Estimation procedure allowing for mismeasurements of the covariate in Cox model with grouped data: a case-study of the risk of leukemia death for Hiroshima A-bomb survivors.

S. Huet and M. L. Martin
INRA, Laboratoire de Biométrie
78352 Jouy-en-Josas
e-mail: huet@banian.jouy.inra.fr
mlm@jouy.inra.fr

Abstract

If the risk estimation procedure does not take the mismeasurement of covariates into account, then the risk estimator is biased. We propose a new method of estimating the instantaneous risk in the Cox model with grouped data where the mismeasurement of a categorised covariate is modelled as a misclassification. An easily implemented estimation procedure based on a generalisation of the least squares method is devised then the statistical properties of the estimators are studied. The procedure is applied in estimating the risk of leukemia death in Hiroshima as a function of the dose of radiation. Our simulation study assesses the properties of the estimators. It also shows that the estimation of the instantaneous risk is altered in a very significant way when mismeasurement is not taken into account.

MSC 2000 subject classifications: 62N01, 62N02, 62F40
Key words and phrases: Cox model, Measurement error, Grouped survival data, Censoring, Least squares method.

1 Introduction

Data from the cohort of the Research Radiation Effect Foundation (RERF) lifespan study have been used to investigate the relationship between the radiation exposure and health effects, for example the relationship between the received dose of radiation and the risk of death by cancer. The received dose per survivor of the cohort is calculated with respect to geographical location and shielding conditions. This information includes uncertainties; therefore the error range between the calculated dose and the received dose is sizeable. Taking account of the error in statistical modelling modifies parameter estimations, and could therefore change conclusions about radiation exposure on health. The objective of this paper is to devise a procedure to estimate the instantaneous risk in the Cox model taking the error on the received dose into account. Because available RERF data are grouped survival data, i.e. they are aggregated and stratified by follow-up intervals and dose intervals, our method is developed for grouped data.

Much work has already been done on evaluating the effects of the error on analyses of RERF data. Assuming that the density function of the calculated dose conditional on the received dose is a known parametric, Gilbert(1983) evaluated the bias in the estimated coefficients in linear regression and Poisson regression models. He compared the behaviour of this bias in different data structures (grouped or individual data, ...) and discussed how the error alters the results of the analyses of RERF data. Pierce *et al.* (1990) modelled the distribution of the received dose using a Weibull density, and focussed on the linear model with individual data to estimate cancer risk. Prentice (1982) estimated the incidence of thyroid cancer by modelling Hiroshima and Nagasaki data using a Cox model and assuming that the distribution of the received dose conditional on the calculated dose was Gaussian.

A major weakness of these methods is found in modelling either calculated dose distribution conditional on the received dose, or the received dose distribution conditional on the calculated dose, by a known function. Like Gong *et al.* (1990) have done with categorised covariates, we have considered the error as a misclassification in dose intervals and estimated simultaneously the misclassification probabilities and the risk from the same data. Although they treated data as grouped data to decrease computation time, their method applies if one disposes of individual data which are necessary in some steps of the calculation procedure.

In this paper we propose a procedure for grouped data to estimate the instantaneous risk in the Cox model and the misclassification probabilities of the covariate. A natural idea would be to define a new estimation criterion by replacing the true covariate functions with their expectation conditional on the calculated covariate within the likelihood. Unfortunately maximising this new criterion does not give estimators with good asymptotic properties. Therefore we have devised a three step estimation procedure based on a generalisation of the least squares method.

We model the instantaneous risk by a constant piecewise function but our procedure lends itself as well to other models such as linear or linear- quadratic models which are usually considered. In the following section we present the statistical model and in section 3 the estimation method. Section 4 illustrates the procedure by estimating the instantaneous risk of leukemia death in Hiroshima and a simulation study is performed.

2 Statistical model

Let T be the failure time and C the censoring time. X is the minimum of T and C and V equals 1 if $T \leq C$ and 0 if $T > C$. Let Z be the set of covariates. To simplify the notations, we assume that Z is one dimensional. In the proportional hazard model the instantaneous risk is written:

$$R(t, Z) = \lambda(t) \exp(\theta Z),$$

where $\lambda(t)$, the baseline hazard function, is the conditional survival probability under standard conditions (a received dose of zero Gray in our example) and $\exp(\theta Z)$ is the relative risk, i.e. the risk associated with the value Z of the covariate and relative to the risk under the standard conditions.

In grouped survival data the time scale is divided into J follow-up intervals $\Delta_j = [I_{j-1}, I_j)$ and Z takes its values in S intervals D_s. The interval D_1 represents the standard condition set. We suppose that the baseline risk $\lambda(t)$ is constant within each follow-up interval Δ_j and the relative risk $\exp(\theta Z)$ is constant on each interval D_s. The instantaneous risk is thus defined by:

$$R(t \in \Delta_j, Z \in D_s) = \lambda_j \exp(\theta_s), \quad \text{with } \theta_1 = 0. \tag{1}$$

Under these assumptions the instantaneous risk R is known up to a finite number of unknown parameters $\lambda_1, ..., \lambda_J, \theta_2, ..., \theta_S$. If conditionally on Z, the failure time T and the censoring time C are independent, then maximising the log-likelihood is equivalent to maximising

$$\mathcal{J}_n(\lambda, \theta) = \sum_{s=1}^{S} \sum_{j=1}^{J} N_{sj} \log \{\lambda_j \exp(\theta_s)\} - U_{sj} \lambda_j \exp(\theta_s), \tag{2}$$

where N_{sj} is the number of uncensored individual in the interval D_s that fail in Δ_j and U_{sj} is the person-years at risk just prior to I_j. More precisely, let $\mathbf{1}_{\{x \in A\}}$ be the indicator function ($\mathbf{1}_{\{x \in A\}}$ is equal to 1 if x belongs to the set A and 0 otherwise), then

$$N_{sj} = \sum_{i=1}^{n} \mathbf{1}_{\{X_i \in \Delta_j, V_i = 1, Z_i \in D_s\}}, \tag{3}$$

$$U_{sj} = \sum_{i=1}^{n} \left\{ (X_i - I_{j-1}) \mathbf{1}_{\{X_i \in \Delta_j\}} + \delta_j \mathbf{1}_{\{X_i \geq I_j\}} \right\} \mathbf{1}_{\{Z_i \in D_s\}}, \tag{4}$$

where n is the number of observations and δ_j denotes the length of Δ_j. In case of mismeasurement of the covariate Z, we observe Z' in place of the "true" covariate Z, and the statistics N_{sj} and U_{sj} are not available. Instead of observing N_{sj} and U_{sj}, we observe N'_{sj} and U'_{sj} which are analogous to N_{sj} and U_{sj} with Z_i replaced by Z'_i. In this case, to define an estimator of the instantaneous risk, we have both to model the distribution of Z conditional on Z', and to define a new estimation criterion.

To model the distribution of Z conditional on Z', we suppose that the mismeasurement of the covariate Z is a misclassification in the intervals D_s for $s = 1, ..., S$. The mismeasurement is thus characterised by a $S \times S$ misclassification matrix $P = (p_{su})$, where $p_{su} = P(Z \in D_s | Z' \in D_u)$ is the probability that Z is in D_s given that Z' is observed in D_u. Note that by definition, the probabilities p_{su} verify $\sum_{s=1}^{S} p_{su} = 1$ for $u = 1, ..., S$. When the covariate is a qualitative variate (the stage of illness as in the example treated by Gong *et al.* (1990)) or is categorised into intervals, this modelling of the mismeasurement is natural and general.

To define an estimation criterion, one could adapt the criterion $\mathcal{J}_n$ by simply replacing the unobserved quantities N_{sj} and U_{sj}, by N'_{sj} and U'_{sj}, neglecting so the mismeasurement of the Z. It is well known that the estimators of λ and θ so defined do not converge to the true values of the parameters (λ, θ) when the number of observed individuals increases, and as it is shown by our simulation study, the bias for estimating θ is high. Another simple idea is to replace the quantities $\mathbf{1}_{\{Z_i \in D_s\}}$ by their expectation conditional on the observed covariate Z'. In the linear regression model, this leads to unbiased estimators of the parameters. Unfortunately, this idea does not work in our case, even asymptotically, i.e. when the number of observed individuals increases. Gong *et al.* (1990) proposed an EM algorithm to estimate the parameters λ, θ and P by maximising the likelihood of the observed data conditional on the observed covariate Z'. As it was already noticed in the introduction, their method needs the failure or the censoring time for each individual to be available. Therefore we propose an estimation method based on a generalisation of the least squares criterion that is similar to the quasi-likelihood method. We show in the following section that this method gives us estimators with good asymptotic properties and, is easily implemented.

3 Estimation of the instantaneous risk of failure.

This section is devoted to the estimation method. In the first part we present the estimating equations obtained by modifying the derivatives of the least squares criterion. Because these equations depend on the unknown distribution of the censoring time, and not only on λ, θ and P, we propose an estimation procedure in several steps, described in the second part.

3.1 The estimation method

Let us first introduce the least squares estimator of the vector of parameters γ where γ denotes the vector with L components $\{\lambda_j, j = 1, \ldots J\}$, $\{\theta_s, s = 2, \ldots S\}$ and $\{p_{su}, u = 1, \ldots S, s = 1, \ldots S-1\}$. Let ν_{uj} and μ_{uj} be the expectations conditional on Z' of N'_{uj} and U'_{uj}. The ν_{uj} and μ_{uj} are functions of γ and of the distribution of the censoring time C. Let us assume in this part that the distribution of C is known, then the least squares estimator of γ minimises the following sum of squares:

$$\sum_{j=1}^{J}\sum_{u=1}^{S}\left\{N'_{uj} - \nu_{uj}(\gamma)\right\}^2 + \left\{U'_{uj} - \mu_{uj}(\gamma)\right\}^2.$$

Under some classical regularity assumptions, the least squares estimator is solution of the following estimating equations corresponding to the derivatives of the sum of squares :

$$\sum_{j=1}^{J}\sum_{u=1}^{S}\frac{\partial \nu_{uj}}{\partial \gamma_l}(\gamma)\left\{N'_{uj} - \nu_{uj}(\gamma)\right\} + \frac{\partial \mu_{uj}}{\partial \gamma_l}(\gamma)\left\{U'_{uj} - \mu_{uj}(\gamma)\right\} = 0, \quad \forall l = 1, ..., L \tag{5}$$

These equations (5) could be re-written as follows:

$$B(\gamma)^T\{Y - \eta(\gamma)\} = 0, \tag{6}$$

where Y is the vector of the observations with the N'_{uj} for the first $S \times J$ components and the U'_{uj} for the last $S \times J$ ones, the vector $\eta(\gamma)$ is the expectation of Y conditional on Z': $\eta(\gamma) = E(Y|Z')$. The length of these both vectors is $K = 2 \times S \times J$. The $K \times L$ matrix $B(\gamma)$ is the matrix of derivatives of $\eta(\gamma)$ with respect to γ and $B(\gamma)^T$ its transpose.

The matrix equation system (6) is generalised by introducing a weighting symmetric $K \times K$ matrix $W(\gamma)$ leading to the following estimating equations:

$$B(\gamma)^T W(\gamma)\{Y - \eta(\gamma)\} = 0. \tag{7}$$

It can be proved that the variance of the estimator $\widehat{\gamma}$ of γ defined as the solution of (7) is minimal when $W(\gamma)$ is equal to $\Omega^{-1}(\gamma)$ where $\Omega(\gamma)$ is the covariance matrix of Y conditional on Z'. This generalised least squares estimator is similar to the quasi-likelihood estimator introduced by Mc Cullagh in 1983.

Under some classical regularity assumptions (see Section 5.1 in the appendix), $\widehat{\gamma}$ is asymptotically (when the number of individuals increases) unbiased and its distribution is approximated by the distribution of a L-dimensional Gaussian variable.

At the beginning of this section we assumed that the distribution of the censoring time were known. In the next part we release this assumption and propose an estimating procedure in several steps.

3.2 Estimation method: a three-step procedure.

The method described above involves the calculation of $\eta(\gamma)$, the expectation of Y conditional on Z'. To calculate it we assume that

Ass. 1: the failure time T and the censoring time C conditional on Z and Z' are independent,

Ass. 2: the distribution of the failure time T conditional on Z and Z' does not depend on Z',

Ass. 3: the distribution of the censoring time C conditional on Z and Z' does not depend on Z.

The first two assumptions are commonly done in survival data analysis and underlie most methods dealing with mismeasurement of covariate data (e.g Prentice(1982)). The third one has already been done by Gong *et al.* (1990), and states that the censoring time C may depend on the observed covariate Z', but not on Z. If the censoring time is independent on the covariates then clearly **Ass 3.** is verified, else by assuming that the distribution may depend on Z', we have in mind both that the distribution of C depends on the covariate Z, and that the mismeasurement on Z has less influence over the distribution of the censoring time than over the distribution of the failure. This assumption simplifies the estimation procedure of γ. In fact if the distribution of the censoring time was depending on Z that is not observed, then we would have to consider the misclassification matrix to estimate the censoring time distribution. Therefore the estimation procedure of γ would thus become laborious while we are not specially interested in the distribution of the censoring time that behaves like a nuisance parameter.

The first step of our estimating procedure deals with the modelling of the instantaneous risk of censoring and its estimation. At the end of this step, the estimating equations (7) are written in the following way:

$$\widehat{B}(\gamma)^T\widehat{W}(\gamma)\{Y - \widehat{\eta}(\gamma)\} = 0, \tag{8}$$

underlining thus that unknown parameters coming from the distribution of the censoring time have been replaced by their estimators. These equations could be solved directly, but to make easier the numerical procedure we have prefered to introduce a second step. At this second step we propose a preliminary estimate of γ, say $\widehat{\gamma}^{(1)}$, defined as the solution of equations (8) where the matrix $\widehat{W}(\gamma)$ has been replaced by a matrix $W^{(1)}$ that does not depend on γ. This $\widehat{\gamma}^{(1)}$ is simply a weighted least-squares estimator of γ. The third step consists in estimating γ by solving the equations (8) after replacing $\widehat{W}(\gamma)$ by its estimator $\widehat{W}(\widehat{\gamma}^{(1)})$.

First step: estimation of the censoring distribution. The censoring time distribution is modelled using a parametric model, in the same way as the failure time distribution. We define the instantaneous risk of censoring $\widetilde{R}(t, Z')$, similarly to R defined at Equation (1), by:

$$\widetilde{R}(t \in \Delta_j, Z' \in D_u) = \widetilde{\lambda}_j \exp(\widetilde{\theta}_u), \qquad \text{with } \widetilde{\theta}_1 = 0. \tag{9}$$

$\widetilde{R}$ is known up to the vector of parameters $\xi = (\widetilde{\lambda}_1, ..., \widetilde{\lambda}_J, \widetilde{\theta}_2, \ldots \widetilde{\theta}_S)$. Thanks to **Ass. 1** and **Ass. 3**, maximising the log-likelihood in ξ is equivalent to maximising

$$\widetilde{\mathcal{J}}_n(\xi) = \sum_{u=1}^{S} \sum_{j=1}^{J} C'_{uj} \log\{\widetilde{\lambda}_j \exp(\widetilde{\theta}_u)\} - U'_{uj} \widetilde{\lambda}_j \exp(\widetilde{\theta}_u), \tag{10}$$

where C'_{uj} is the number of censored individuals in the interval D_u that fail in Δ_j. Let us denote by $\widehat{\xi}$ the estimator of ξ. The expectation of Y, depending on the censoring time distribution only through the parameters ξ, is written $\eta(\gamma, \xi)$ (see Section 5.3 in the appendix). The function $\widehat{\eta}(\gamma)$ introduced at Equation (8) is thus defined by $\widehat{\eta}(\gamma) = \eta(\gamma, \widehat{\xi})$. Similarly $\widehat{B}(\gamma) = B(\gamma, \widehat{\xi})$ is the matrix of derivatives of η with respect to γ calculated in $(\gamma, \widehat{\xi})$ and $\widehat{W}(\gamma) = \Omega^{-1}(\gamma, \widehat{\xi})$.

Second step: calculation of a preliminary estimate of γ. The objective of this step is to define an easy-to-calculate preliminary estimator of γ, $\widehat{\gamma}^{(1)}$, to replace $\widehat{W}(\gamma)$ in Equation (8) by $\widehat{W}(\widehat{\gamma}^{(1)})$, making the equations thus defined easiest to solve. $\widehat{\gamma}^{(1)}$ is a weighted least squares estimator, solution in γ of the following equations:

$$B(\gamma, \widehat{\xi}) W^{(1)} \{Y - \eta(\gamma, \widehat{\xi})\} = 0, \tag{11}$$

where the weighting matrix $W^{(1)}$ is a diagonal matrix with known coefficients. We propose to choose coefficients as the order of magnitude of the statistics N'_{uj} and U'_{uj}, for $u = 1, ..., S$ and $j = 1, ..., J$.

Third step: estimation of γ. The third step consists in solving the resulting estimating equations defined by

$$B(\gamma, \widehat{\xi})^T W(\widehat{\gamma}^{(1)}, \widehat{\xi}) \{Y - \eta(\gamma, \widehat{\xi})\} = 0, \tag{12}$$

where $W(\widehat{\gamma}^{(1)}, \widehat{\xi}) = \Omega^{-1}(\widehat{\gamma}^{(1)}, \widehat{\xi})$. Let us denote by $\widehat{\gamma}$ this estimator.

The advantage in doing this third step is to get an estimator of γ with smaller variance than $\widehat{\gamma}^{(1)}$ when the number of individuals increases. To illustrate this fact in a simple way, let us assume again for a while that ξ is known. It can be shown (see Section 5.4)

1. that the distribution of $\widehat{\gamma}^{(1)}$ can be approximated by the distribution of a Gaussian vector with expectation γ and covariance matrix $\Psi^{(1)}$ where

$$\Psi^{(1)} = \left(B^T W^{(1)} B\right)^{-1} B^T W^{(1)} \Omega W^{(1)} B \left(B^T W^{(1)} B\right)^{-1},$$

with $B = B(\gamma, \xi)$, $\Omega = \Omega(\gamma, \xi)$ and

2. that the distribution of $\widehat{\gamma}$ can be approximated by the distribution of a Gaussian vector with expectation γ and covariance matrix Ψ where

$$\Psi = \left(B^T \Omega^{-1} B\right)^{-1}.$$

Some simple matrices calculations show that the matrix Ψ is *smaller* than $\Psi^{(1)}$ in the sense that $\Psi^{(1)} - \Psi$ is a positive semi-definite matrix.

3.3 Calculating confidence intervals by bootstrap

We propose a method to calculate confidence intervals for the parameters of interest $\gamma_1, \ldots \gamma_L$.

Starting from the asymptotic theory we know that the distribution of $\widehat{\gamma}$ conditional on Z' can be approximated by the distribution of a Gaussian vector with expectation γ and covariance matrix $\widehat{\Psi}$: more precisely, $B^T \Omega^{-1} B(\widehat{\gamma} - \gamma)$ is approximated by $B^T \Omega^{-1} M(Y_c - \eta_c)$, where Y_c is the vector of the observations with the N'_{uj} for the first $S \times J$ components, C'_{uj} for the following next $S \times J$ components and the U'_{uj} for the last $S \times J$ ones, the vector $\eta_c = \eta_c(\gamma, \xi)$ is the expectation of Y_c conditional on Z' and the matrix M is depending on the derivatives of η_c with respect to ξ and on the covariance matrix of Y_c (see Section 5.5 in the appendix). The calculation of $\widehat{\Psi}$ is tedious particularly if the dimensions S and J are large. Therefore we propose to estimate the covariance of $\widehat{\gamma}$ by the bootstrap method, replacing so a difficult numerical calculation by a resampling procedure.

In our approach, the following parametric procedure is used to generate the pseudo observations $(C^\star, T^\star)$. The censoring time and failure time distributions conditional on $Z' \in D_u, u = 1, \ldots S$ (see Section 5.2 in the appendix) are calculated in $(\widehat{\gamma}, \widehat{\xi})$. Let n'_u be the total number of individuals that are observed in D_u. For u varying from 1 to S, a n'_u-sample of the pseudo variables $C^\star$ and $T^\star$ is simulated from these distributions, leading to the calculation of $X^\star = \min\{C^\star, T^\star\}$ and $V^\star = \mathbf{1}_{\{T^\star \leq C^\star\}}$. The statistics $N^\star_{uj}$, $C^\star_{uj}$ and $U^\star_{uj}$ are then calculated: they are analogous to N'_{uj}, C'_{uj} and U'_{uj} with X and V replaced by $X^\star$ and $V^\star$. Following the three estimating steps described above, the vector of parameters γ is estimated by $\widehat{\gamma}^\star$. This procedure is replicated a large number of times, say $\mathcal{B}$, to obtain a $\mathcal{B}$-sample of variables, say $\{\widehat{\gamma}^{\star,b}, b = 1, \ldots \mathcal{B}\}$ with the same distribution as $\widehat{\gamma}^\star$. The bootstrap estimate of the covariance of $\widehat{\gamma}$ is the empirical covariance of the sample $\{\widehat{\gamma}^{\star,b}, b = 1, \ldots \mathcal{B}\}$:

$$\widehat{\psi}^\star_{\mathcal{B}} = \frac{1}{\mathcal{B}} \sum_{b=1}^{\mathcal{B}} \left(\widehat{\gamma}^{\star,b} - \widehat{\gamma}^{\star,\cdot}\right)\left(\widehat{\gamma}^{\star,b} - \widehat{\gamma}^{\star,\cdot}\right)^T$$

where $\widehat{\gamma}^{\star,\cdot}$ is the empirical mean.

According to Bickel and Freedman (1981), conditionally on the observations N'_{uj}, C'_{uj} and U'_{uj}, $n\widehat{\psi}^\star - n\widehat{\psi}$ tends to zero in probability when the $n'_u, u = 1, \ldots S$ and $\mathcal{B}$ tend to infinity. We apply this result for calculating a confidence interval for each component of γ, say for example the first one γ_1. Let $\widehat{\sigma}^\star_1$ be the square-root of the bootstrap empirical variance of γ_1:

$$\widehat{\sigma}^\star_1 = \sqrt{\frac{1}{\mathcal{B}} \sum_{b=1}^{\mathcal{B}} \left(\widehat{\gamma}^{\star,b}_1 - \widehat{\gamma}^{\star,\cdot}_1\right)^2}.$$

Then the distribution of $(\widehat{\gamma}_1 - \gamma_1)/\widehat{\sigma}_1^\star$ is approximated by a standard centered Gaussian distribution. Let q_α be the α-quantile of a standard centered Gaussian distribution. The interval

$$[\widehat{\gamma}_1 - q_{1-\alpha}\widehat{\sigma}_1^\star \; ; \; \widehat{\gamma}_1 + q_{1-\alpha}\widehat{\sigma}_1^\star]$$

is a confidence interval for γ_1 with asymptotic level $1 - 2\alpha$.

4 Application

4.1 Analysis of risk of leukemia death among survivors of atomic bombing of Hiroshima

We consider data on mortality among survivors of Hiroshima atomic bombing during the period October 1, 1950 through December 31, 1985. These data are available on the web at http://www.rerf.or.jp/eigo/archives/lss11md.htm. We relate leukemia death to the received dose of radiation. A total of 165 leukemia deaths is recorded in the Hiroshima cohort of 51 390 persons. For this analysis we keep four dose intervals D_s ($S = 4$) and seven follow-up intervals Δ_j (J=7). This classification is done such that the dose intervals correspond respectively to null, weak, medium and high doses. Data are given in Table 1 and Table 2.

Dose intervals in grays	[0; 0.005)	[0.005; 0.1)	[0.1; 1)	[1; 6)
Total number	20346	16121	12678	2245

Table 1: Total number of individuals observed by dose intervals

The instantaneous risk of leukemia death is assumed to be a function of the received dose Z and is modelled as in equation (1). The received dose of radiation is unobserved, and we have only at our disposal a dose calculated from information about location and shielding conditions. We denote by Z' this dose and we name it the observed dose. The error between Z and Z' is sizeable and we take it into account in the proportional hazard model to estimate the risk of leukemia death. We model this error as a misclassification in the dose intervals by introducing a $S \times S$ misclassification matrix P. We assume that the error only leads to underestimate the received dose of one interval. Under this assumption, the dose interval $D_4 = [1; 6)$ being the last interval, the probability that the received doses are in D_4 given that the observed doses are in D_4 is one, i.e. $p_{44} = 1$. As Gilbert (1984) has already assumed with RERF data, a received dose of zero Gray is assumed to have no measurement error, hence $p_{11} = 1$. Moreover, we assume that the probability of correct classification is identical in all intervals, consequently the misclassification matrix P, reduced to a scalar p, is written :

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 1-p & p & 0 \\ 0 & 0 & 1-p & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Follow-up intervals	Observations	Dose intervals in Grays			
		[0; 0.005)	[0.005; 0.1)	[0.1; 1)	[1; 6)
(50; 54]	N'	6	1	8	13
	U'	104106.35	82565.57	65084.85	11492.75
	W'	1019	781	596	111
[55; 59]	N'	6	4	12	10
	U'	93796.38	74597.76	58775.68	10348.83
	W'	1119	833	656	130
[60; 64]	N'	2	5	8	4
	U'	88359.58	70285.20	55325.08	9681.31
	W'	1068	898	753	135
[65; 69]	N'	3	0	2	5
	U'	82847.36	65930.65	51560.88	8966.13
	W'	1155	842	737	159
[70; 74]	N'	7	6	2	4
	U'	76871.63	61646.63	47757.04	8207.15
	W'	1232	873	785	152
[75; 79]	N'	14	6	2	3
	U'	70466.15	57115.10	43763.22	7378.35
	W'	1272	932	793	156
[80; 84]	N'	6	6	0	4
	U'	64514.45	52392.91	39796.76	6657.67
	W'	1150	941	791	154

N' is the number of leukemia deaths, U' the number of person-year at risk and W' is the number of deaths.

Table 2: Data on mortality among 51390 survivors in Hiroshima.

where p is the unknown parameter to be estimated. Thus an observed dose in D_2 (resp. D_3) relates either to a received dose in the interval D_2 (resp. D_3) with probability p or to a received dose in the interval D_3 (resp. D_4) with probability $1-p$.
Note that p is by definition less than one. It is assumed greater than 0.5, so that the probability of correct classification is always greater than the probability of misclassification. Before carrying out the estimation procedure, a logistic transformation is applied on p insuring that the distribution of the estimator of p is defined only within the compact set $[0.5\ ;\ 1]$. The vector of parameters γ involved in the estimation procedure is thus $(\lambda_1, ..., \lambda_J, \theta_2, \theta_3, \theta_4, \pi)$, where $\pi = \log\{(2p-1)/(2-2p)\}$.

We apply the three-step procedure to estimate γ: in the first step the instantaneous risk of censoring is modelled as in equation (9), it is a function of the observed dose Z', known up to the parameter vector ξ which is estimated by maximising (10). In the second step $\widehat{\gamma}^{(1)}$, solution of the estimating equations (11) is calculated. The weighting matrix $W^{(1)}$ is a diagonal matrix having coefficients in the same scale-sizes as the inverse of the variance of N'_{uj} and U'_{uj} for $u = 1, ..., S$ and, $j = 1, ..., J$: the first $S \times J$ diagonal coefficients equal to 1 and the last $S \times J$ ones to 10^{-4}. In the third step, the estimating equations (12) are slightly modified: the weighting matrix $\widehat{W}(\gamma)$ is defined as the matrix constituted only by the diagonal elements of the matrix $\Omega^{-1}(\widehat{\gamma}^{(1)}, \widehat{\xi})$ since some preliminary calculations have shown that the correlations between the components of Y are nearly zero.

The criterion (10) is maximised using a Newton algorithm available in the Numerical Algorithms Group's (1988) library. The estimators $\widehat{\gamma}^{(1)}$ and $\widehat{\gamma}$ are calculated using nls2, a software implemented in the statistical system S-Plus (Huet *et al.* (1996)).

We use the bootstrap method described in Section 3.3 with $\mathcal{B} = 500$ to calculate a 95 % confidence interval of the relative risk and of the probability of correct classification p. The estimated values of the baseline risk and the relative risk of censoring and leukemia death are given Tables 3 and 4. The probability of correct classification p is estimated at 0.746 and its confidence interval with level 95 % is [0.554; 0.942]. Therefore we conclude that p is significantly different from 1.

	Follow-up intervals						
	(1950; 54]	[55; 59]	[60; 64]	[65; 69]	[70; 74]	[75; 79]	[80; 84]
$\widetilde{\lambda} \times 10^3$	9.47	11.46	12.75	13.85	15.61	17.59	18.56
$\lambda \times 10^5$	6.75	8.32	5.21	3.76	7.64	9.39	8.71

$\widetilde{\lambda}$ is the baseline risk of censoring and λ is the baseline risk of leukemia death.

Table 3: Estimation of the baseline risks.

In Table 1 we note that the percentage of censored individuals is about 5 % in each dose interval and from the results of Table 4, we note that the risk of censoring $\tilde{R}$ does not vary with the dose, therefore the censoring time does not depend on the received dose, and consequently assumption **Ass. 3** is verified in this application.

	Dose intervals in Grays		
	[0.005; 0.1)	[0.1; 1)	[1; 6)
$\widetilde{\theta}$	-0.05	0.02	0.1
$\widetilde{r}$	0.951	1.02	1.105
θ	-0.194	0.317	2.264
r	0.824	1.373	9.621
CI(r)	[0.498; 1.363]	[0.821; 2.296]	[6.238; 14.843]

$r = \exp(\theta)$ is the relative risk of leukemia death (see Equation (1)), $\widetilde{r} = \exp(\widetilde{\theta})$ is the relative risk of censoring (see Equation (9)). The last line CI(r) gives a confidence interval for r with level 95 %.

Table 4: Estimation of the relative risks.

4.2 A simulation study

We carry out a simulation study to assess the properties of $\widehat{\gamma}$ when the number of individuals is fixed and to evaluate how the estimation of the instantaneous risk is altered when the error of the covariate is not taken into account.

We run two simulation experiments of 500 trials. The first experiment shows how behaves $\widehat{\gamma}$ when the number of individuals is close to the size of the Hiroshima cohort. The second experiment is done with a number of individual close to five times as big as the size of Hiroshima cohort to illustrate the behaviour of $\widehat{\gamma}$ for increasing cohort size. The total number of individuals by interval for each experiment is given Table 5 (compare to Table 1). We keep $S = 4$ dose intervals and $J = 7$ follow-up intervals.

	n'_1	n'_2	n'_3	n'_4
Experiment 1	21000	16500	13000	2400
Experiment 2	105000	82500	65000	12000

Table 5: Total number of individuals by interval for each experiment.

In each trial, the parameter values of the baseline risks and the relative risks are equal to the values given Table 6, and π is fixed at 0 (the corresponding probability p is 0.75). Note that the parameter values are close to the estimated values given Tables 3 and 4. We choose the relative risk of censoring in this simulation study equal to zero, since **Ass. 3** is verified in the analysis of the risk of leukemia death in Hiroshima. For each interval D_u with $u \in \{1, ..., S\}$, the quantities N'_{uj}, U'_{uj} and C'_{uj} are simulated from the cumulative functions of the failure time T and the censoring time C conditional on $Z' \in D_u$ (given Section 5.2 in the appendix). Firstly, we estimate ξ and $\gamma = (\lambda_1, ..., \lambda_J, \theta_2, \theta_3, \theta_4, \pi)$ using the three-step procedure. Secondly, we estimate the parameter vectors λ and θ by maximum likelihood from the criterion $\mathcal{J}_n(\lambda, \theta)$.

From 500 trials in each experiment, we calculate the empirical mean $\widehat{\gamma}^{\cdot}$ and the relative bias (the difference between the empirical mean and the parameter value divided by the absolute value of the parameter). The results for the estimator of γ of the three-step procedure are given Table 7. We note that the biases of the parameters of the baseline risk is about -5%

	Follow-up intervals								Dose intervals		
	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7		D_2	D_3	D_4
$\widetilde{\lambda} \times 10^3$	10	11	13	14	16	18	19	$\widetilde{\theta}$	0	0	0
$\lambda \times 10^5$	7	8	5	4	8	9	9	θ	-0.2	0.3	2

$\widetilde{\lambda}$ is the baseline risk of censoring, and λ is the baseline risk of leukemia death. $\widetilde{\theta}$ is the parameter of the baseline risk of censoring, and θ is the parameter of the baseline risk of failure.

Table 6: Parameter values for the simulation study.

in the first experiment, and they are close to zero in the second experiment. The bias of $\widehat{\theta}$ is high in the first experiment, but decreases sharply in the second one. Since we are especially interested in the relative risk $r = \exp(\theta)$, we have calculated the empirical mean of the relative risk and the bias: the relative risk is underestimated in all intervals. The biases are high in the first experiment (about 8%), but they decrease sharply in the second experiment (about 2 %). In both experiments, the bias of the probability $\widehat{p}$ is small, 2.50 % in the first one and 0.19 % in the second one.

According to Section 3.2, the distribution of $\widehat{\gamma}$ can be approximated by the distribution of a L-dimensional Gaussian vector with expectation γ and covariance matrix $\widehat{\Psi}$, where $\widehat{\Psi}$ is some estimator of Ψ. To verify this result, we calculate from the second experiment the empirical skewness and kurtosis of each estimated parameter, and we perform the test of normality. For all parameters except π we accept the null hypothesis at level 2%. The distribution of $\widehat{\pi}$ appears to be slightly dissymmetric with a skewness equal to -3 and is heavily tailed with a kurtosis equal to 5.

		Follow-up intervals						
		Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7
Experiment 1	$\widehat{\lambda}^{\cdot} \times 10^5$	6.81	7.57	4.79	3.89	7.72	8.55	8.70
	% bias	-3.65	-5.36	-4.19	-2.80	-3.43	-4.97	-3.33
Experiment 2	$\widehat{\lambda}^{\cdot} \times 10^5$	6.94	7.93	4.94	3.94	7.88	8.89	8.87
	% bias	-0.90	-0.84	-1.20	-1.39	-1.54	-1.25	-1.40

$\widehat{\lambda}^{\cdot}$ is the vector of the empirical mean of the baseline risk, the percentage of bias corresponding to $100 * (\widehat{\lambda}^{\cdot} - \lambda)/|\lambda|$

		Covariate intervals		
		D_2	D_3	D_4
Experiment 1	$\widehat{\theta}^{\cdot}$	-0.15	0.35	2.07
	% bias	23.35	17.53	3.41
	$\widehat{r}^{\cdot}$	0.88	1.47	8.12
	% bias	8.10	8.64	9.91
Experiment 2	$\widehat{\theta}^{\cdot}$	-0.19	0.32	2.01
	% bias	6.38	5.42	0.69
	$\widehat{r}^{\cdot}$	0.83	1.38	7.53
	% bias	1.92	2.23	1.85

$\widehat{\theta}^{\cdot}$ is the empirical mean of θ, $\widehat{r}^{\cdot} = \widehat{\exp(\theta)}^{\cdot}$ is the empirical mean of the relative risk.

	Experiment 1	Experiment 2
$\widehat{p}^{\cdot}$	0.769	0.751
% bias	2.50	0.19

$\widehat{p}^{\cdot}$ is the empirical mean of p.

Table 7: Estimation by the three-step procedure.

Table 8 presents the estimation of the instantaneous risk by maximising the criterion $\mathcal{J}_n(\lambda, \theta)$. The biases of the baseline risks are very close to zero. In contrast, the biases of the parameters of the relative risk are very high in both experiments, especially in the intervals D_2 and D_3, where the covariate is subject to mismeasurement. This bias does not decrease even if the size of the cohort increases. Consequently, it is essential to take the error of measurement of the covariate into account to evaluate correctly the relative risk. Note that mismeasurements of the covariate do not alter the estimation of the baseline risk.

		Follow-up intervals						
		Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7
Experiment 1	$\widehat{\lambda}^{\cdot} \times 10^5$	7.11	7.91	4.95	3.99	8.08	8.96	9.09
	% bias	1.64	-1.11	-0.89	-0.10	1.01	-0.48	0.97
Experiment 2	$\widehat{\lambda}^{\cdot} \times 10^5$	6.99	7.99	4.97	3.96	7.94	8.96	8.94
	% bias	-0.09	-0.01	-0.64	-0.91	-0.77	-0.47	-0.69

$\widehat{\lambda}^{\cdot}$ is the vector of the empirical mean of the baseline risk, the percentage of bias corresponding to $100 * (\widehat{\lambda}^{\cdot} - \lambda)/|\lambda|$

		Covariate intervals		
		D_2	D_3	D_4
Experiment 1	$\widehat{\theta}$	-0.17	0.51	2.01
	% bias	**13.94**	**71.67**	0.58
	$\widehat{r}$	0.87	1.71	7.66
	% bias	5.90	27.02	3.69
Experiment 2	$\widehat{\theta}$	-0.16	0.53	2.002
	% bias	**19.00**	**77.23**	0.12
	$\widehat{r}$	0.85	1.71	7.44
	% bias	4.45	26.64	0.69

$\widehat{\theta}$ is the empirical mean of θ, $\widehat{r} = \widehat{\exp(\theta)}$ is the empirical mean of the relative risk.

Table 8: Estimation by maximum likelihood.

5 Appendix

5.1 Asymptotic properties of estimators

Let T and C be defined as in Section 2. We assume in this paragraph that the distribution of C is known, and that the instantaneous risk is modelled as in Equation (1). Moreover the mismeasurement of the covariate is modelled through the misclassification matrix P. Let γ be the vector of parameters defined in Section 3.1. The following proposition gives the asymptotic properties of $\widehat{\gamma}$, the solution of the estimating equations(7). In our case, since the covariate is categorised into intervals D_u, n tends to infinity means that the number of observations n'_u in each interval D_u tends to infinity.

Proposition 1

Let (P1) be the following set of assumptions :

- *γ is within a compact Γ of R^L.*
- *$\eta(\alpha)$ is twice continuously differentiable with respect to α on Γ.*
- *$W(\alpha)$ is any regular symmetric matrix $K \times K$ and is once continuously differentiable with respect to α on Γ.*
- *$B(\alpha) = d\eta(\alpha)/d\alpha$ is a regular matrix.*
- *Let define $\Sigma_n(\gamma) = n(B^T WB)^{-1}(B^T W)\Omega(B^T W)^T(B^T WB)^{-1}$.*
  When n'_u tends to infinity for all $u \in \{1, ..., S\}$, $\Sigma_n(\gamma)$ has a positive definite limit in $\gamma, \Sigma(\gamma)$.
- *Let define $q(\gamma, \alpha) = B(\alpha)^T W(\alpha)\{\eta(\gamma) - \eta(\alpha)\}$, $q(\gamma, \alpha) = 0$ if and only if $\alpha = \gamma$.*

Under (P1) $\widehat{\gamma}$ converges to γ and $\sqrt{n}(\widehat{\gamma} - \gamma)$ converges in distribution to a L-dimensional Gaussian with expectation zero and covariance matrix $\Sigma(\gamma)$.

The asymptotic distribution is derived from the Mc Cullagh's paper (1983) about quasi-likelihood functions and the consistency is proved by application of the theorem of consistency of minimum of contrast estimators, see Dacunha-Castelle and Duflo (1983). The used contrast function is the norm of the function $q(\gamma, \alpha)$.

5.2 Cumulative functions

The density of T conditional on $Z' \in D_u$ could be written as the marginal distribution of (T, Z) conditional on $Z' \in D_u$, therefore

$$P(T \leq t|Z' \in D_u) = \sum_{s=1}^{S} P(T \leq t, Z \in D_s|Z' \in D_u)$$

Using Bayes's formula and **Ass. 2** , we find that :

$$P(T \leq t|Z' \in D_u) = \sum_{s=1}^{S} p_{su} P(T \leq t|Z \in D_s)$$

Let $\bar{F}_{T|Z \in D_s}(t)$ denote the survivor function of T conditional on $Z \in D_s$, then

$$P(T \leq t|Z' \in D_u) = \sum_{s=1}^{S} p_{su} \left\{1 - \bar{F}_{T|Z \in D_s}(t)\right\}$$

By definition(see eg Cox and Oakes(1984) p. 14),

$$\bar{F}_{T|Z \in D_s}(t) = \exp\left\{-H(t, Z \in D_s)\right\}, \tag{13}$$

where $H(t, Z \in D_s)$ is the cumulative risk defined as $\int_0^t R(u, Z \in D_s) du$.

Introducing the relation (13) into the above equality yields :

$$P(T \leq t|Z' \in D_u) = \sum_{s=1}^{S} p_{su} \left[1 - \exp\left\{-H(t, Z \in D_s)\right\}\right]$$

If the instantaneous risk is modelled as in equation (1), then

$$H(t, Z \in D_s) = \sum_{j=1}^{J} \mathbf{1}_{\mathrm{t} \in \Delta_{\mathrm{j}}} \sum_{\mathrm{k}=1}^{\mathrm{j}-1} \{\delta_{\mathrm{k}} \lambda_{\mathrm{k}} + (t - I_{j-1})\lambda_j\} \exp(\theta_{\mathrm{s}}) \tag{14}$$

The approach is easier to define the cumulative function of the censoring time, since the censoring time depends only on the observed covariate Z' :

$$\begin{aligned} P(C \leq c|Z' \in D_u) &= 1 - P(C \geq c|Z' \in D_u) \\ &= 1 - \exp\left\{-\tilde{H}(c, Z \in D_u)\right\} \end{aligned}$$

The definition of the cumulative risk of censoring is analogous to $H(t, Z \in D_s)$,

$$\tilde{H}(t, Z' \in D_u) = \sum_{j=1}^{J} \mathbf{1}_{\mathrm{t} \in \Delta_{\mathrm{j}}} \sum_{\mathrm{k}=1}^{\mathrm{j}-1} \left\{\delta_{\mathrm{k}} \tilde{\lambda}_{\mathrm{k}} + (t - I_{j-1})\tilde{\lambda}_j\right\} \exp(\tilde{\theta}_{\mathrm{u}}) \tag{15}$$

5.3 Expectation of Y conditional on Z': $\eta(\gamma, \xi)$

We first give the density expression of (X, V) conditional on $Z' \in D_u$, and then the expectation of $\eta(\gamma, \xi)$. Let ξ be the parameter vector of the censoring risk defined in Equation (9). Let us first define some notations : $f_{Y|W=w}(y)$ is the density of Y conditional on $W = w$ at a point y.

The density $f_{X,V|Z'\in D_u}(x, v)$ could be written as the marginal distribution of the triplet (X, V, Z) conditional on $Z' \in D_u$:

$$f_{X,V|Z'\in D_u}(x, v) = \sum_{s=1}^{S} f_{X,V,Z|Z'\in D_u}(x, v, z)\mathbf{1}_{\mathrm{z}\in\mathrm{D}_s}$$

Using Bayes's formula, we obtain :

$$f_{X,V|Z'\in D_u}(x, v) = \sum_{s=1}^{S} p_{su} f_{X,V|Z\in D_s, Z'\in D_u}(x, v)$$

According to **Ass. 1, Ass. 2, Ass. 3**

$$f_{X,V|Z\in D_s,Z'\in D_u}(x, v) = \left\{f_{C|Z'\in D_u}(x)\bar{F}_{T|Z\in D_s}(x)\right\}^{1-v}\left\{f_{T|Z\in D_s}(x)\bar{F}_{C|Z'\in D_u}(x)\right\}^{v}, \quad (16)$$

By definition (see Cox and Oakes, p. 15)

$$f_{T|Z\in D_s}(t) = R(t, Z \in D_s)\exp\left\{-H(t, Z \in D_s)\right\},$$

In the same way, according to the model of the censoring time,

$$f_{C|Z'\in D_u}(t) = \tilde{R}(t, Z' \in D_u)\exp\left\{-\tilde{H}(t, Z' \in D_u)\right\}$$

Inserting these relations and the relations (14) and (15) in (16), the density of (X, V) conditional on $Z' \in D_u$ is written:

$$f_{X,V|Z'\in D_u}(x, v) = \tilde{R}(x, Z' \in D_u)^{1-v}\sum_{s=1}^{S} p_{su} R(x, Z \in D_s)^{v}\mathcal{H}(y, u, s),$$

where $\mathcal{H}(y, u, s) = \exp\left[-\left\{\tilde{H}(x, Z' \in D_u) + H(x, Z \in D_s)\right\}\right]$.

After straightforward calculations, we obtain:

$$\begin{aligned} \nu_{uj} &= n'_u \sum_{s=1}^{S} p_{su} R(t \in \Delta_j, Z \in D_s)\int_{\Delta_j}\mathcal{H}(y, u, s)dy \\ \mu_{uj} &= n'_u \sum_{s=1}^{S} p_{su}\int_{\Delta_j}\mathcal{H}(y, u, s)dy \end{aligned}$$

ν_{uj} are the first $S \times J$ components of $\eta(\gamma, \xi)$ and μ_{uj} the last $S \times J$ ones.

The expressions of $\tilde{H}(y, Z' \in D_u)$and $H(y, Z \in D_s)$ are given in section 5.1. By simply integrating we get :

$$\int_{\Delta_j}\mathcal{H}(y, u, s)dy = \frac{\exp\left\{-\Lambda(j-1, u, s)\right\} - \exp\left\{-\Lambda(j, u, s)\right\}}{\tilde{\lambda}_j\exp(\tilde{\theta}_u) + \lambda_j\exp(\theta_s)},$$

where $\Lambda(j,u,s)=\sum_{k=1}^{j}\delta_k\left\{\tilde{\lambda}_k\exp(\tilde{\theta}_u)+\lambda_k\exp(\theta_s)\right\}$ and $\Lambda(0,u,s)=0$.

The variance of Y conditional on Z' needs tedious calculations. The variance of N'_{uj} is written from the expectation of N'_{uj} : $Var(N'_{uj})=n'_u\nu_{uj}(1-\nu_{uj})$, but the conditional variance of U'_{uj} on Z' is much more complicated. A program is available on the web at http://www-bia.inra.fr/J/MLM/ENGLISH/abstract.html to calculate $\Omega(\xi,\gamma)$ from any ξ and γ.

5.4 Asymptotic variance

The distribution of C is assumed to be known. The covariance matrix $\Psi^{(1)}$ is calculated by applying the proposition 1 given in section 5.1. The expression of the covariance matrix Ψ is established using a Taylor expansion and the law of large numbers: let $\mathcal{C}_n(\gamma,\gamma^{(1)},\xi)=\frac{1}{\sqrt{n}}B(\gamma,\xi)^T\Omega^{-1}(\gamma^{(1)},\xi)\left\{Y-\eta(\gamma,\xi)\right\}$. Using a Taylor expansion of $\mathcal{C}_n(\widehat{\gamma},\widehat{\gamma}^{(1)},\xi)$ in a neighbourhood of (γ,γ,ξ) we have the following result :

$$
\begin{aligned}
0=\mathcal{C}_n(\widehat{\gamma},\widehat{\gamma}^{(1)},\xi) \quad =\mathcal{C}_n(\gamma,\gamma,\xi) \quad & +\frac{\partial\mathcal{C}_n(\gamma,\gamma,\xi)}{\partial\gamma}(\widehat{\gamma}-\gamma) \\
& +\frac{\partial\mathcal{C}_n(\gamma,\gamma,\xi)}{\partial\gamma^{(1)}}(\widehat{\gamma}^{(1)}-\gamma)+o_p(1),
\end{aligned}
$$

where $o_p(1)$ denotes a sequence of random vectors that converges to zero in probability.

By the law of large numbers, when n'_u tends to infinity for all u,

$$
\frac{1}{\sqrt{n}}\frac{\partial\mathcal{C}_n(\gamma,\gamma,\xi)}{\partial\gamma}-\frac{1}{n}B(\gamma,\xi)^T\Omega^{-1}(\gamma,\xi)B(\gamma,\xi)\to 0
$$

and

$$
\frac{1}{\sqrt{n}}\frac{\partial\mathcal{C}_n(\gamma,\gamma,\xi)}{\partial\gamma^{(1)}}\to 0
$$

We deduce then

$$
\widehat{\gamma}-\gamma=(B^T\Omega^{-1}B)^{-1}B^T\Omega^{-1}\left\{Y-\eta(\gamma,\xi)\right\}+o_p(n^{-1/2}),
$$

with $B=B(\gamma,\xi)$ and $\Omega=\Omega(\gamma,\xi)$.
The variance of Y conditional on Z' is $\Omega(\gamma,\xi)$, therefore the result follows: $\Psi=(B^T\Omega^{-1}B)^{-1}$.

5.5 Covariance matrix

Let us assume now that the distribution of C is known up to the parameter vector ξ.
Let $\mathcal{C}_n(\gamma,\gamma^{(1)},\widehat{\xi})=\frac{1}{\sqrt{n}}B(\gamma,\widehat{\xi})^T\Omega^{-1}(\gamma^{(1)},\widehat{\xi})\left\{Y-\eta(\gamma,\widehat{\xi})\right\}$. As in Section 5.4, after a Taylor expansion of $\mathcal{C}_n(\widehat{\gamma},\widehat{\gamma}^{(1)},\widehat{\xi})$ in a neighbourhood of (γ,γ,ξ) and using the law of large numbers we have the following result :

$$
0=B^T\Omega^{-1}\left\{Y-\eta(\gamma,\xi)\right\}+B^T\Omega^{-1}B(\widehat{\gamma}-\gamma)+B^T\Omega^{-1}\frac{\partial\eta(\gamma,\xi)}{\partial\xi}(\widehat{\xi}-\xi)+o_p(1), \qquad (17)
$$

with $B=B(\gamma,\xi)$ and $\Omega=\Omega(\gamma,\xi)$.
$\widehat{\xi}$ is the estimator of the parameter vector ξ calculated at the first step of the three-step procedure. Owing to maximising the criterion $\widetilde{\mathcal{J}}_n(\xi)$ is equivalent to maximising the likelihood,

$\widehat{\xi}$ is also the maximum likelihood estimator.
Let $\widetilde{Y}$ be the vector with C'_{uj} for the first $S \times J$ components and U'_{uj} for the last $S \times J$ ones, and $\widetilde{\eta}(\xi)$ denote the expectation of $\widetilde{Y}$ conditional on Z'. According to the equation (10), $\widetilde{\mathcal{J}}_n(\xi)$ is a linear function of $\widetilde{Y}$, and we have the following approximation

$$\widehat{\xi} - \xi = \frac{\partial \widetilde{\eta}}{\partial \xi}^T \widetilde{\Omega}(\xi)^{-1}(\widetilde{Y} - \widetilde{\eta}(\xi)) + o_p(n^{-1/2}) \tag{18}$$

where $\widetilde{\Omega}(\xi)$ is the variance of $\widetilde{Y}$ conditional on Z'.
Inserting (18) in the expression (17), we get :

$$\widehat{\gamma} - \gamma = (B^T \Omega^{-1} B)^{-1} B^T \Omega^{-1} M(Y_c - \eta_c) + o_p(n^{-1/2}),$$

where M is a matrix depending on the derivative of $\eta_c(\gamma, \xi)$ with respect to ξ and on the covariance matrix of Y_c conditional on Z'.

References

Bickel, P. J. and Freeman, D. A. (1981) Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9**, 1196-1217.

Cox, D.R. and Oakes, D. (1984) *Analysis of Survival Data.* London: Chapman Hall

Dacunha-Castelle, D. and Duflo, M. (1983) *Probabilités et Statistiques.Problèmes à temps mobile. Tome 2.* Paris: Masson

Gilbert, E. S. (1984) Some effects of random dose measurement errors on the analyses of atomic bomb survivor data. *Radiation Research* **98**, 591-605.

Gong, G., Whittemore, A. S. and Grosser, S. (1990) Censored Survival Data With Misclassified Covariates: A Case Study of Breast-Cancer Mortality. *Journal of the Americain Statistical Association*,**90**, 20-28.

Huet, S., Bouvier, A., Gruet, M. A. and Jolivet, E. (1996) *Statistical tools for nonlinear regression. A practical guide with S-PLUS examples.*New-York: Springer-Verlag.

McCullagh, P. (1983) Quasi-likelihood functions. *The Annals of Statistics*, **11**, 59-67.

Numerical Algorithm Group (1988) NAG Fortran Library Manual, Mark 13. Oxford: Numerical Algorithm Group.

Pierce, D. A., Stram, D. O. and Væth, M. (1990) Allowing for random errors in radiation dose estimates for the atomic bomb survivor data. *Radiation Research***123**, 275-284.

Prentice, R. L. (1982) Covariate measurement errors and parameter estimation in a failure time regression model, *Biometrika*, **69**, 331-342.