymbol{x})e^{-\eta(\beta_0+\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}\ d\boldsymbol{x} \end{pmatrix} = \mathbf{0}.$$

References

Bertrand, A., Legrand, C., Carroll, R. J., De Meester, C., Van Keilegom, I., 2015. Inference in a survival cure model with mismeasured covariates using a SIMEX approach. Biometrika (under revision).

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M., 2006. Measurement Error in Nonlinear Models: A Modern Perspective. 2nd edition. Chapman and Hall/CRC, Boca Raton.

Cook, J. R., Stefanski, L. A., 1994. Simulation-extrapolation in parametric measurement error models. Journal of the American Statistical Association 89, 1314–1328.

Cox, D. R., 1972. Regression models and life-tables. Journal of the Royal Statistical Society, Series B 34, 187–220.

Crowther, M. J., Lambert, P. C., Abrams, K. R., 2013. Adjusting for measurement error in baseline prognostic biomarkers included in a time-to-event analysis: a joint modelling approach. BMC Medical Research Methodology 13, 1–8.

Hughes, M. D., 1993. Regression dilution in the proportional hazards model. Biometrics 49, 1056–1066.

Li, Y., Lin, X., 2000. Covariate measurement errors in frailty models for clustered survival data. Biometrika 87, 849–866.

Ma, Y., Yin, G., 2008. Cure rate model with mismeasured covariates under transformation. Journal of the American Statistical Association 103, 743–756.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `https://www.R-project.org/`

Taylor, J. M. G., 1995. Semi-parametric estimation in failure time mixture models. Biometrics 51, 899–907.

Tsodikov, A., 1998. A proportional hazards model taking account of long-term survivors. Biometrics 54, 1508–1516.

Zeng, D., Yin, G., Ibrahim, J. G., 2006. Semiparametric transformation models for survival data with a cure fraction. Journal of the American Statistical Association 101, 670–684.