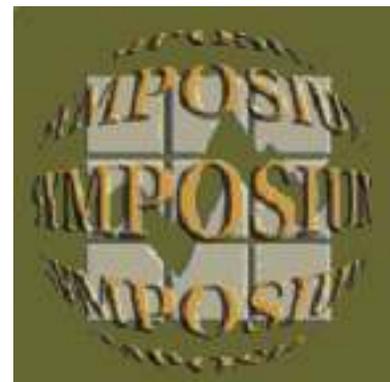


No 11-522-XIF au catalogue

**La série des symposiums internationaux
de Statistique Canada - Recueil**

**Symposium 2006 : Enjeux
méthodologiques reliés à la
mesure de la santé des
populations**



2006



Statistics
Canada

Statistique
Canada

Canada

Erreur de mesure dans les données sur le cycle de vie

Grace Y. Yi¹ et Wenqing He²

Résumé

En pratique, il arrive souvent que certaines données recueillies comportent une erreur de mesure. Parfois, des covariables (ou facteurs de risque) d'intérêt sont difficiles à observer avec précision en raison de l'emplacement physique ou du coût. D'autres fois, il est impossible de mesurer précisément les covariables à cause de leur nature. Dans d'autres situations, une covariable peut représenter la moyenne d'une certaine grandeur mesurable au cours du temps, et tout moyen pratique de mesurer cette grandeur comporte nécessairement une erreur de mesure. Lorsqu'on procède à des inférences statistiques dans de telles conditions, il est important de tenir compte des effets des covariables mesurées incorrectement; sinon, les résultats risquent d'être incorrects, voire même trompeurs. Dans le présent article, nous discutons de plusieurs exemples d'erreur de mesure survenant dans des contextes distincts. Nous accordons une attention particulière aux données sur la survie en présence de covariables sujettes à une erreur de mesure. Nous discutons d'une méthode de simulation-extrapolation pour la correction des effets de l'erreur de mesure. Enfin, nous présentons une étude par simulation.

MOTS-CLÉS : censuration; erreur de mesure; simulation-extrapolation; données sur la survie.

1. Introduction

L'erreur de mesure fréquente dans les covariables peut avoir des causes diverses. Parfois, les covariables d'intérêt sont difficiles à observer avec précision en raison de l'emplacement physique ou du coût. Par exemple, le rétrécissement du diamètre des artères coronaires pourrait révéler le risque de défaillance cardiaque, mais les médecins pourraient choisir de mesurer le rétrécissement des artères carotides en raison du caractère moins effractif de cette méthode d'évaluation. Parfois, il est impossible de mesurer précisément les covariables en raison de leur nature. Ainsi, le niveau d'exposition à des éventuels facteurs de risque de cancer, comme le rayonnement, ne peut être mesuré exactement (Pierce et coll., 1992). Dans d'autres situations, une covariable peut correspondre à la moyenne d'une certaine grandeur mesurable au cours du temps, mais tout moyen pratique de mesurer cette grandeur comporte nécessairement une erreur de mesure. Les erreurs de mesure des covariables sont fréquentes dans les études longitudinales, les études cas-témoins, les enquêtes par sondage et les analyses de données sur la survie. Le présent article décrit des exemples d'erreur de mesure.

Exemple 1 : Données longitudinales

L'Étude sur la santé cardiaque de Framingham est une étude prospective longitudinale permanente des facteurs de risque de maladie cardiovasculaire (MCV). L'objectif de l'étude de Framingham était de déterminer quels sont les caractéristiques ou facteurs communs qui contribuent à la MCV en suivant sa manifestation et son évolution au cours d'une longue période chez un grand groupe de sujets qui ne présentaient pas encore de symptômes évidents de MCV ou n'avaient pas encore fait de crise cardiaque ni d'accident vasculaire cérébral (Kannel et coll., 1986). L'un des facteurs de risque importants, la pression artérielle systolique, est sujet à une erreur de mesure considérable. Il est impossible de mesurer la pression artérielle systolique à long terme x . À la place, on dispose d'une mesure spécifique W faite durant une visite dans une clinique. La mesure à long terme x et la mesure unique lors d'une

¹ Grace Y. Yi, Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1, courriel : yyi@uwaterloo.ca

² Wenqing He, Department of Statistics and Actuarial Science, University of Western Ontario, 1151 Richmond Street North, London, Ontario, Canada N6A 5B7, courriel : whe@stats.uwo.ca

visite W sont généralement différentes, en raison des variations quotidiennes et saisonnières et de facteurs confusionnels.

Exemple 2 : Données cas-témoins

Les données dont il est question dans Carroll, Gail et Lubin (1993) ont été recueillies dans le contexte d'une étude cas-témoins dont l'objectif principal était d'examiner l'association entre le cancer invasif du col de l'utérus et l'exposition à l'herpès simplex virus type 2 (HSV-2). L'exposition au HSV-2 a été évaluée par une méthode de transfert de Western perfectionnée (x) et par une méthode de transfert de Western moins précise (W) pour les cas ($Y = 1$) et les témoins ($Y = 0$). Toutefois, le résultat x du test n'a été observé directement que pour moins de 6 % des sujets. Les mesures fondées sur la méthode de transfert de Western standard moins précise (W) étaient disponibles pour tous les sujets. Selon ce scénario, une erreur de classification de la covariable x survient si l'on utilise toutes les mesures disponibles W pour étudier la relation entre Y et x .

Exemple 3 : Données d'enquête

Hwang (1986) discute d'un ensemble de données d'enquête recueillies auprès de 5 979 ménages sélectionnés au hasard aux États-Unis. Ces ménages ont déclaré leur consommation annuelle d'énergie. Les caractéristiques du ménage, comme le nombre de fenêtres, la zone chauffée cloisonnée, le nombre de pouces d'isolant dans les murs, l'isolation du toit, l'isolation des planchers, etc., ont été recueillies. Les renseignements sur d'autres variables, comme le revenu familial, la présence ou non de personnes dans le logement durant la journée, la présence ou non de certains gros appareils électroménagers, l'indice de région géographique et les conditions climatiques locales ont également été enregistrés. Il est important de comprendre la relation entre la consommation d'énergie d'un ménage aux États-Unis et les caractéristiques du logement de ce ménage. Toutefois, afin d'assurer la protection des renseignements personnels, certains prédicteurs x risquant de permettre d'identifier les membres d'un ménage ont été multipliés par des variables aléatoires générées à partir d'une loi connue. Autrement dit, x n'est pas disponible, et une variable substitut $W = xU$ a été enregistrée à sa place, où U est une variable aléatoire générée à partir d'une loi connue.

Exemple 4 : Données sur des événements récurrents

Les données sur des événements récurrents sont fréquentes en sciences biomédicales, dans les études démographiques, et en recherche industrielle. Les crises chez les patients épileptiques, les tumeurs successives dans les études sur le cancer, les naissances multiples au cours de la vie d'une femme, et le temps écoulé jusqu'aux réclamations au titre de la garantie pour certains types d'articles manufacturés en sont des exemples. Ainsi, le Nutritional Prevention of Cancer Trial (essai NPC) a été conçu pour étudier la sécurité et l'efficacité à long terme d'un supplément nutritionnel quotidien de 200 μg de sélénium (Se) en vue de prévenir le cancer. Un résultat présentant un intérêt important est celui concernant le carcinome squameux de la peau (CS). Comme une personne peut présenter des récurrences multiples de ce carcinome au cours du temps, il est intéressant d'étudier l'effet d'un supplément de sélénium (Se) sur le carcinome squameux (Jiang, Turnbull et Clark, 1999). Divers facteurs, y compris des variables démographiques, des variables de comportement et des variables médicales de référence peuvent avoir une incidence sur la récurrence du carcinome squameux. Il est bien connu que certains facteurs de risque, comme la concentration plasmatique de sélénium, mesurée en ng/mL, sont sujets à une variabilité biologique et à une erreur de mesure.

Exemple 5 : Données sur la survie

Les données de la Busselton Health Study ont été recueillies à l'aide d'une enquête transversale répétée réalisée dans la collectivité de Busselton en Australie occidentale de 1966 à 1988. L'un des objectifs de l'étude était d'évaluer l'effet des facteurs de risque de maladie cardiovasculaire sur le risque de décès dû à une maladie

coronarienne (Knuiman et coll., 1994). L'ensemble de données analysées dans Yi et Lawless (2006) contient des données sur la survie pour 2 306 couples. Pour 2 266 de ces couples, au moins une réponse est censurée, c'est-à-dire qu'au moins un membre du couple était encore en vie au moment de la dernière observation. Les facteurs de risque de mortalité, surtout la mortalité par maladie coronarienne, comprennent la pression artérielle systolique (PAS), le taux de cholestérol (CHOL), l'âge, la situation d'usage du tabac et l'indice de masse corporelle. On sait que les mesures des facteurs de risque PAS et CHOL sont sujettes à une importante erreur de mesure due à la nature même de ces covariables.

L'erreur de mesure est un problème fréquent en pratique. Ses effets sont complexes et dépendent généralement de la forme du modèle d'erreur et de l'association entre la réponse et les covariables. Afin de faire des inférences valides, il faut pouvoir apporter des corrections appropriées aux problèmes individuels. Pour un examen complet du sujet, voir Carroll et coll. (2006). La discussion qui suit a trait aux données sur la survie avec covariables sujettes à l'erreur de mesure.

2. Données sur la survie avec erreur dans les covariables

L'analyse de survie joue un rôle important dans de nombreux domaines, y compris en sciences biomédicales, en génie et en sciences sociales. Ainsi, dans les études sur les maladies éventuellement mortelles, nous voulons déterminer la durée de survie des personnes atteintes de la maladie. Dans ces circonstances, nous nous intéressons à la relation entre la survie ou la durée de vie des sujets et un certain nombre de covariables qui ont une incidence sur cette survie. Certaines de ces covariables sont toutefois sujettes à une erreur de mesure. Par exemple, dans l'étude de l'infection par le VIH, il est de toute première importance d'estimer la distribution du temps écoulé entre l'infection par le VIH et le diagnostic du SIDA. Toutefois, l'évaluation d'un facteur important, la numération des CD4, présente une erreur de mesure importante, et toute analyse statistique valide devrait tenir compte de ce problème.

Depuis les premiers résultats concernant les covariables sujettes à l'erreur de mesure en analyse de survie discutés dans Prentice (1982), l'intérêt des chercheurs pour ce sujet a grandi. Selon les modèles à risques proportionnels de Cox pour les données univariées sur la survie, l'approche dite de calage par la régression a été proposée par Prentice (1982) pour tenir compte des covariables mesurées incorrectement, et a été suivie par d'autres discussions, dont celles de Nakamura (1992), Wang et coll. (1997), de Buzas (1998), de Hu, Tsiatis et Davidian (1998), de Zhou et Wang (2000), de Huang et Wang (2000), de Tsiatis et Davidian (2001), de Xie, Wang et Prentice (2001), d'Augustin (2004), ainsi que de Yi et Lawless (2006). Dans le cas de données sur la survie regroupées, Li et Lin (2000, 2003) ont discuté de méthodes d'inférence tenant compte de l'erreur de mesure dans les covariables grâce à l'utilisation de modèles de fragilité. Hu et Lin (2004) proposent des méthodes de régression semi-paramétriques pour les temps de défaillance multivariée. De nouveau, l'accent est mis sur les modèles à risques proportionnels de Cox qui sont employés pour caractériser les distributions marginales des temps de défaillance.

L'effet de l'erreur de mesure des covariables est bien compris dans le cas des modèles à risques proportionnels de Cox. Par contre, la discussion de son effet sur les modèles de défaillance accélérée est très limitée, alors que ces modèles ont le grand avantage d'une interprétation transparente (Lawless, 2003). Dans le présent article, nous étudions l'effet de l'erreur de mesure sur les modèles de défaillance accélérée. Nous explorons spécifiquement une méthode de simulation et d'extrapolation (SIMEX) en vue de faire une inférence valide. Nous utilisons le modèle d'erreur additif classique pour faciliter l'obtention de covariables incorrectement mesurées, autrement dit nous supposons que $W = x + e$, où x est la covariable réelle qui est souvent observée de façon imprécise, W est une mesure observée de x , et e suit une loi normale de moyenne 0 (p. ex., Nakamura, 1992; Li et Lin, 2000, 2003).

3. Notation et formulation du modèle

Soit T_i et C_i les temps de défaillance et de censure pour le sujet i , respectivement, et δ_i la variable indicatrice de censure prenant la valeur 1 si $T_i < C_i$ et 0 autrement, $i = 1, 2, \dots, n$. Nous supposons que la censure est

indépendante. Soit Y_i le logarithme du temps de défaillance pour le sujet i , x_i les covariables éventuellement sujettes à une erreur de mesure, et z_i le vecteur des covariables exemptes d'erreur, $i = 1, 2, \dots, n$. La variable de réponse Y_i est caractérisée par le modèle de défaillance accélérée (MDA) donné par

$$Y_i = \beta'_x x_i + \beta'_z z_i + \varepsilon_i \quad (1)$$

où $\beta = (\beta'_x, \beta'_z)'$ est le vecteur des paramètres de régression d'intérêt, et β_z peut inclure le coefficient de l'ordonnée à l'origine. Dénotons les dimensions de β_x et β par p et q , respectivement. Ici, les ε_i suivent une loi $G(\cdot)$ de paramètres α , disons.

Soit W_i une version observée de la covariable. Nous supposons que x_i , z_i et W_i suivent un modèle d'erreur de mesure additif classique, modèle dont l'usage est peut-être le plus répandu en pratique. C'est-à-dire, sachant x_i et z_i ,

$$W_i = x_i + e_i \quad (2)$$

où e_i suit, indépendamment de x_i et z_i , une loi normale de moyenne 0 et de matrice de covariance $\Sigma_e = \text{diag}(\sigma_j^2, j = 1, 2, \dots, p)$.

Les paramètres contenus dans Σ_e peuvent être estimés d'après un ensemble de données de validation ou de mesures répétées de W_i . Si l'on ne dispose ni de données de validation ni de mesures répétées, alors on peut mener des analyses de sensibilité basées sur l'information contextuelle au sujet du processus de mesure pour évaluer l'effet de divers degrés d'erreur de mesure sur l'estimation des paramètres de réponse. Dans le présent article, nous supposons que les paramètres de la distribution de l'erreur sont connus, ou estimés à partir d'un échantillon indépendant.

4. Méthode d'estimation

À la présente sous-section, nous décrivons une méthode d'estimation fonctionnelle basée sur une simulation résultant de l'adaptation de l'approche de simulation-extrapolation (SIMEX) décrite dans Cook et Stefanski (1994). La méthode SIMEX comprend deux étapes : une étape de simulation et une étape subséquente d'extrapolation. La simulation permet d'établir les estimations naïves pour les cas où la variance du terme d'erreur pour chaque mesure W_{ij} , $j = 1, 2, \dots, p$, est $(1 + \lambda)\sigma_j^2$. Ici, W_{ij} est la j^{e} composante de W_i , et $\lambda > 0$ et prend des valeurs dans un ensemble spécifié $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_M\}$, disons. Pour chaque $j = 1, 2, \dots, p$, nous générons un grand nombre, B , de données simulées, $W_{ij}(b, \lambda)$, $b = 1, 2, \dots, B$, en ajoutant à chaque mesure observée W_{ij} une variable aléatoire de moyenne 0 et de variance $\lambda\sigma_j^2$ selon l'équation

$$W_{ij}(b, \lambda) = W_{ij} + \sqrt{\lambda}\sigma_j U_{ijb},$$

où les U_{ijb} sont des observations indépendantes de loi $N(0, 1)$. Pour des valeurs données de λ et b , nous implémentons la fonction **survreg** de **R** pour ajuster le modèle (1) à l'ensemble de données constitué de $\{Y_i, \delta_i, W_i(b, \lambda), z_i\}$. Soit $\hat{\beta}(b, \lambda)$ l'estimation correspondante et $\hat{\Omega}_r(b, \lambda)$, l'estimation de la variance pour la r^{e} composante $\hat{\beta}_r(b, \lambda)$ de $\hat{\beta}(b, \lambda)$, $r = 1, 2, \dots, q$. Soit

$$\begin{aligned}\hat{\beta}_r(\lambda) &= \frac{1}{B} \sum_{b=1}^B \hat{\beta}_r(b, \lambda), \quad \hat{\Omega}_r(\lambda) = \frac{1}{B} \sum_{b=1}^B \hat{\Omega}_r(b, \lambda), \\ \hat{S}_r(\lambda) &= \frac{1}{B-1} \sum_{b=1}^B (\hat{\beta}_r(b, \lambda) - \hat{\beta}_r(\lambda))^2, \text{ et } \hat{\tau}_r(\lambda) = \hat{\Omega}_r(\lambda) - \hat{S}_r(\lambda).\end{aligned}$$

À l'étape de l'extrapolation, pour chaque composante $\hat{\beta}_r(\lambda)$, nous calculons la régression de l'estimation $\hat{\beta}_r(\lambda)$ sur λ et extrapolons la moyenne prévue résultante à $\lambda = -1$ pour obtenir l'estimation $\hat{\beta}_r$; de même, pour chaque r , nous calculons la régression de l'estimation $\hat{\tau}_r(\lambda)$ sur λ et extrapolons la moyenne prévue résultante à $\lambda = -1$ pour obtenir l'estimation de la variance $\hat{\tau}_r$.

Puisque la forme exacte de la fonction d'extrapolation est inconnue, la question qui se pose souvent est de savoir quelle approximation est appropriée. En principe, tout choix entre diverses fonctions d'extrapolation concurrentes peut s'appuyer sur l'examen d'un tracé de $\hat{\beta}_r(\lambda)$ (ou $\hat{\tau}_r(\lambda)$) en fonction de λ . En pratique, les choix de fonctions de régression courants à l'étape de l'extrapolation comprennent la régression linéaire, la régression quadratique et la régression inverse (Cook et Stefanski, 1994). Nous notons qu'en général, l'estimateur SIMEX $\hat{\beta}_r$ est seulement approximativement convergent pour β_r , parce qu'on utilise une fonction d'extrapolation approximative (plutôt qu'exacte) à l'étape de l'extrapolation.

5. Étude par simulation

Nous utilisons une étude par simulation pour étudier l'effet de l'omission de l'erreur de mesure dans l'estimation et pour évaluer la performance de l'approche SIMEX comparativement à la méthode naïve. Dans l'étude par simulation qui suit, nous fixons $n = 200$ et générons 1000 simulations pour chacune des configurations paramétriques. Prenons $B = 200$ et $M = 20$ pour l'approche SIMEX. Nous générons les temps de défaillance pour le modèle

$$Y_i = \beta_0 + \beta_x x_i + \beta_z z_i + \varepsilon_i / \alpha$$

où ε_i suit une loi standard des valeurs extrêmes dont la densité de probabilité est $f(u) = e^u \exp(-e^u)$.

Nous posons que α est égal à 0,5. Nous considérons les cas de censure à 30% et à 50%, où un temps de censure fixe C est généré pour chaque sujet. La covariable réelle x_i est simulée à partir d'une distribution normale $N(1,1)$. Une valeur observée W_i est générée pour la distribution normale conditionnelle $N(x_i, \sigma^2)$, sachant x_i . La covariable z_i est générée à partir de la loi $Bin(1, 0,5)$ pour représenter un plan équilibré. Prenons $\beta_0 = 0$, $\beta_x = -\log 2$ et $\beta_z = 0,5$. Nous considérons des configurations différentes de σ avec $\sigma = 0,15, 0,25, 0,75$, représentant une erreur de mesure faible, moyenne et grande, respectivement.

Dans le tableau, nous présentons les résultats du biais des estimations, l'erreur-type empirique, l'erreur-type fondée sur le modèle et la probabilité de couverture des intervalles de confiance à 95%. Quand $\sigma = 0,25$, l'effet de l'erreur de mesure n'est pas frappant. Par contre, à mesure qu'augmente la grandeur de l'erreur, l'effet de cette dernière devient manifestement visible. L'approche naïve ne produit pas d'estimation convergente, ce qui est évident si l'on examine les estimations de β_x ainsi que de β_0 . Les probabilités de couverture correspondantes pour les intervalles de confiance à 95% s'écartent fortement du niveau nominal de 95%. Il n'est pas étonnant que les estimations de β_z ne soient pas très influencées par l'erreur de mesure, car β_z est le coefficient de la covariable

observée avec précision z_i qui n'est pas corrélée dans la simulation envisagée ici. L'approche SIMEX donne de nettement meilleurs résultats, les biais étant beaucoup plus faibles et les probabilités de couverture, considérablement meilleures. Lorsque l'erreur de mesure est faible ou moyenne, les estimations obtenues par l'approche SIMEX sont satisfaisantes pour tous les paramètres. Les biais d'échantillon fini sont raisonnablement faibles, et les probabilités de couverture concordent relativement bien avec la valeur nominale de 95%. Lorsque l'erreur de mesure est importante, la performance de la méthode SIMEX semble moins impressionnante. Cela pourrait tenir au fait que l'on n'utilise pas la fonction d'extrapolation exacte, mais seulement une approximation à l'étape de l'extrapolation. Toutefois, l'approche SIMEX donne de nettement meilleurs résultats que la méthode naïve, moyennant un certain degré d'ajustement de l'erreur de mesure. On peut voir que la méthode SIMEX exagère les erreurs de mesure, contrairement à la méthode naïve. Les erreurs-types produites par les méthodes naïve et SIMEX augmentent parallèlement à l'erreur de mesure. Il est intéressant de constater que l'accroissement de la proportion de censurement n'augmente pas nécessairement le biais. Toutefois, la précision des estimations diminue.

6. Discussion

L'effet de l'erreur de mesure sur les covariables est bien décrit pour les données sur la survie postulées par les modèles à risques proportionnels de Cox, mais son effet dans le cas des modèles de défaillance accélérée, qui représente un outil utile pour l'analyse des données sur la survie, a fait l'objet de peu de discussions. Dans le présent article, nous nous concentrons sur les modèles de défaillance accélérée et étudions les effets de l'erreur de mesure sur l'estimation des paramètres de réponse. Yi et He (2006) ont exploré le problème de l'erreur de mesure dans le cas de données sur la survie bivariées selon les modèles de défaillance accélérée, mais leur discussion portait sur le modèle de défaillance accélérée avec distribution normale de l'erreur. Ici, notre discussion s'applique aux modèles de défaillance accélérée associés à des formes générales de distribution, au lieu d'être limitée à une distribution particulière, comme la distribution normale. Nous décrivons une méthode basée sur la simulation pour corriger le biais induit par l'erreur dans les covariables. Cette méthode est séduisante en ce sens qu'elle est facile à appliquer et qu'elle ne nécessite pas la spécification de la distribution suivie par les covariables réelles sujettes à l'erreur de mesure, distribution qui est généralement inobservable. Notre simulation démontre que l'approche SIMEX donne de nettement meilleurs résultats que la méthode naïve. Ses propriétés sont raisonnablement satisfaisantes, spécialement dans le cas d'une erreur de mesure faible ou moyenne.

En nous inspirant de Li et Lin (2000), nous pourrions étudier une méthode basée sur la vraisemblance pour tenir compte de l'erreur de mesure dans le cadre de modèles de défaillance accélérée. Dans un tel contexte, nous devons opter pour une forme hypothétique de la distribution des covariables sujettes à l'erreur. Il serait intéressant de comparer les propriétés de cette approche à celles de la méthode SIMEX.

Remerciement

Cette étude a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada.

Références

- Augustin, T. (2004), "An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions", *Scandinavian Journal of Statistics*, 31, pp.43-50.
- Buzas, J. F. (1998), "Unbiased scores in proportional hazards regression with covariate measurement error", *Journal of Statistical Planning and Inference*, 67, pp.247-257.
- Carroll, R. L., Gail, M. H., et Lubin, J. H. (1993), "Case-control studies with errors in covariates", *Journal of the American Statistical Association*, 88, pp.185-199.

- Carroll, R. J., Ruppert, D., Stefanski, L. A., et Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models*, 2nd ed., Chapman & Hall.
- Cook, J. et Stefanski, L. A. (1994), "A simulation extrapolation method for parametric measurement error models", *Journal of the American Statistical Association*, 89, pp. 464-467.
- Hu, C. et Lin, D. Y. (2004), "Semiparametric failure time regression with replicates of mismeasured covariates", *Journal of the American Statistical Society*, 99, pp.105-118.
- Hu, P., Tsiatis, A. A., et Davidian, M. (1998), "Estimating the parameters in the Cox model when covariate variables are measured with error", *Biometrics*, 54, pp.1407-1419.
- Huang, Y. et Wang, C. Y. (2000), "Cox regression with accurate covariates unascertainable: a nonparametric correction approach", *Journal of the American Statistical Association*, 95, pp.1209-1219.
- Hwang, J. T. (1986), "Multiplicative errors-in-variables models with application to recent data released by the U.S. department of energy", *Journal of the American Statistical Association*, 81, pp.680-688.
- Jiang, W., Turnbull, B. W., et Clark, L. C. (1999), "Semiparametric regression models for repeated events with random effects and measurement error", *Journal of the American Statistical Association*, 94, pp.111-124.
- Kannel, W. B., Neaton, J. D., Wentworth, D., Thomas, H. E., Stamler, J., Hulley, S. B., et Kjelsberg, M. O. (1986), "Overall and coronary heart disease mortality rates in relation to major risk factors in 325,348 men screened for MRFIT", *American Heart Journal*, 112, pp.825-836.
- Knuiman, M. W., Cullent, K. J., Bulsara, M. K., Welborn, T. A., et Hobbs, M. S. T. (1994), "Mortality trends, 1965 to 1989, in Busselton, the site of repeated health surveys and interventions", *Australian Journal of Public Health*, 18, pp.129-135.
- Lawless, J. F. (2003), *Statistical Models and Methods for Lifetime Data*, 2nd ed., Johns Wiley & Sons, New York.
- Li, Y. et Lin, X. (2000), "Covariate measurement errors in frailty models for clustered survival data", *Biometrika*, 87, pp.849-866.
- Li, Y. et Lin, X. (2003), "Functional inference in frailty measurement error models for clustered survival data using the SIMEX approach", *Journal of the American Statistical Association*, 98, pp.191-203.
- Nakamura, T. (1992), "Proportional hazards model with covariates subject to measurement error", *Biometrics*, 48, pp.829-838.
- Pierce, D. A., Stram, D. O., Vaeth, M., et Schafer, D. (1992), "Some insights into the errors in variables problem provided by consideration of radiation dose-response analyses for the A-bomb survivors", *Journal of the American Statistical Association*, 87, pp.351-359.
- Prentice, R. L. (1982), "Covariate measurement errors and parameter estimation in a failure time regression model", *Biometrika*, 69, pp.331-342.
- Tsiatis, A. A. et Davidian, M. (2001), "A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error", *Biometrika*, 88, pp.447-458.
- Wang, C. Y., Hsu, L., Feng, Z. D., et Prentice, R. L. (1997), "Regression calibration in failure time regression", *Biometrics*, 53, pp.131-145.
- Xie, S. H., Wang, C. Y., et Prentice, R. L. (2001), "A risk set calibration method for failure time regression by using a covariate reliability sample", *Journal of the Royal Statistical Society B*, 63, pp.855-870.

Yi, G. Y. et He, W. (2006), “Methods for bivariate survival data with mismeasured covariates under an accelerated failure time model”, *Communications in Statistics - Theory and Methods*, 35, pp.1539-1554.

Yi, G. Y. et Lawless, J. F. (2006), “A corrected likelihood method for the proportional hazards model with covariates subject to measurement error”, En cours d’impression dans *Journal of Statistical Planning and Inference*.

Zhou, H. et Wang C. Y. (2000), “Failure time regression with continuous covariates measured with error”, *Journal of the Royal Statistical Society B*, 62, pp.657-665.

Tableau : Résultats de la simulation

	σ	Méthode	β_0				β_x				β_z			
			Biais	E.-T. (M)	E.-T. (E)	PC	Biais	E.-T. (M)	E.-T. (E)	PC	Biais	E.-T. (M)	E.-T. (E)	PC
I	0,25	NAÏVE	0,067	0,320	0,309	0,961	-0,071	0,216	0,219	0,916	0,018	0,396	0,391	0,964
	0,50		0,312	0,325	0,319	0,862	-0,276	0,199	0,211	0,647	-0,021	0,405	0,401	0,955
	0,75		0,511	0,326	0,322	0,677	-0,506	0,176	0,181	0,195	0,014	0,408	0,402	0,956
	0,25	SIMEX	-0,018	0,327	0,316	0,967	0,014	0,232	0,236	0,948	0,015	0,398	0,392	0,961
	0,50		0,075	0,341	0,345	0,945	-0,032	0,241	0,265	0,913	-0,032	0,410	0,410	0,955
	0,75		0,190	0,342	0,358	0,912	-0,175	0,231	0,259	0,812	0,001	0,413	0,415	0,944
II	0,25	NAÏVE	0,079	0,397	0,411	0,947	-0,074	0,259	0,258	0,924	-0,007	0,473	0,486	0,950
	0,50		0,256	0,403	0,397	0,945	-0,273	0,237	0,235	0,738	0,015	0,479	0,480	0,943
	0,75		0,460	0,410	0,419	0,853	-0,516	0,208	0,213	0,319	0,026	0,481	0,489	0,944
	0,25	SIMEX	0,001	0,401	0,417	0,939	0,010	0,277	0,276	0,952	-0,010	0,475	0,488	0,950
	0,50		0,033	0,412	0,410	0,945	-0,029	0,287	0,294	0,937	0,006	0,483	0,487	0,945
	0,75		0,162	0,416	0,441	0,942	-0,193	0,272	0,298	0,818	0,019	0,486	0,500	0,947

Cas I : censurement à 30 %; cas II : censurement à 50 %.

E.-T. (M) : erreur-type basée sur le modèle; E.-T. (E) : erreur-type empirique; PC : probabilité de couverture à 95 %.