Modèles de regression en présence de compétition

Aurélien Latouche

THESE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité

Biomathématique (ED 0393)

présentée par

M. Aurélien Latouche

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITE PARIS 6

Sujet de la thèse :

Modèles de régression en présence de compétition

Soutenue le 20 décembre 2004

devant le jury composé de :
Mme Catherine Quantin, Rapporteur (PU-PH)
M. Daniel Commenges, Rapporteur (DR1 INSERM)
Mme Sylvie Chevret, Directeur de thèse (PU-PH)
M. Raphaël Porcher, Examinateur (MCU-PH)
M. Guy Thomas, Examinateur (PU-PH)
M. Alain Jacques Valleron, Examinateur (PU-PH)

Table des matières

Chapitre 1

Introduction

Le terme de risques compétitifs (ou concurrents) se rapporte au domaine de l'analyse de survie où, en plus d'un temps (ou délai) d'événement, on observe aussi un type (ou une cause) unique d'événement. Si ce concept remonte au XVIII^{e} siècle, avec l'estimation des avantages de la vaccination contre la variole sur l'espérance de vie par Bernoulli (1760), il a connu un regain d'intérêt récent, notamment avec de nombreuses applications en épidémiologie et en recherche clinique. Les exemples en sont nombreux : lors d'une grossesse simple, la césarienne empêche un accouchement par voie basse et inversement. En hématologie, les patients leucémiques recevant une allogreffe de moëlle peuvent présenter une maladie du greffon contre l'hôte (les lymphocytes du greffon attaquant l'organisme du receveur), rechuter ou décéder. Il est alors possible de considérer que ces trois événements (et par extension leur risque associé) sont en compétition dans la mesure où la survenue du décès post-greffe empêche la survenue d'une rechute et de la maladie du greffon contre l'hôte.

Un modèle à risques compétitifs à K causes mutuellement exclusives est représenté sur la Figure (1.1). De nombreuses méthodes ont été proposées pour étudier les risques compétitifs (Tsiatis, 1998). En particulier, l'approche par temps d'événements latents proposée par Sampford (1952) a souvent été utilisée dans la caractérisation des modèles de survie puis des risques compétitifs. Cette approche considère K variables aléatoires positives, $T_1, \ldots, T_K$, chacune correspondant au temps de survenue d'un des événements considérés, dans la situation hypothétique où seul un type d'événement peut survenir (les événements étant exclusifs). Le temps T est alors le minimum des T_k et la cause d'événement ε est k lorsque $T = T_k$. Ces temps latents ne sont généralement pas observables et doivent plutôt être vus comme une représentation théorique. Cependant, ils sont aussi adaptés à la modélisation de la première question posée dans le cadre

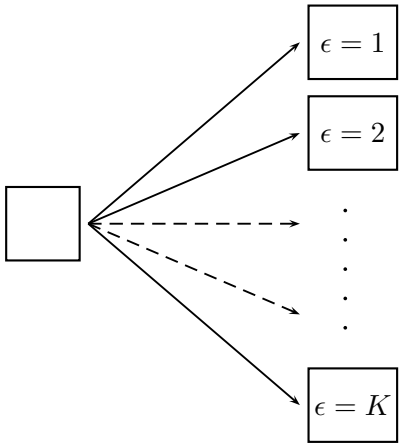


FIG. 1.1 – Modèle à K risques compétitifs

de risques compétitifs, à savoir que se passerait-il si l'une des causes d'événement pouvait être supprimée (par exemple, la mortalité liée au tabagisme, dans l'évaluation d'un programme de sensibilisation). Au contraire, les observations sont uniquement constituées de réalisations du couple (T, ε) (éventuellement indicées par i en référence à l'individu) et il a été montré que plusieurs distributions jointes des temps latents sous différentes conditions d'indépendance mutuelle pouvaient conduire à la même vraisemblance des observations (Prentice et al., 1978). Enfin, il est courant dans les études cliniques que ces données soient soumises à un processus de censure à droite, indépendant de (T, ε), qui complique l'estimation.

Pour analyser des données dans un cadre de risques compétitifs, deux types de probabilités peuvent être définies :

- la probabilité "brute" d'événement de type k en présence des autres risques d'événement, dite aussi fonction d'incidence cumulée ou sous-fonction de répartition de l'événement de type k, $F_k(t) = \Pr(T \leq t, \varepsilon = k)$
- la probabilité "nette" d'événement de type k, dans la situation où seul ce risque agirait sur la population.

Il a été montré que les probabilités nettes n'étaient pas des quantités identifiables à partir des observations, pas plus que la distribution jointe des temps latents, à moins de supposer par exemple que les temps latents sont mutuellement indépendants. Cependant, cette hypothèse ne peut être vérifiée (Tsiatis, 1975).

Dans le cadre de risques compétitifs, l'objet central est la fonction de risque cause-spécifique d'événement de type k, qui s'interprète comme la probabilité de survenue de l'événement de type

k dans un intervalle infinitésimal, sachant que cet événement ne s'est pas encore produit au début de l'intervalle. D'une manière plus générale, le modèle à risques compétitifs est un cas particulier des modèles multi-états (Commenges, 1999) où à partir d'un état "vivant", les individus peuvent expérimenter K causes d'événements exclusives (figure 1.1). Les taux de transition (ou intensité) entre chaque état sont des fonctions de risque cause-spécifique. La somme de toutes ces intensités correspond au risque global de quitter l'état "vivant".

Un modèle de régression entend établir une relation entre une variable réponse (ici un délai d'événement) et des covariables explicatives pouvant ou non dépendre du temps. De nombreux modèles de régression pour la fonction de risque cause-spécifique ont été proposés, en particulier le modèle semi-paramérique à risques proportionnels de Cox (1972), qui relie la fonction de risque cause-spécifique à une covariable Z (qui peut être un vecteur) par la relation : $\lambda_k(t;Z) = \lambda_{k0}(t)\exp(\beta Z)$, où $\lambda_{k0}(t)$ est une fonction continue non spécifiée, représentant le risque de base. Toutefois, la fonction de survie marginale, $S(t) = \Pr(T > t)$, dépend ici de toutes les fonctions de risque cause-spécifique des K causes considérées à travers

$$S(t) = \exp\left\{-\sum_{k=1}^{K}\int_0^t \lambda_k(u)du\right\}.$$

Des modèles de régression pour la fonction d'incidence cumulée ont aussi été proposés (Fine and Gray, 1999; Fine, 2001; Andersen et al., 2003). En particulier, Fine et Gray ont proposé un modèle semi-paramérique à risques proportionnels de formulation similaire au modèle de Cox, pour la fonction de risque associée à la fonction d'incidence cumulée proposée par Gray (1988) :

$$\alpha_k(t) = -\frac{d}{dt}\log\{1 - F_k(t)\} \tag{1.1}$$

Le modèle à risques proportionnels s'écrit $\alpha_k(t;Z) = \alpha_{k0}(t)\exp(\gamma Z)$ où $\alpha_{k0}(t)$ est une fonction continue non spécifiée. La fonction de risque associée à la fonction d'incidence cumulée est aussi appelée fonction de risque de sous-répartition (*subdistribution hazard*).

On dispose donc de deux modèles de formulation similaire, mais dont l'interprétation diffère et qui ne peuvent pas être simultanément vrais, sauf dans le cas dégénéré où seule une cause agit sur la population. Dans un contexte pratique d'analyse de données en présence de compétition, il est alors important de pouvoir examiner plus précisement les implications d'un choix de modélisation. Il s'agit de l'objectif de ce travail de thèse.

Plan de Thèse

Nous présentons, dans un premier temps, les définitions des objets utilisés en analyse de survie en présence de compétition, en insistant sur l'existence, l'identifiabilité et l'interprétation des objets dans une modélisation à temps latents. Le modèle de Fine & Gray (1999) est le modèle de référence pour ce travail de thèse.

Lors de la planification d'essai clinique, l'évaluation du nombre de sujets à inclure est un problème crucial. Une telle formule n'existant pas pour le modèle de Fine & Gray, nous avons donc établi un calcul d'effectif dans un contexte de compétition pour une inférence basée sur la fonction d'incidence cumulée. Ce travail constitue le premier développement de cette thèse.

Nous avons ensuite étudié les propriétés du modèle de Fine & Gray lorsqu'il est mal spécifié, notamment, si un modèle de Cox pour la fonction de risque cause-spécifique est postulé, et qu'une analyse est conduite en modélisant la fonction de risque de sous-répartition. C'est le deuxième développement que présente ce travail.

D'autre part, lors d'études longitudinales de cohorte, certains marqueurs sont souvent mesurés de manière répétée. Les covariables sont alors dépendantes du temps et il est de plus fréquent que les individus soient confrontés à des événements en compétition. Nous avons donc éudié la possibilité d'inclure des covariables dépendantes du temps dans un modèle de Fine & Gray. Ceci constitue le troisième volet de ce travail de thèse.

Enfin, nous présentons un article de vulgarisation à l'intention de médecins et d'épidémiologistes, et constitue une synthèse concernant la stratégie d'utilisation des modèles de régression en présence de compétition.

Chaque développement apporté au modèle de Fine & Gray (1999) est présenté sous la forme d'une introduction au problème soulevé, puis de l'article (soumis ou accepté) qui en constitue le développement.

Chapitre 2

Modélisation basée sur les fonctions de risque cause-spécifique et d'incidence cumulée

2.1 Temps latents d'événements

On s'intéresse ici aux problèmes d'identifiabilité liés à une modélisation par temps latents. Les observations consistent en $T = \min(T_1, \ldots, T_K)$ et $\varepsilon = k$ si $T = T_k$ *i.e.*, $\varepsilon = \text{Argmin}(T_1, \ldots, T_K)$.

Causes multiples agissant sur la population

L'objet principal de ce type de modélisation est la fonction jointe de survie, lorsque tous les K risques agissent sur la population, qui est définie par

$$H(t_1, \ldots, t_K) = \Pr(T_1 \geq t_1, \ldots, T_K \geq t_K). \tag{2.1}$$

En présence de tous les risques, on peut définir différentes fonctions :

- la fonction de sous-répartition "brute" : $F_k(t) = \Pr(T \leq t, \varepsilon = k)$
- la fonction de risque cause-spécifique : $\lambda_k(t) = \lim_{h \to 0} \Pr(t \leq T \leq t + h, \varepsilon = k | T \geq t)/h$

Si l'on considère la fonction de répartition $F(t) = \sum_{k=1}^{K} F_k(t) = \Pr(T \leq t)$ et la fonction de survie $S(t) = 1 - F(t)$, on obtient les relations suivantes

$$\lambda_k(t) = \frac{dF_k(t)}{dt} / S(t), \quad k = 1, \ldots, K \tag{2.2}$$

et

$$\lambda(t) = \sum_{i=k}^{K} \lambda_k(t) = \frac{dF(t)}{dt}/S(t). \tag{2.3}$$

La fonction de sous répartition étant reliée à la fonction jointe de survie, H, par :

$$\frac{dF_k(t)}{dt} = -\frac{\partial H(t_1, \ldots, t_K)}{\partial t_k} | t_1 = \ldots = t_K = t,$$

on a

$$\lambda_k(t) = \frac{-\frac{\partial H(t_1, \ldots, t_K)}{\partial t_k} | t_1 = \ldots = t_K = t}{H(t, \ldots, t)}. \tag{2.4}$$

Cause unique agissant sur la population

On considère maintenant qu'une seule cause agit sur la population, l'exposant k indique alors le seul risque agissant. La distribution marginale pour le temps latent T_k est $\Pr(T_k \geq t) = H(t_1, \ldots, t_K)|t_k = t, t_j = 0, j \neq k$.

Définissons la fonction de risque instantané marginale pour la cause k

$$\lambda_k^k(t) = \lim_{h \to 0} \Pr(t \leq T_k \leq t + h | T_k \geq t)/h$$

et notons $S_k^k(t) = \Pr(T_k \geq t)$, la fonction de survie nette. On a

$$\lambda_k^k(t) = -\frac{dS_k^k(t)}{dt}/S_k^k(t) \tag{2.5}$$

Si les temps latents sont indépendants, alors la fonction jointe de survie se factorise en produit des fonctions de survie nettes, soit

$$H(t_1, \ldots, t_K) = \prod_{k=1}^{K} S_k^k(t_k). \tag{2.6}$$

Donc, sous cette hypothèse, les fonctions de risque causes-spécifiques en présence de toutes les causes et en seule présence de la cause k, sont égales, *i.e.*

$$\lambda_k = \lambda_k^k.$$

Sous l'hypothèse d'indépendance des temps latents, les formules (2.1) et (2.3) nous permettent d'obtenir la relation suivante entre la fonction de survie nette, H_k^k, et la fonction de sous répartiton "brute"

$$H_k^k(t) = \exp\left[-\int_0^t \frac{dF_k(u)}{S(u)} du\right] \tag{2.7}$$

Tsiatis (1975) a montré qu'en l'absence d'hypothèse d'indépendance des $(T_1, \ldots, T_K)$, les probabilité brutes ne peuvent être identifiées aux probabilités nettes.

Pour illustrer ce résultat et l'importance de l'hypothèse d'indépendance des temps latents, nous présentons ici l'exemple donné par Prentice et al. (1978). Considérons deux risques en compétition, avec la fonction de survie jointe suivante

$$H(t_1, t_2) = \exp\left[1 - \alpha_1 t_1 - \alpha_2 t_2 - \exp\{\alpha_{12}(\alpha_1 t_1 + \alpha_2 t_2)\}\right] \tag{2.8}$$

où $\alpha_1, \alpha_2 > 0$ et $\alpha_{12} > -1$.

Les fonctions de risque cause-spécifique sont alors

$$\lambda_k(t) = \alpha_k \left[1 + \alpha_{12} \exp\{\alpha_{12}(\alpha_1 + \alpha_2)t\}\right], \quad k = 1, 2$$

Si $\alpha_{12} = 0$, les temps latents sont indépendants car la fonction jointe de survie se factorise en produit des fonctions de survie marginales, $H(t_1, t_2) = H_1^1(t_1) \times H_2^2(t_2)$. Cependant, si on considère la fonction jointe de survie H^* définie par

$$H^*(t_1, t_2) = \exp\left[1 - \alpha_1 t_1 - \alpha_2 t_2 - \sum_{k=1}^{2} \alpha_k \exp\{\alpha_{12}(\alpha_1 + \alpha_2)t_k)\}/(\alpha_1 + \alpha_2)\right],$$

on obtient des fonctions cause-spécifique identiques au modèle (2.8), mais des temps latents indépendants pour toutes les valeurs de a_{12}. On constate donc qu'il convient d'insister sur l'importance des hypothèses (d'indépendance) qui permettent de relier les quantités "brutes" ou "nettes".

2.2 Fonction de risque de sous-répartition

La fonction de risque de sous-répartition (1.1), pour l'événement de type 1, peut se définir aussi comme

$$\alpha_1(t) = \lim_{h \to 0} \frac{\Pr\{t \leq T \leq t + h, \varepsilon = 1 | T \geq t \cup (T \leq t \cap \varepsilon \neq 1)\}}{h}, \tag{2.9}$$

ce qui correspond à la fonction de risque instantané de la pseudo-variable aléatoire

$$T^* = 1_{[\varepsilon=1]} \times T + 1_{[\varepsilon \neq 1]} \times \infty. \tag{2.10}$$

Les relations (1.1) et (2.2) nous fournissent le système suivant

$$\begin{cases} \lambda_1(t) = \frac{1}{S(t)} \frac{dF_1(t)}{dt} \\ \alpha_1(t) = \frac{1}{1 - F_1(t)} \frac{dF_1(t)}{dt} \end{cases}$$

d'où :

$$\lambda_1(t) = \frac{1 - F_1(t)}{S(t)} \alpha_1(t). \tag{2.11}$$

Le rapport, g, de la fonction de risque cause-spécifique et de la fonction de risque de sous-répartition est une fonction croissante.

La preuve est donnée par :

$1 - F_1(t) = \exp\left[-\int_0^t \alpha_1(u)du\right]$, de plus $S(t) = \exp\left[-\int_0^t \lambda_1(u) + \lambda_2(u)du\right]$, d'où

$$g(t) = \exp\left[\int_0^t \lambda_1(u) + \lambda_2(u) - \alpha_1(u)du\right],$$

donc $g'(t) = [\lambda_1(t) + \lambda_2(t) - \alpha_1(t)] \exp\left[\int_0^t \lambda_1(u) + \lambda_2(u) - \alpha_1(u)du\right]$. Or $S(t) \leq 1 - F_1(t)$ (Peterson, 1976), ce qui implique $\lambda_1(t) \geq \alpha_1(t)$, par suite $g' \geq 0$.

2.3 Modèles de Cox et de Fine & Gray

Nous présentons la vraisemblance partielle provenant du modèle de Fine et Gray, ainsi que le comportement asymptotique de l'estimateur du paramètre de régression en absence de censure, puis en présence d'une censure à droite. Ces propriétés seront utilisées lors des Chapitres 3 et 4.

Nous introduisons les notations et objets qui seront utiles dans la suite. Une revue exhaustive des processus de comptage et de l'analyse de survie peut être trouvée dans Andersen et al. (1993) et Fleming and Harrington (1991). On considère n individus, soumis à K causes d'événement exclusives.

Soit $Y_i(t) = I(T_i \geq t)$, la variable qui indique si l'individu i est "à risque" avant l'instant t et $Y(t) = \sum_{i=1}^n Y_i(t)$. On introduit le processus qui compte la survenue d'événement de type 1, $N_i(t) = I(T_i \leq t, \varepsilon_i = 1)$ et $N(t) = \sum_{i=1}^n N_i(t)$.

Absence de censure

En l'absence de censure, la vraisemblance partielle pour l'échantillon de taille n porte sur le triplet $(T_i, \varepsilon_i, Z_i)_{i=1,\ldots,n}$. L'ensemble des individus à risque (*risk-set*) s'exprime comme

$$R_i = \{j : (T_j > T_i) \cup (T_j \leq T_i \cap \varepsilon \neq 1)\}. \tag{2.12}$$

La vraisemblance partielle pour la fonction d'incidence cumulée de l'événement de type 1 ($\varepsilon_i = 1$) est

$$\mathcal{L}(\beta) = \prod_{i=1}^n \left[\frac{\exp(\beta Z_j)}{\sum_{j \in R_i} \exp(\beta Z_j)}\right]^{I(\varepsilon_i = 1)}.$$

Le score s'exprime alors

$$\mathcal{U}(\beta) = \sum_{i=1}^n I(\varepsilon_i = 1)\left[Z_i - \frac{\sum_{j \in R_i} Z_j \exp(\beta Z_j)}{\sum_{j \in R_i} \exp(\beta Z_j)}\right],$$

ou, en utilisant les processus de comptage,

$$\mathcal{U}(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \left[Z_i(s) - \frac{\sum_{j=1}^{n} Y_j(s) Z_j \exp \beta Z_j}{\sum_{j=1}^{n} Y_j(s) \exp \beta Z_j} \right] dN_i(s). \tag{2.13}$$

L'estimateur de β vérifie $\mathcal{U}(\hat{\beta}) = 0$. Soit β_0 la valeur théorique de β. L'utilisation classique des processus de comptage et du théorème de limite centrale pour les martingales (Rebodello, 1978) permet d'établir le comportement asymptotique de $\hat{\beta}$: $\sqrt{n}(\hat{\beta} - \beta_0)$ est gaussien centré, de matrice de variance-covariance Ω^{-1} où Ω est estimée par

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{S_1^{(2)}(\hat{\beta}, T_i)}{S_1^{(0)}(\hat{\beta}, T_i)} - \bar{Z}(\hat{\beta}, T_i)^{\otimes 2} \right]$$

où pour un vecteur v, $v^{\otimes 0} = 1$, $v^{\otimes 1} = 1$ et $v^{\otimes 2} = vv'$, $S_1^{(p)}(\beta, u) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) Z_i^{\otimes p} \exp(\beta Z_i) \quad (p = 0, 1, 2)$ et $\bar{Z}(\beta, u) = \frac{S_1^{(1)}(\beta,u)}{S_1^{(0)}(\beta,u)}$.

On constate que, hormis la formulation des individus à risque, les méthodes d'estimation se déduisent de celles employées pour le modèle de Cox.

Censure à droite

Dans le cas où les données sont soumises à une censure à droite, la méthode de pondération de la probabilité inverse de censure ("inverse probability of censoring weighting technique"(Robins and Rotnitzky, 1992)) est utilisée. Cette méthode permet d'obtenir un estimateur du score (2.13) en présence de censure à droite. Si $G(t)$ est la probabilité de ne pas être censuré à l'instant t, on définit la pondération suivante

$$w_i(t) = \begin{cases} \frac{\hat{G}(t)}{\hat{G}(t \wedge T_i)}, \text{si} \quad N_i(t) \quad \text{est observé} \\ 0, \quad \text{sinon} \end{cases}$$

où $\hat{G}$ est l'estimateur de Kaplan-Meier de G, qui est convergent quand la censure est indépendante.

Le score pondéré s'écrit

$$\tilde{\mathcal{U}}(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \left[Z_i(s) - \frac{\sum_{j=1}^{n} w_j(t) Y_j(s) Z_j \exp \beta Z_j}{\sum_{j=1}^{n} w_j(t) Y_j(s) \exp \beta Z_j} \right] w_i(t) dN_i(s). \tag{2.14}$$

Un estimateur consistant de β est obtenu en resolvant $\tilde{\mathcal{U}}(\beta) = 0$.

Un développement en série de Taylor au voisinage de la solution théorique, β_0 nous fournit l'approximation au premier ordre

$$\sqrt{n}(\hat{\beta} - \beta_0) \equiv \mathcal{I}^{-1} \{ \tilde{\mathcal{U}}(\beta_0) / \sqrt{n} \}.$$

Un estimateur consitent de $\mathcal{I}$ en présence de censure à droite est donné par

$$\hat{\mathcal{I}} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{S_2^{(2)}(\hat{\beta}, T_i)}{S_2^{(0)}(\hat{\beta}, T_i)} - \mathbb{E}(\hat{\beta}, T_i)^{\otimes 2}\right]$$

où

$$\mathbb{E}(\beta, T_i) = \frac{S_2^{(1)}(\beta, u)}{S_2^{(0)}(\beta, u)}$$

et

$$S_2^{(p)}(\beta, u) = \frac{1}{n}\sum_{i=1}^{n} w_i(t) Y_i(t) Z_i^{\otimes p} \exp(\beta Z_i), \quad p = 0, 1, 2$$

On montre ensuite que $\tilde{\mathcal{U}}(\beta_0)/\sqrt{n}$ converge vers une loi gaussienne centrée, de matrice de variance-covariance Σ. La forme de $\hat{\Sigma}$ est donnée dans (Fine and Gray, 1999, page 500).

2.4 Etude détaillée dans un cadre paramétrique

Afin d'illuster les différences entre les fonctions de risque présentées, nous avons considéré que les distributions des temps latents sont spécifiées de manière paramétrique. Nous avons explicité les relations qui existent entre ces fonctions de risques si l'on considère 2 causes d'événements. On dispose alors d'un couple de temps latents (T_1, T_2) dont on observe $T = min(T_1, T_2)$ et l'indicateur du type d'événement $\varepsilon = 2 - 1_{(T_1 \leq T_2)}$.

Supposons que (T_1, T_2) suive une loi ACBVE (*Absolutely Continuous BiVariate Exponential*) introduite par Block and Basu (1974) et notée $(T_1, T_2) \sim$ ACBVE(a_1, a_2, a_{12}). Soit $T = \min(T_1, T_2)$, alors $T \sim Exp(a)$ où $a = a_1 + a_2 + a_{12}$, et T est indépendante de $T_1 - T_2$. La fonction de sous-répartition de l'événement de type 1, F_1, les fonctions de risque cause-spécifique, λ_1, marginale, λ_1^1, et de sous-répartition, α_1, s'expriment alors respectivement par :

$$F_1(t) = P(T \leq t, T_1 \leq T_2) = P(T \leq t)P(T_1 - T_2 \leq 0) = \frac{a_1}{a_1 + a_2}[1 - \exp(-at)],$$

$$\lambda_1(t) = \frac{aa_1}{a_1 + a_2},$$

$$\lambda_1^1(t) = a\left[\frac{a_1 + a_{12} - a_{12}\exp(-a_2 t)}{a - a_{12}\exp(-a_2 t)}\right],$$

$$\alpha_1(t) = \frac{aa_1}{a_1 + a_2\exp(at)}.$$

Il apparaît clairement que la fonction λ_1^1 est égale à λ_1 sous l'hypothèse d'indépendance des temps latents ($a_{12} = 0$). La fonction de risque cause-spécifique est une fonction constante du temps, que les deux temps latents soient indépendants ou non, alors que la fonction de risque marginale ne l'est que dans le cas indépendant.

Dans notre cadre paramétrique, on obtient la forme suivante pour le rapport des fonctions λ_1 et α_1 :

$$g(t) = \frac{a_1}{a_1 + a_2} + \frac{a_2}{a_1 + a_2} \exp(at).$$

Le comportement des autres fonctions présentées plus haut en fonction du paramètre de dépendance entre les deux temps latents T_1 et T_2, a_{12}, est représenté sur la figure 2.1. Le graphe (a)

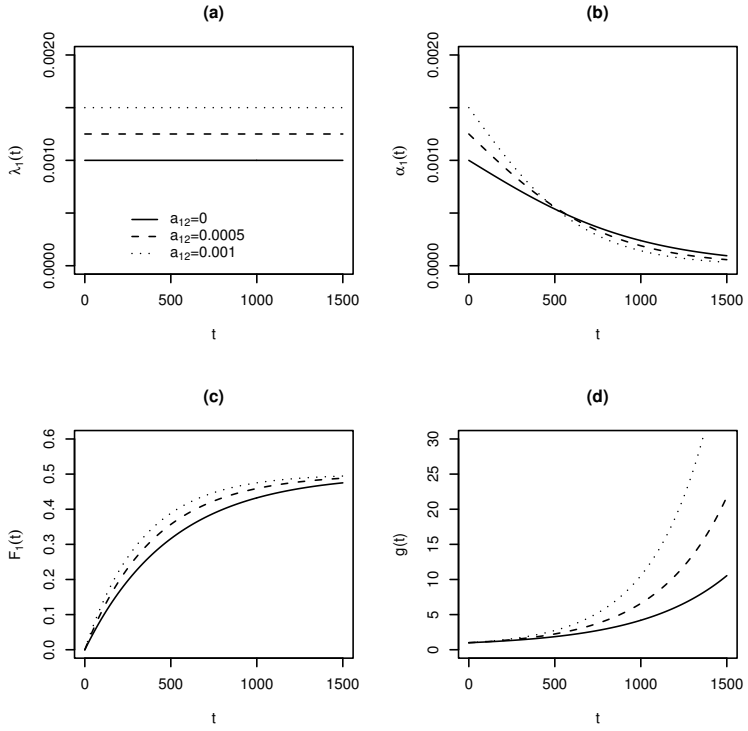


FIG. 2.1 – Fonction de risque cause-spécifique **(a)**, fonction de risque de sous-répartition **(b)**, fonction d'incidence cumulée **(c)** et rapport des deux fonctions de risque **(d)**

de la figure 2.1 illustre cette égalité dans le cas indépendant, toutes deux sont égales à a_1, T_1 étant distribué selon une loi exponentielle de paramètre a_1. Les graphes (b), (c) et (d) tracent respectivement les fonctions α_1, F_1 et g dans les mêmes conditions. Il apparaît que l'influence de la dépendance qui semble importante sur les fonctions de risque (graphe a) n'a finalement que peu d'influence sur la fonction d'incidence cumulée (graphe c).

Prise en compte de covariables

En considérant une covariable binaire Z dans le cadre du modèle de Cox, la fonction de risque cause-spécifique d'événement de type 1 est :

$$\lambda_1(t;Z) = \frac{aa_1}{a_1+a_2}\exp(\beta_1 Z),$$

et la fonction de sous-répartition de l'événement de type 1 est donnée par

$$F_1(t;Z) = \frac{a_1\exp(\beta_1 Z)}{a_1\exp(\beta_1 Z)+a_2\exp(\beta_2 Z)}[1-\exp[-\Psi(Z)t]],$$

où $\Psi(Z) = \lambda_1\exp(\beta_1 Z)+\lambda_2\exp(\beta_2 Z)$ et $\lambda_i = \frac{aa_i}{a_1+a_2}$. La fonction de risque de sous-répartition est alors

$$\alpha_1(t;Z) = \frac{a_1\exp(\beta_1 Z)\Psi(Z)\exp\{-\Psi(Z)t\}}{a_2\exp(\beta_2 Z)+a_1\exp(\beta_1 Z)\exp\{-\Psi(Z)t\}}.$$

Dans ce contexte, il est possible de donner une forme explicite à la fonction $\gamma(.)$ conditionnellement à $Z=1$. En effet

$$\gamma(t) = \log\left[\frac{\alpha_1(t;1)}{\alpha_{10}(t)}\right] = \beta_1 + [a-\Psi(1)]t + \log\left[\frac{\Psi(1)}{a}\times\frac{a_2+a_1\exp(-at)}{a_2\exp(\beta_2)+a_1\exp(\beta_1)\exp[-\Psi(1)t]}\right]. \tag{2.15}$$

Les différents profils de la fonction $\gamma(t)$ sont représentés sur la figure 2.2. Cette figure montre clairement que $\gamma(0)$ est bien égal à β_1, mais que dans aucun de ces cas $\gamma(t)$ n'est constante

La figure 2.3 représente différentes fonctions d'incidence cumulée obtenues pour les deux valeurs de la covariable Z. Sur ces quatre graphes, le paramètre β_1 est fixé à 0.5, et seul le paramètre β_2 varie. D'un point de vue pratique, cette figure illustre les résultats déjà énoncés par Gray (1988), à savoir que l'effet de la covariable sur λ_1 peut être très différent de celui sur F_1, en fonction de l'effet de Z sur λ_2.

Il apparait donc que les relations entres ces deux modèles sont non-triviales et que les covariables agissent de manière différente sur les fonctions de risque cause-spécifique et d'incidence cumulée.

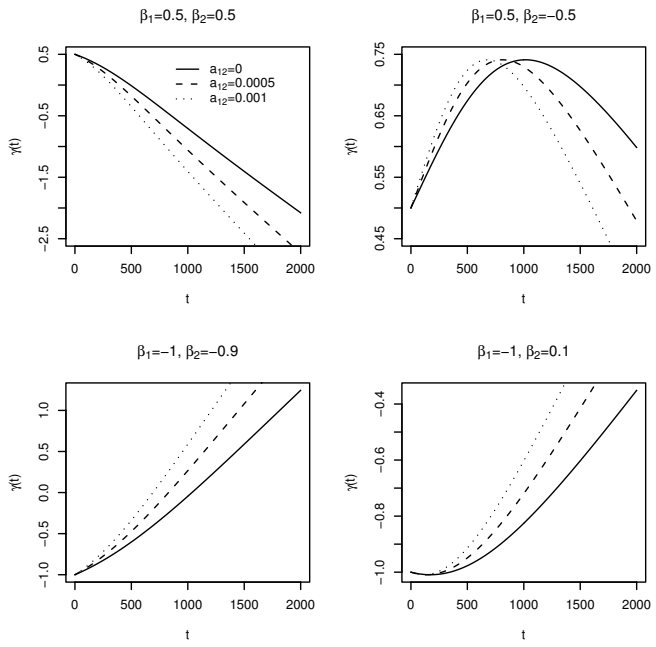


FIG. 2.2 – Tracé de $\gamma(.)$ pour plusieurs valeurs du paramètre de dépendance des temps latents (a_{12}) et pour différentes combinaisons des coefficients de régression du modèle de Cox pour les fonctions de risque cause-spécifique β_1 et β_2.

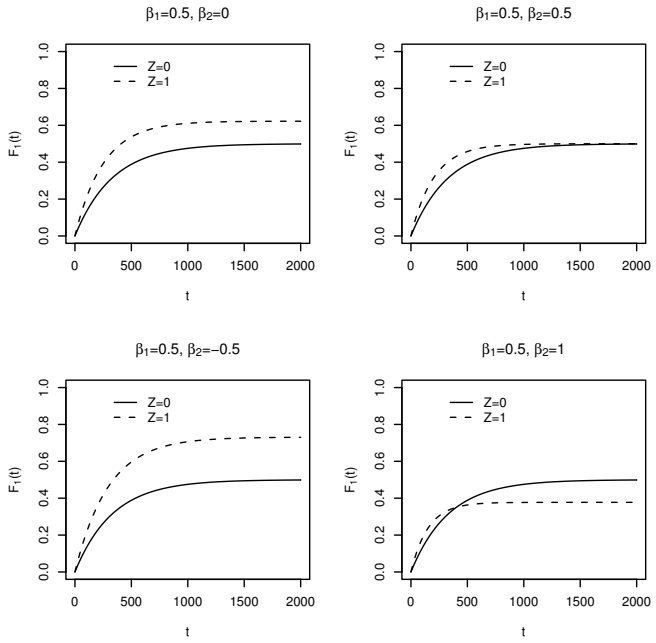


FIG. 2.3 – $F_1(.; Z = 0)$ et $F_1(.; Z = 1)$ pour différentes combinaisons des coefficients de régression β_1 et β_2 du modèle de Cox pour les fonctions de risque cause-spécifique.

Chapitre 3

Implications dans la planification

3.1 Objectifs

En épidémiologie clinique, la planification d'une étude nécessite la justification de la taille de l'échantillon à constituer. Sur un plan statistique, cette justification repose sur le contrôle des risques d'erreur de type I et II. C'est une évidence dans le cadre des essais thérapeutiques randomisés (approche consensuelle internationale sur l'évaluation d'une nouvelle stratégie thérapeutique). C'est moins souvent le cas dans le cadre des études de cohorte à des fins pronostiques. L'objectif de ce chapitre est de permettre le calcul du nombre de sujets nécessaire au contrôle de ces risques d'erreur dans ces deux contextes, en situation de compétition.

3.2 Méthodes

La surestimation de la probabilité brute d'événement d'intérêt en situation de compétition par l'estimateur de Kaplan-Meier est reconnue par tous. Pour tester l'influence d'une covariable Z binaire sur l'événement d'intérêt en situation de compétition, les avis apparaissent plus divergents.

En pratique, dans le calcul de l'effectif à inclure, l'approche la plus fréquente est d'utiliser un modèle de Cox pour la fonction de risque instantané cause-spécifique de l'événement d'intérêt. Le nombre total d'événements d'intérêt nécessaire est obtenu directement à partir du rapport spécifié θ des fonctions de risque instantané cause-spécifique (Schoenfeld, 1983) :

$$e_\theta = \frac{\left(u_{\alpha/2} + u_\beta\right)^2}{(\log\theta)^2\, p(1-p)} \tag{3.1}$$

où u_x est le $100 \times (1-x)^{\text{ème}}$ percentile d'une distribution Gaussienne centrée réduite, et p est la

proportion de patients avec $Z = 1$.

Au contraire, d'autres auteurs proposent d'utiliser une statistique de test comparant les fonctions d'incidence cumulée (Gray, 1988; Pepe and Mori, 1993). Si l'on suppose que la covariable Z agit sur la fonction d'incidence cumulée selon le modèle de Fine & Gray, ce qui revient pratiquement à utiliser le test de Gray (1988) dans le cas de d'une covariable binaire, celui-ci peut servir de base au calcul du nombre d'événements d'intérêt nécessaire. Nous avons montré qu'une formulation similaire du nombre d'événements d'intérêt pouvait être obtenue à partir du rapport γ des fonctions de risque α_i associées aux fonctions d'incidence cumulée. On obtient :

$$e_\gamma = \frac{\left(u_{\alpha/2} + u_\beta\right)^2}{(\log \gamma)^2 \, p(1-p)} \tag{3.2}$$

Dans les deux cas, le nombre n total de sujets à inclure correspondant au nombre d'événements d'intérêt nécessaire se déduit alors directement de $e_.$ selon :

$$n_. = \frac{e_.}{P(\varepsilon = 1)} \tag{3.3}$$

Pour le calcul de n_θ, une estimation de $P(\varepsilon = 1)$ à partir de l'estimation des fonctions d'incidences cumulées a été proposée par Pintilie (2002), sous l'hypothèse de risques exponentiels et indépendants et en supposant des inclusions uniformes.

Pour le calcul de n_γ, nous avons utilisé l'estimation de la fonction d'incidence cumulée de l'événement d'intérêt pour estimer $P(\varepsilon = 1)$, dans la mesure où le modèle de Fine & Gray ne fait pas intervenir les fonctions de risque instantané cause-spécifique. Il est de plus possible d'utiliser les résultats de Lachin and Foulkes (1986) pour prendre en compte une inclusion non uniforme des patients dans l'essai. Enfin, dans le contexte d'une étude de cohorte pronostique, il n'est pas rare qu'il existe une corrélation entre Z, la covariable dont on veut montrer la valeur pronostique sur l'événement d'intérêt, et une autre covariable mesurée Y. Nous avons donc secondairement étendu la formule (3.2) à cette situation :

$$e_\gamma = \frac{\left(u_{\alpha/2} + u_\beta\right)^2}{(\log \gamma)^2 \, p(1-p)(1-\rho^2)} \tag{3.4}$$

où ρ est le coefficient de corrélation entre Z et Y. Ce résultat conduit à une expression symétrique de celle obtenue pour le modèle de Cox en absence de compétition (Schmoor et al., 2000).

Les performances des formules développées ont été étudiées par simulation numérique, y compris dans le cas d'une covariable Y continue.

Enfin, deux exemples ont illustré ces calculs, notamment pour un calcul de puissance.

3.3 "'Sample size formula for proportional hazards modelling of competing risks"'

Cet article a été publié dans *Statistics in Medicine* 2004,**23** :3263-3274

STATISTICS IN MEDICINE
Statist. Med. 2004; **23**:3263–3274
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.1915

Sample size formula for proportional hazards modelling of competing risks

Aurélien Latouche*,†, Raphaël Porcher and Sylvie Chevret

Département de Biostatistique et Informatique Médicale, Hôpital Saint-Louis, Université Paris 7, Inserm Erm 321, Paris, France

SUMMARY

To test the effect of a therapeutic or prognostic factor on the occurrence of a particular cause of failure in the presence of other causes, the interest has shifted in some studies from the modelling of the cause-specific hazard to that of the subdistribution hazard. We present approximate sample size formulas for the proportional hazards modelling of competing risk subdistribution, considering either independent or correlated covariates. The validity of these approximate formulas is investigated through numerical simulations. Two illustrations are provided, a randomized clinical trial, and a prospective prognostic study. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: competing risks; sample size; subdistribution hazard; cumulative incidence

1. INTRODUCTION

In cohort studies, patients are often observed to fail from several distinct causes, the eventual failure being attributed to one cause exclusively to the others, which defines a competing risks situation [1]. In this setting, we are often interested in testing the effect of some covariate on the risk of a specific cause. For example in cancer studies, one could wish to assess the effect of a new treatment or age on delaying relapse, while some patients will die before relapse.

The effect of such a covariate on specific failure types is usually analysed through the modelling of the cause-specific hazard function [2]. However, the interpretation of the effects of the covariate on the specific types of failure is restricted to actual study conditions, and there is no implication that the same effect would be observed under a new set of conditions, notably when certain causes of failure would have been eliminated. Except in circumstances of complete biologic independence among the biological mechanisms giving rise of the various failure types, it is unrealistic to suppose that general statistical methods can be put forward that will encompass all possible mechanisms for cause removal [2]. Notably, the instantaneous

*Correspondence to: A. Latouche, Département de Biostatistique et Informatique Médicale, Hôpital Saint-Louis, 1 avenue Claude Vellefaux, 75475 Paris Cedex 10, France.
†E-mail: aurelien.latouche@chu-stlouis.fr

Received December 2003
Accepted May 2004

risk of specific failure type is sometimes of less interest than the overall probability of this specific failure. Such a probability could be formulated as either the marginal distribution of the specific failure type, or the cumulative incidence function, i.e. the overall probability of the specific type of failure in the presence of competing causes [3–6]. In some situations, the marginal distribution could be the relevant target of estimation. However, this marginal distribution is not identifiable from available data without additional assumptions, such as statistical independence between competing failure types (in which case the marginal hazard equals the cause-specific hazard). In the situation where the competing risks arise from the underlying biology of the disease, and not from the observation process, such an assumption is not clinically relevant. This will be exemplified below on real data (see Section 5). Moreover, it has been proven that this assumption cannot be verified [7]. Therefore, the cumulative incidence functions may appear more relevant than marginal probabilities. However, no one-to-one relationship exists between the cause-specific hazard and the cumulative incidence function. Therefore, in such cases, the emphasis has shifted from the conventional modelling of the cause-specific hazard function to the modelling of quantities directly tractable to the cumulative incidence function [8–10].

Let T be the time of failure, ε the cause of failure ($\varepsilon = 1$, denoting the failure cause of interest) and $F_1(.)$ the cumulative incidence function for failure from cause of interest, i.e. $F_1(t) = \Pr(T \leqslant t, \varepsilon = 1)$. Gray [8] defined the subdistribution hazard for cause 1 as:

$$\lambda_1(t) = \lim_{\mathrm{d}t \to 0} \frac{1}{\mathrm{d}t} \Pr\{t \leqslant T \leqslant t + \mathrm{d}t, \varepsilon = 1 | T \geqslant t \cup (T \leqslant t \cap \varepsilon \neq 1)\}$$

by contrast to the cause-specific hazard:

$$\alpha_1(t) = \lim_{\mathrm{d}t \to 0} \frac{1}{\mathrm{d}t} \Pr\{t \leqslant T \leqslant t + \mathrm{d}t, \varepsilon = 1 | T \geqslant t\}$$

By construction, the subdistribution hazard $\lambda_1(t)$ is explicitely related to the cumulative incidence function for failure from cause 1, $F_1(t)$, through

$$\lambda_1(t) = -\mathrm{d}\log\{1 - F_1(t)\}/\mathrm{d}t \tag{1}$$

while the relation between the cause-specific hazard $\alpha_1(t)$ and $F_1(t)$ involves the cause-specific hazards of all failures types [11].

A semi-parametric proportional hazards model was proposed to test covariate effects on the subdistribution hazard [9]. Using the partial likelihood principle and weighted estimated equations, consistent estimators of the covariate effects were derived, either in the absence of censoring (referred as 'complete data'), or in the presence of administrative censoring, i.e. when the potential censoring time is observed on all individuals ('censoring complete data'). A weighted estimator for right-censored data ('incomplete data') was also proposed, properties of which were investigated through numerical simulations.

This paper focuses on computing sample size for the Fine and Gray's model [9], with two main goals. First, we provide a sample size formula to design a parallel arm randomized clinical trial with a right-censored competing endpoint, contrasting this computation with the standard Cox modelling of the cause-specific hazard [12]. Secondly, we compute the required sample size (or statistical power) to detect a relevant effect in a prognostic study, when dealing with a competing risks outcome.

Section 2 of this paper describes the sample size formula for evaluation of a therapeutical effect. In Section 3, we extend this sample size formula to the prognostic situation, where the prognostic factor of interest is possibly correlated with another factor. The validity of the resulting approximate asymptotic formulas is investigated with respect to their finite sample behaviour by numerical simulations in Section 4. Our approach is illustrated by two examples, one is the randomized clinical trial, and the other is a prognostic study. We close the paper with some discussion.

2. SAMPLE SIZES FOR EVALUATION OF THERAPEUTIC EFFECT IN THE COMPETING RISK SETTING

Assume a randomized clinical trial is designed to compare two treatments with respect to a competing risks endpoint. Usually, one treatment is a standard treatment or a placebo, whereas the other treatment is an experimental treatment.

Let X be a binary covariate representing the treatment group (with $X=1$ denoting the experimental group, and $X=0$ the standard group) and Y, any additional covariate, independent of X. Let p, be the proportion of patients randomly allocated to the experimental treatment group, and n the required sample size of the trial. We wish to test the benefit of the experimental group over the standard group with regards to the occurrence of the failure of interest, either roughly or adjusted on Y. We extend Schoenfeld's sample size formula [13], developed in the conventional survival case, to the competing risks setting.

We assume the Fine and Gray's model [9] for the subdistribution hazard of failure times of interest,

$$\lambda_1(t;X,Y)=\lambda_0(t)\,\exp(aX+bY)$$

In the case of 'complete data', the partial likelihood of the model is:

$$L(\beta)=\prod_i^n\left[\frac{\exp(ax_i+by_i)}{\sum_{j\in\mathscr{R}_i}\exp(ax_j+by_j)}\right]^{I(\varepsilon_i=1)}$$

which, besides the definition of the risk-set $\mathscr{R}_i=\{j:(T_i\leqslant T_j)\cup(T_j\leqslant T_i\cap\varepsilon_j\neq 1)\}$, is similar to the Cox partial likelihood. Fine and Gray [9] have shown that the Wald (partial) statistic to test the null hypothesis $\{a=0\}$ is $\sqrt{n}\hat{a}$ where $\hat{a}$ is the maximum partial likelihood estimate (MPLE) of a, which is asymptotically Gaussian with zero mean and estimated variance $\mathscr{V}_0$. Using the same terminology as in Reference [13, p. 502], this variance expresses:

$$\mathscr{V}_0=\frac{n}{\sum_{i\in\mathscr{E}}M_i(x,\hat{b})\{1-M_i(x,\hat{b})\}-\{\sum_{i\in\mathscr{E}}M_i(xy,\hat{b})-M_i(x,\hat{b})\times M_i(y,\hat{b})\}^2/\sum_{i\in\mathscr{E}}M_i(y,\hat{b})\{1-M_i(y,\hat{b})\}}$$

with $M_i(x,b)=\sum_{j\in\mathscr{R}_i}x_j\exp(by_j)/\sum_{j\in\mathscr{R}_i}\exp(by_j)$ and $\mathscr{E}=\{i:\varepsilon_i=1\}$ and $\hat{b}$ is the MPLE of b.

Assuming that the variance of the Wald statistic under the alternative hypothesis is approximately equal to $\mathcal{V}_0$, as in Reference [14, p. 442], it is straightforward that:

$$\mathcal{V}_0 \approx \frac{n}{e\,p(1-p)} \tag{2}$$

where e is the number of failures of interest.

Of note, the subdistribution hazard ratio, $\theta = \exp(a)$, can be expressed as

$$\theta = \frac{\log\{1 - \mathbf{F}_1(\mathbf{t}; \mathbf{X} = 0, \mathbf{Y})\}}{\log\{1 - \mathbf{F}_1(\mathbf{t}; \mathbf{X} = 1, \mathbf{Y})\}}$$

To determine the sample size for this trial, we first calculate the number e of failures of interest required to control for both type I and type II error rates of α and β, respectively, as follows:

$$e = \frac{(u_{\alpha/2} + u_\beta)^2}{(\log\theta)^2\, p(1-p)} \tag{3}$$

where u_γ denotes the $(1-\gamma)$-quantile of the standard Gaussian distribution.

Since the 'censoring complete data' analysis relies on a proper partial likelihood, previous results are readily inherited from classical Cox model with censoring [9, p. 499]. Formula (3) thus holds in the cases of 'complete data' and 'censoring complete data'.

The total required sample size is,

$$n = \frac{(u_{\alpha/2} + u_\beta)^2}{(\log\theta)^2\, p(1-p)\psi} \tag{4}$$

where ψ is the proportion of failures of interest at the time of analysis, T_a.

Formula (4) looks similar to that derived by Schoenfeld [13] for the survival Cox regression model, with θ, the subdistribution hazard ratio instead of the hazard ratio. Nevertheless, since the effect of a covariate on the cause-specific hazard for a particular failure can be quite different than its effect on the subdistribution function [8, 15], actually, formulas are different.

Finally, note that, in the absence of any censoring, the proportion ψ of failures of interest (that is identical to the proportion of non-censored observations in the Schoenfeld's formula [13]) reduces to the value at the time of analysis of the subdistribution function for the failure of interest in the whole cohort, $F_1(T_a)$. In the case of 'censoring complete data', ψ can be estimated roughly by $(1-c)F_1(T_a)$, where c is the expected proportion of censored observations.

3. SAMPLE SIZES FOR THE EVALUATION OF PROGNOSTIC FACTORS FOR COMPETING RISKS OUTCOMES

We now consider the planning of a prospective cohort study aiming at assessing whether or not a particular exposure is associated with the subsequent occurrence of the failure of interest. Let X be a binary covariate representing the exposure (with $X = 1$ if exposed and $X = 0$ otherwise), and Y another prognostic covariate. For simplicity, we assume that Y is

binary. Let $p = \Pr(X = 1)$, $q = \Pr(Y = 1)$, $p_0 = \Pr(X = 1 \mid Y = 0)$ and $p_1 = \Pr(X = 1 \mid Y = 1)$. The correlation coefficient of X and Y, ρ, is

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{p(1-p)q(1-q)}} = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}}$$

Using the same approximation as Schmoor *et al.* [14], we obtain,

$$n = \frac{(u_{\alpha/2} + u_{\beta})^2}{(\log \theta)^2 p(1-p)\psi(1-\rho^2)} \tag{5}$$

Mathematical details are given in Appendix A. Formula (5) is similar to formulas derived by Schmoor *et al.* [14] for the Cox model, and by Hsieh *et al.* [16] for linear and logistic regression models, with the same variance inflation factor $1/(1-\rho^2)$.

As in the independent case, formula (5) holds in the cases of 'complete data' and 'censoring complete data'. In the case of 'incomplete data' (i.e. in case of right censoring), Fine and Gray [9] showed that the variance of the estimator was very close to that based on the 'censoring complete data', suggesting that both our proposed formulas (4) and (5) could apply in such a case. Hence, we decided to perform a simulation study to assess the validity of the sample size formulas for right-censored data in finite samples.

4. SIMULATION STUDY

In this section, we present the results of numerical investigations. In each set of simulations, we considered a failure cause of interest and a competing cause of failure, and two possibly correlated binary covariates (X, Y).

Failure times data were generated through the method described by Fine and Gray [9]. Briefly, the subdistribution for the failures of interest are given by,

$$\Pr(T_i \leqslant t, \varepsilon_i = 1; X_i, Y_i) = 1 - [1 - \Pr(\varepsilon = 1; X_i = 0, Y_i = 0)(1 - \exp(-t))]^{\exp(aX_i + bY_i)}$$

which is a unit exponential mixture with mass $1 - \Pr(\varepsilon = 1; X, Y)$ at ∞ when $(X, Y) = (0, 0)$, and uses the proportional subdistribution hazards model to obtain the subdistribution for non-zero covariate values. The subdistribution for the competing risks failure cause was obtained using an exponential distribution with rate $\exp(a_2 X_i + b_2 Y_i)$. Covariate values of (X, Y) were generated from a bivariate Bernoulli process with parameters p, q, and ρ defined above.

We first assumed that there was no censoring. Next, we considered 'incomplete data' with right-censoring times generated from Uniform distribution, to reach an average of 30 per cent of censored observations. In each simulated set, we generated an n-sample of subjects, where n was computed according to equation (5) for predetermined $\alpha = 0.05$ and $\beta = 0.20$. We took $\Pr(\varepsilon = 1; (X, Y) = (0, 0)) = 0.5$, $\psi = (1 - c)F_1(Ta) = 0.2$, $\theta \in \{1, 1.5, 2, 3, 4\}$, $b = 1$, $a_2 = b_2 = 1$, $p = q = 0.5$, and $\rho \in \{0, 0.2, 0.3, 0.4\}$.

In each situation, a total of 10 000 independent data sets were generated, with estimation of the actual level and power of the Wald test for $H_0 : \{a = 0\}$. Moreover, the respective effects of ρ , θ and b with other parameters fixed were investigated.

Table I displays the observed level and power as obtained from the 10 000 simulations, together with computed sample size, corresponding to several combinations of parameters ρ,

Table I. Sample size for nominal type I error rate 0.05 and power 0.80, observed level and power of the partial Wald test according to the correlation ρ between both covariates, the subdistribution hazard ratio θ and the censoring rate.

ρ	$\theta = \exp(a)$	Observed level	Observed power	n
No censoring				
0	1.5	0.0525	0.8010	382
0.2		0.0520	0.8083	398
0.4		0.0519	0.8032	455
0	2	0.0445	0.8311	131
0.2		0.0461	0.8322	137
0.4		0.0483	0.8302	156
0	3	0.0512	0.8624	53
0.2		0.0498	0.8652	55
0.4		0.0535	0.8619	62
0	4	0.0520	0.8771	33
0.2		0.0530	0.8868	35
0.4		0.0640	0.8729	39
30 per cent censoring				
0	1.5	0.0488	0.7848	546
0.2		0.0496	0.7804	569
0.4		0.0506	0.7853	650
0	2	0.0461	0.8169	187
0.2		0.0519	0.8175	195
0.4		0.0502	0.8242	223
0	3	0.0485	0.8629	75
0.2		0.0495	0.8555	78
0.4		0.0500	0.8541	89
0	4	0.0417	0.8683	47
0.2		0.0531	0.8728	49
0.4		0.0560	0.8715	56

θ, in the uncensored and 30 per cent-censored cases. In the uncensored setting, results show that our formula performs well for small values of the subdistribution hazard ratio θ, while it overpowers the study for increased values of θ. The level of the Wald test was found to be approximately equal to the nominal level of 5 per cent, except for small sample sizes corresponding to high values of θ. Results were very similar in the censored case, with a relatively slight decrease in power as compared to the uncensored case.

The more specific effects of the formula's parameters (ρ and a) as well as that of the other regression parameter b are detailed in Figures 1 and 2. For fixed covariates effect ($a = \log(2)$ and $b = -a$), Figure 1 illustrates that the formula accounts well for a non-zero correlation ρ. Both the observed level and power remained stable around their nominal values for a range of correlation coefficient values ranging from -0.9 to 0.9.

The left plot of Figure 2 represents the observed level and power when a varies, while b is set to zero. Results show that our formula correctly estimates the sample size for small (absolute) values of a. For large negative values of a, however, the computed sample size leads to under-powered studies whereas for large positive values of a, the studies are over-powered. The right plot of Figure 2 displays observed level and power when b varies, when $a = \log(2)$ and $\rho = 0.2$. As expected, when the effect of Y is in the same direction as that

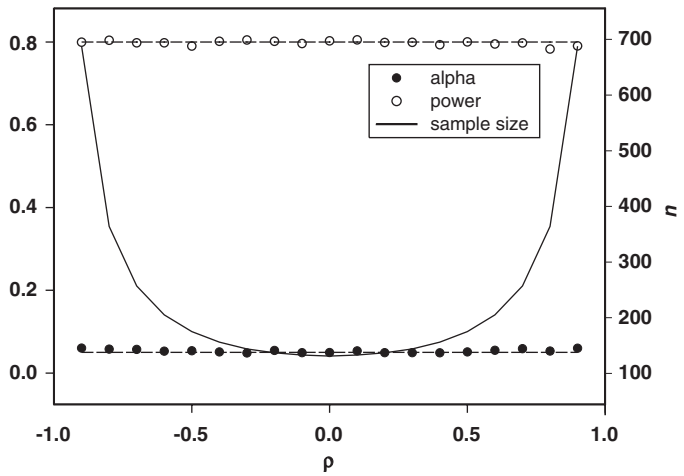


Figure 1. Sample size, observed level and power, according to the correlation coefficient ρ between X and Y.

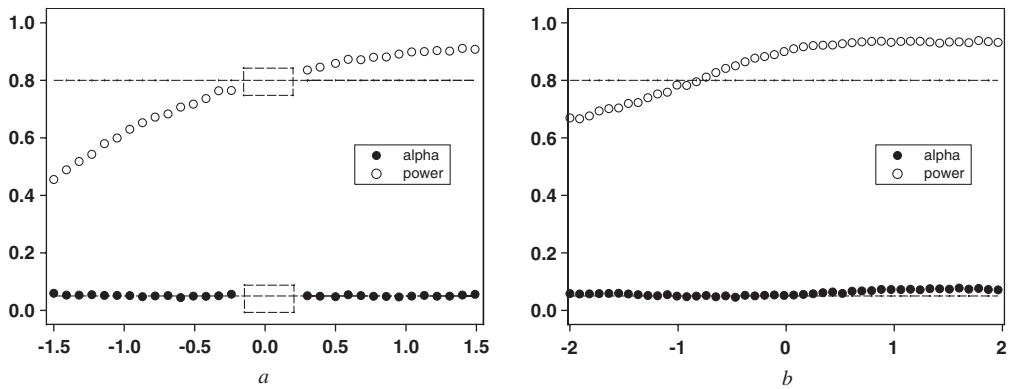


Figure 2. Observed level and power according to the values of the regression coefficients a and b associated with the covariate of interest and the other covariate, respectively.

of X, the power of the trial is higher than its nominal value, and opposite effects of both correlated variables (a situation which is however rather theoretical) lead to a decrease in statistical power.

5. EXAMPLES

The sample size formulas presented in the previous sections are illustrated by two examples.

Table II. Comparison of sample size computation for a clinical trial comparing an experimental (E) and a standard (S) treatment groups, based on estimates of F_1 computed either from $1-\widehat{\mathrm{KM}}$ or $\widehat{\mathrm{CIF}}$.

T_a (h)	Estimates of F_1	S	E	$\theta=\frac{\log(1-\hat{F}_E(T_a))}{\log(1-\hat{F}_S(T_a))}$	e	ψ (per cent)	n
12	$1-\widehat{\mathrm{KM}}$	0.20	0.30	1.6	191	25.0	218
	$\widehat{\mathrm{CIF}}$	0.20	0.30	1.6	191		218
24	$1-\widehat{\mathrm{KM}}$	0.60	0.75	1.5	245	67.5	364
	$\widehat{\mathrm{CIF}}$	0.60	0.67	1.2	1160		1596
48	$1-\widehat{\mathrm{KM}}$	0.865	0.90	1.2	1357	87.75	1546
	$\widehat{\mathrm{CIF}}$	0.77	0.78	1.03	47339		70131

5.1. Planning a randomized clinical trial

We retrospectively redesigned a phase III randomized clinical trial, to illustrate the use of our proposed sample size formula contrasting with the use of the approach based on the cause-specific hazard ratio.

This double-blind randomised clinical trial was conducted to compare the efficacy in the induction of labour of vaginal misoprostol (experimental arm, $X=1$) with vaginal dinoprosone (standard arm, $X=0$) [17]. The primary outcome of the trial was vaginal delivery within 24 h while time to vaginal delivery defined a secondary outcome. A total of 370 women were enrolled.

Of note, in this setting, caesarian sections act as competing risks outcomes. To measure the effect of misoprostol, the cause-specific hazard, that is the instantaneous risk of vaginal delivery, is of less interest than the overall probability of vaginal delivery, which quantifies the overall success rate of the method used for labour induction.

Analysis of the trial was based on the complement of the Kaplan–Meier survival estimates (denoted $1-\widehat{\mathrm{KM}}$) of the outcome in each treatment group. Since it has been shown that caeserian sections are related to prolonged labour, i.e. when vaginal birth is not possible or not safe for the mother or the child, both risks are obviously not independent, and marginal probabilities could not be validly estimated. We thus computed nonparametric estimates of the cumulative incidence functions (denoted $\widehat{\mathrm{CIF}}$), treating caesarian sections as competing risks outcomes. We then attempted to show how the use of either approach would have modified the computed sample size of the trial.

Therefore, on the basis of these estimates, we computed the sample size in each setting accordingly, using three possible time points for the outcome that could have been chosen by the investigator to plan a new trial (vaginal delivery at either $T_a=12,24$ or 48 h). Table II displays the main results of these computations. The cause-specific hazard ratio differed from the subdistribution hazard ratio due to differences between $1-\widehat{\mathrm{KM}}$ and $\widehat{\mathrm{CIF}}$ (Figure 3). These differences result in differences in sample size computations. For instance, as reported in Table II, estimates of vaginal delivery rate at 24 h yield to a sample size of 364 based on $1-\widehat{\mathrm{KM}}$ and 1596 based on $\widehat{\mathrm{CIF}}$.

5.2. Posterior power assessement in a cohort study

The second example is based on a cohort study involving 107 patients with acute or chronic leukemia who underwent allogeneic stem cell transplantation, that was conducted to identify

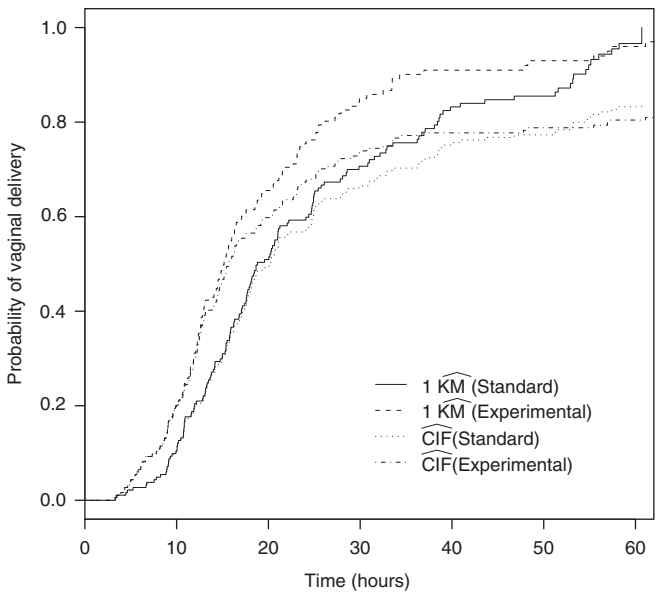


Figure 3. Estimation of the cumulative incidence of vaginal delivery in both randomized groups according to the statistical handling of caesarian sections: either censored using the 1 minus Kaplan–Meier estimate ($1 - \widehat{KM}$), or considered as a competing event in cumulative incidence function estimation ($\widehat{CIF}$).

prognostic factors for early outcomes including infections, hematological recovery, acute graft versus host disease (aGVHD) and survival [18]. We focused on one particular endpoint, that is aGVHD.

Statistical analysis was initially based on the Fine and Gray model, considering death prior to aGVHD as a competing risk event. Indeed, from a clinical point of view, the overall probability of developing aGVHD was found more meaningful than the instantaneous risk of aGVHD, owing to the short exposure period for aGVHD (restricted to the first 100 days post-transplant). Marginal probability of aGVHD could have been of interest if the removal of aGVHD was expected to have no effect on failure rates for the remaining causes such as death prior to aGVHD. In our setting, this was unlikely and we focused on the cumulative incidence functions, and subdistribution hazards.

The study identified both female donor to male recipient and a particular gene polymorphism for the donor interleukin-1 (IL-1) cytokine as predictive of aGVHD. More precisely, the estimated subdistribution hazard ratio for both variables was 1.91 (95 per cent confidence interval 1.05–3.47) and 2.07 (95 per cent confidence interval 1.09–3.91), respectively. Otherwise, the analysis failed to identify any additional prognostic factor for aGVHD.

However, investigators would have expected a prognostic effect of interleukin-6 (IL-6) gene polymorphism, approximately of the same magnitude as that of IL-1 gene polymorphism. We wondered whether this could not be explained by a lack of statistical power.

Let X be a binary covariate, denoting the presence of donor IL-6 gene polymorphism, and Y be a discrete score (taking four distinct values) defined as a linear combination of the two

binary variables female donor to male recipient and donor IL-1 gene polymorphism. Since the variance inflation factor still holds in the Cox model with non-binary covariates [19], we decided to apply formula (5) and compute statistical power to detect a subdistribution hazard ratio of 2. The correlation coefficient ρ, between X and Y in equation (5) using estimated regression coefficients was estimated at 0.132 from the data.

Given the proportion of subjects with IL-6 gene polymorphism of $p = 0.39$, $\psi = 0.505$, inverting formula (5) yielded that the study had a 69 per cent power to detect a subdistribution hazard ratio of 2.

In other words, if one wish to plan a future study, a sample size of $n = 139$ (resp, $n = 186$) patients would be necessary to reach a power of 80 per cent (resp, 90 per cent).

6. DISCUSSION

Computation of sample size is important in designing experiments. In cohort studies, analysis is frequently complicated by the presence of competing risks. However, despite the increasing literature devoted to competing risks [1, 3–6, 8, 9, 15, 20], no specific method for sample size computation has been proposed, besides that based on cause-specific hazards [21]. As mentioned above, the use of cause-specific hazards in the competing risks setting can be restrictive, so that developing sample size formulas based on comparison of subdistribution hazards was of prime interest. Two situations of planning were considered, a randomized clinical trial and a prognostic cohort study, since both may have to deal with competing risks outcomes and to focus on cumulative incidence functions rather than cause-specific hazards.

In both situations, the main objective of the study was to estimate the effect of a therapeutic or prognostic factor, respectively. We proposed to test such effects on the subdistribution hazard, using the proportional hazards model proposed by Fine and Gray [9]. Estimation in this model from uncensored data is easily provided by using a Cox model where all competing risks failures have been censored at $+\infty$ [11]. Moreover, due to the structure of the partial likelihood for the subdistribution, the results provided in the setting of the standard Cox model [13, 14] were adapted in the current set-up. Of note, several assumptions were required. Notably, formulas (4) and (5) only hold for contiguous alternatives. Nevertheless, similar results have been stressed for the standard Cox model [13, 14]. Finally, the formulas only apply without right censoring at least theoretically. Indeed, as reported in the simulation study, the sample size formula for right-censored data could be reasonably approximated by that for the 'censoring complete data'. This allows a greater applicability of the formulas.

Actually, when computing sample size for time-to-failure data in the presence of competing risks, a meaningful expected covariate effect on the subdistribution hazard ratio must be specified. Nevertheless, the subdistribution hazard does not have a natural interpretation. Hence, in the examples, we formulated the effect size directly in terms of cumulative incidences functions instead of subdistribution hazards, which may be easier to interpret. This also exemplifies the fundamental difference between both hazard ratios, and the reasons why sample sizes computation may be affected by the model choice. As illustrated in the first example, we showed that a covariate may have a non-negligible influence on the cause-specific hazard (with estimated $\theta = 1.2$), but no effect on the cause-specific subdistribution (with estimated subdistribution hazard ratio $\theta = 1.03$), as previously reported [9].

Since the first aim of the paper was to provide a sample size formula for clinical trials, we investigated the situation of two independent binary factors. The second objective of the paper was to deal with the planning of prognostic studies, which is often overlooked leading to inconsistency of prognostic studies [14]. Therefore, we extended sample size formulas to handle for multiple models with correlated covariates. For simplicity, we considered binary covariates and we ignored the accrual pattern. Nevertheless, this could be easily extended to more general situations, as previously done by Hsieh and Lavori [19] and Lachin and Foulkes [22], respectively.

APPENDIX A

Under $H_0 : \{a = 0\}$, it is possible to approximate $M_i(x)$ by $\eta p_1 q_i + p_0(1 - q_i)/\eta q_i + (1 - q_i)$, where $\eta = \exp(b)$ and q_i is the probability of Y equal to 1 at time t_i. Similarly, we approximate $M_i(y)$ as $\eta q_i/\eta q_i + (1-q_i)$ and $M_i(xy)$ as $\eta p_1 q_i/\eta q_i + (1-q_i)$. This yields to $M_i(xy) = p_1 M_i(y)$ and $M_i(x) = p_1 M_i(y) + p_0(1-M_i(y))$. Thus, $M_i(xy) - M_i(x)M_i(y) = (p_1 - p_0)M_i(y)[1-M_i(y)]$ and

$$\mathscr{V}_0 = \frac{n}{\sum_{i\in\mathscr{E}}[M_i(y)(p_1 - p_0)(1 - p_1 - p_0) + p_0(1 - p_0)]}$$

As $\sum_{i\in\mathscr{E}} M_i(y)$ approximately equals to $q \times e$, $\mathscr{V}_0$ reduces to $\mathscr{V}_0 = n/e\,p(1 - p)(1 - \rho^2)$. The latter is equal to the variance in the independent case multiplied by the variance inflation factor $1/(1 - \rho^2)$.

To illustrate that $\sum_{i\in\mathscr{E}} M_i(y) \approx q \times e$, we considered that time to failures from the cause of interest were exponentially distributed, i.e. $\lambda_{10}(t) = \lambda$, as did Schmoor *et al.* [14]. We also assumed that there is no censoring. Under this assumption, the subdistribution function is

$$F_1(t) = 1 - q\,\exp(-\lambda\eta t) - (1 - q)\exp(-\lambda t)$$

with derivative with respect to t, $f(t) = \lambda[q\eta\,\exp(-\lambda\eta t) + (1 - q)\exp(-\lambda t)]$
It follows that,

$$q_i = \frac{q\,\exp(-\lambda\eta t_i)}{q\,\exp(-\lambda\eta t_i) + (1 - q)\exp(-\lambda t_i)}$$

and

$$\mathrm{e}^{-1}\sum_{i\in\mathscr{E}} M_i(y) \to \int_0^\infty \frac{q\eta\exp(-\lambda\eta t)}{q\eta\exp(-\lambda\eta t) + (1 - q)\exp(-\lambda t)} f(t)\,\mathrm{d}t$$

leading to $\sum_{i\in\mathscr{E}} M_i(y) \to q \times e$.

ACKNOWLEDGEMENTS

We are grateful to Doctor Patrick Rozenberg and Professor Eliane Gluckman for providing access to data. We also want to thank the referees for their helpful comments that led to important improvements.

REFERENCES

1. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. Wiley: New York, 1980.
2. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics* 1978; **34**:541–554.
3. Aalen OO. Nonparametric estimation of partial transition probabilities in multiple decrement models. *The Annals of Statistics* 1978; **6**:534–545.
4. Benichou J, Gail MH. Estimates of absolute cause-specific risk in cohort studies. *Biometrics* 1992; **46**: 813–826.
5. Korn EL, Dorey FJ. Applications of crude incidence curves. *Statistics in Medicine* 1992; **11**:813–829.
6. Pepe MS, Mori M. Kaplan–Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine* 1993; **12**:737–751.
7. Tsiatis AA. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* 1975; **72**:20–22.
8. Gray RJ. A class of *k*-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics* 1988; **116**:1141–1154.
9. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**:496–509.
10. Klein JP, Andersen PK. Regression modelling of competing risks data based on pseudo-values of the cumulative incidence function. *Technical Report*, Department of Biostatistics, University of Copenhagen, 2003.
11. Andersen PK, Abildstrøm SZ, Rosthøj S. Competing risks as a multi-state model. *Statistical Methods in Medical Research* 2002; **11**:203–215.
12. Cox DR. Regression models and life tables. *Journal of the Royal Statistical Society*, *Series A* 1972; **34**: 187–220.
13. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics* 1983; **39**: 499–503.
14. Schmoor C, Sauerbrei W, Schumacher M. Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine* 2000; **19**:441–452.
15. Pepe MS. Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association* 1991; **86**:770–778.
16. Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine* 1998; **17**:1623–1634.
17. Rozenberg P, Chevret S, Goffinet F, Durand-Zaleski I, Ville Y, Vayssière C, Roberto A, Lahna Z, Nisand I, Fisch C, Chaumet-Riffaud P, Chastang C. Induction of labour with a viable infant: a randomised clinical trial comparing intravaginal misoprostol and intravaginal dinoprostone. *British Journal of Obstetrics and Gynaecology* 2001; **108**:1255–1262.
18. Rocha V, Franco R, Porcher R, Bittencourt H, Silva Jr W, Latouche A, Devergie A, Espérou H, Ribaud P, Socié G, Zago M, Gluckman E. Host defense inflammatory gene polymorphism are associated with outcomes after HLA-identical sibling bone marrow transplantation. *Blood* 2002; **100**:3908–3918.
19. Hsieh FY, Lavori PW. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials* 2000; **21**:552–560.
20. Fine JP. Regression modelling of competing crude failure probabilities. *Biostatistics* 2001; **2**:85–97.
21. Pintilie M. Dealing with competing risks: testing covariates and calculating sample size. *Statistics in Medicine* 2002; **21**:3317–3324.
22. Lachin JM, Foulkes MA. Evaluation of sample size and power for analysis of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics* 1986; **42**:507–519.

Chapitre 4

Propriétés du modèle de Fine & Gray pour une fonction de risque mal spécifiée

4.1 Objectifs

Nous avons vu dans le Chapitre 2 les relations existant entre les fonctions de risque instantané qui peuvent être modélisées lorsque l'on s'intéresse à l'évaluation de l'influence d'une covariable Z sur un événement d'intérêt en situation de compétition. En premier lieu, il est nécessaire de supposer que cette covariable agit préférentiellement sur l'une de ces fonctions de risque instantané. En effet, il apparait impossible que l'effet de la covariable soit multiplicatif sur ces deux fonctions sauf cas dégénérés (absence d'effet sur tous les risques ou absence de compétition). En d'autres termes, au mieux, l'un des deux modèles peut être postulé (Andersen et al., 2002). Dans la mesure où un effet multiplicatif sur la fonction de risque cause-spécifique apparaît plus réaliste, on s'est interrogé sur les propriétés de l'estimateur de l'effet dans un modèle de Fine & Gray dans cette situation de modèle mal spécifié.

4.2 Méthodes

La figure (4.1) illustre, sur un exemple donné par Gray, la différence de l'effet d'une covariable sur la fonction de risque instantané cause-spécifique et la fonction d'incidence cumulée en présence de 2 types d'événements, avec des risques cause-spécifiques constants ($\lambda_{11} = \lambda_{21} = 3$ groupe 1,

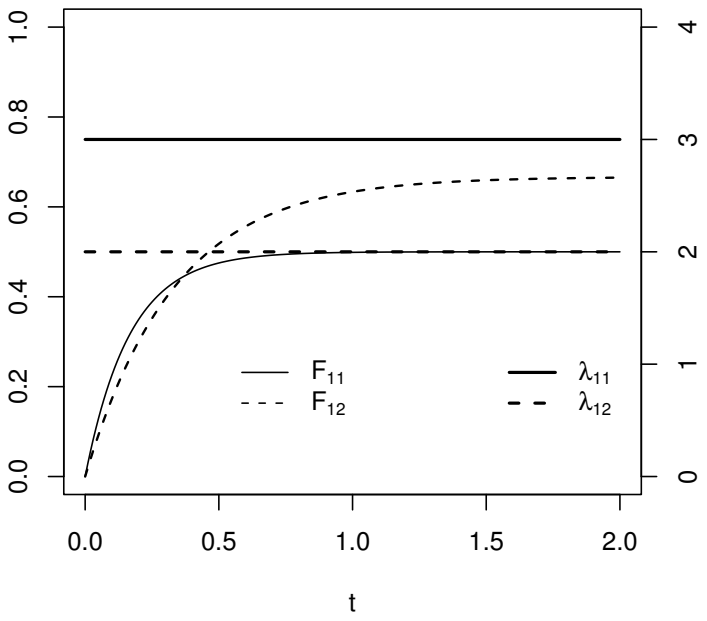


FIG. 4.1 – Incidence cumulée et risque cause-spécifique pour l'événement de de type en 1 (groupe 1 (plein) et groupe 2 (pointillé))

$\lambda_{22} = 1$ et $\lambda_{12} = 2$ groupe 2). Les incidence cumulées sont donc $F_{11}(t) = (1 - \exp(-6t))/2$ dans le groupe 1 et $F_{12} = 2(1 - \exp(-3t))/3$. Pour $t \geq \log(3)/3$ $F_{11}(t) < F_{12}(t)$ alors que $\lambda_{11} > \lambda_{12}$. Nous avons choisi de postuler un modèle de Cox, où β_1 est l'effet de la covariable Z sur le risque cause-spécifique de l'événement d'intérêt et β_2, son effet sur le risque cause-spécifique de l'événement en compétition.

D'une part, le modèle de Cox est largement utilisé dans la littérature. D'autre part, il semble légitime de supposer que la covariable Z modifie intrinsèquement la probabilité d'événement d'intérêt, par exemple en représentant un traitement qui diminue la masse tumorale, et ainsi le risque de rechute, à chaque instant. Au contraire, un effet de la covariable Z sur la fonction d'incidence cumulée représente une combinaison de cet effet intrinsèque et de son effet sur les autres événements en compétition. Par exemple, dans un modèle à deux événements, un traitement qui supprimerait complètement l'un des événements augmenterait nécessairement la probabilité de l'autre, quelque soit son mode d'action. L'analyse inverse n'a donc pas été effectuée.

Nous avons montré que, dans ces conditions, l'estimateur $\hat{\gamma}$ de l'effet de Z dans le modèle de Fine & Gray convergeait vers une quantité γ^*, pour laquelle une solution approchée à l'ordre 1

a été calculée :

$$\gamma* \simeq c_0 + c_1\beta_1 \tag{4.1}$$

Une solution explicite de cette approximation a été obtenue (secondairement), en absence de censure et dans le cadre d'un modèle paramétrique, le modèle ACBVE pour les temps latents d'événements (Block and Basu, 1974). Cette formulation explicite la valeur de γ^* comme une fonction affine de β_1, l'ordonnée à l'origine dépendant à la fois de β_2 et de la variance de Z, alors que la pente (ne) dépend (que) de β_2 et $P(\varepsilon = 1)$. A noter que le modèle ACBVE permet de prendre en compte une éventuelle corrélation entre ces temps latents ; en cas de temps latents indépendants, on retrouve le cas exponentiel indépendant souvent utilisé (voir Pintilie (2002), Chapitre 3).

La qualité de cette approximation a été étudiée par simulation, en particulier lorsque β_1 s'éloigne de zéro.

4.3 "'Misspecified regression model for the cumulative incidence function"'

Cet article a été soumis pour publication à *Biometrics*.

Misspecified regression model for the cumulative incidence function

A. Latouche,[1,*] **V. Boisson,**[2] **S. Chevret**[1] **and R. Porcher**[1]

[1]Département de Biostatistique et Informatique Médicale, Hôpital Saint-Louis,
Université Denis Diderot, Inserm Erm 321, Paris, France

[2]Laboratoire de Statistique Théorique et Appliquée
Université Pierre et Marie Curie, Paris, France

SUMMARY. We considered a competing risks setting, when evaluating the prognostic influence of an exposure on a specific cause of failure. Two main regression models are used in such analyses, the Cox cause-specific proportional hazards model and the subdistribution proportional hazards model. We examine the properties of the estimator based on the latter model when the true model is the former. An explicit relationship between subdistribution hazards ratio and cause-specific hazards ratio is derived, assuming a parametric distribution for latent failure times.

KEY WORDS: Model misspecification; Cumulative incidence; Proportional hazards.

1. Introduction

In the competing risks setting, subjects may fail from distinct and exclusive causes. Thus, observed data typically consist in both a failure time $T \geq 0$ and a failure cause $\epsilon \in \{1, \ldots, K\}$, which is unobserved if T is right-censored (this will be denoted further $\epsilon = 0$). Consider that we are interested in estimating the effect of an exposure, Z, on a particular cause of failure, identified by $\epsilon = 1$.

To examine the effect of an exposure on a particular cause of failure in the setting

* *email:* aurelien.latouche@chu-stlouis.fr

of competing risks, two main approaches are used (Andersen et al., 2002). The most common approach is to focus on the modelling of the cause-specific hazard of this failure cause (Prentice et al., 1978), widely through the use of a Cox model. The second approach is to compare the cumulative incidence functions (CIF) of this failure cause between exposed and unexposed groups, either directly (Gray, 1988; Pepe, 1991), or by modelling the hazard function associated with the CIF, the so-called subdistribution hazard (Fine and Gray, 1999; Fine, 2001).

Let us denote the CIF for failure of cause k by:

$$F_k(t) = \Pr(T \leq t, \epsilon = k), \tag{1}$$

and the marginal survival function:

$$S(t) = \Pr(T > t) = 1 - \sum_k F_k(t).$$

For simplicity, the $F_k(t)$ are assumed to be continuous with subdensities $f_k(t)$ (with respect to Lebesgue measure). The cause-specific hazard function is defined by:

$$\lambda_k(t) = \lim_{\delta t \to 0} \frac{1}{\delta t} \Pr(t \leq T < t + \delta t, \epsilon = k | T \geq t) = f_k(t)/S(t), \tag{2}$$

while the subdistribution hazard (Gray, 1988) is defined by:

$$\alpha_k(t) = \lim_{\delta t \to 0} \frac{1}{\delta t} \Pr(t \leq T < t+\delta t, \epsilon = k | T \geq t \cup (T < t \cap \epsilon \neq k)) = f_k(t)/[1-F_k(t)]. \tag{3}$$

All these functions are identifiable from observations (Tsiatis, 1975; Prentice et al., 1978). To relate the cause-specific hazard (Prentice et al., 1978) on the exposure covariate Z, the Cox proportional hazards model (Cox, 1972) is often used while a similar model was proposed for the subdistribution hazard (Fine and Gray, 1999). From equations (2) and (3), it is clear that the subdistribution hazard can be directly obtained

from the CIF, whereas the relationship between the cause-specific hazard and the CIF involves the marginal survival function, *i.e.*, the competing risks. One further consequence of the use of these two different approaches is that the effect of a covariate on either function can be different, as illustrated by the example given in Gray (1988, p 1142).

In practice, when reporting analysis of competing risks failure time data, a graphical display of the probabilities of failure causes against time is useful. These probabilities can be defined in two ways: the "crude" probability, which corresponds to the CIF and the "net" probability, which corresponds to the probability of the failure cause of interest in the hypothetical situation where it is the only cause of failure acting on the population (Tsiatis, 1975; Tsiatis, 1998). If "crude" probabilities can be estimated using observed data, "net" probabilities are not identifiable from observations, unless additional assumptions on the mechanism underlying the competing risks situation, such as the so-called independent competing risks framework. In this case, the failure time T is assumed to be the minimum of K mutually independent latent failure times, each corresponding to a failure of a particular cause (Sampford, 1952). This assumption is however unverifiable (Tsiatis, 1975), and strong evidence of complete biologic independence among the physiological mechanisms giving rise to the various failure causes would be required to justify such an assumption (Prentice et al., 1978). Cause-specific hazards could also be used for graphical representation, but they are less interpretable, particulary in terms of magnitude of the proportion of patients failing from the different failure causes (Pepe and Mori, 1993). Therefore, CIFs are well suited to summarize competing risks failure time data and a direct modelling of such quantities seems relevant to avoid presenting conflicting or contradictory results.

However, when assessing the prognostic value of an exposure on a specific failure

cause, a multiplicative effect of the exposure on the cause-specific hazard appears more clinically understandable. First, as exposed above, the cause-specific hazard can be expressed using the Cox model, which is widely used in the medical literature. Secondly, it seems natural that the physiological effect of a treatment or any prognostic exposure would be to reduce or increase the probability of the failure cause of interest at any time, conditionally on being still alive at that time. On the contrary, a decrease (resp. increase) in the cumulative incidence function could be due either to a physiological effect of the exposure or to an increase (resp. decrease) in the probability of the competing failure causes.

This is first exemplified on a real data example. Then, we place ourselves in a the setting where the true model for the covariate effect is a Cox proportional (cause-specific) hazards model, and derive asymptotic properties of regression parameter estimates reached by fitting a regression model for the subdistribution hazard.

2. A real example

In this section, we reanalyzed the `mgus` data set presented in Therneau (2000, pp 175–177), using the two regression approaches described above. This data set represents observations from 241 patients with monoclonal gammopathy of undetermined significance (MGUS) identified at the Mayo Clinic before Jan. 1, 1971, with a 20 to 35 years follow-up for each patient (Kyle, 1993). MGUS is considered as a potential precursor to several plasma cell malignancies. From a competing risks point of view, several competing causes of failure were reported: Failure from multiple myeloma ($n = 39$), amyloidosis ($n = 8$), macroglobulineamia ($n = 7$), other lymphoproliferative diseases ($n = 5$) or first death ($n = 130$). Due to the rather small number of failures, we considered two main failure causes, namely "death", and "plasma cell malignancy". We

focused on the estimation of the effect on failure of age, distinguishing two categories according to the sample's median, that is 64 years. The Figure 1 displays the estimated CIF of the two failure causes in each age category. Clearly, patients aged 64 ys or more were less likely to develop plasma cell malignancy than those aged 64 years or less. This is in accordance with the estimated subdistribution hazard ratio of 0.43 (95% confidence interval: 0.25-0.74) in this age category (Table 1). By contrast, the cause-specific hazard ratio of age above 64 years category was estimated at 0.80 (95% confidence interval: 0.45-1.40). Actually, as depicted in Figure 1, those patients aged 64 years or more had a much higher cumulative incidence of first death. At least, as exposed above, this highlights the necessity to display the probability of the competing failure causes across the covariate groups before any interpretation of the effect of that covariate on the CIF of the failure cause of interest. Of note, whatever the model used, assumption of proportional hazards was roughly checked through the display of cumulative hazards in each age category (Figure 2).

[Figure 1 about here.]

[Table 1 about here.]

[Figure 2 about here.]

3. Misspecified model for the cumulative incidence function

For simplicity, we considered two failure causes ($\epsilon = 1, 2$), where $\epsilon = 1$ denoted the failure cause of interest, and assumed that the true underlying model generating the data is a proportional hazards model for the cause-specific hazard function, *i.e.* :

$$\lambda_k(t; Z) = \lambda_{k0}(t) \exp(\beta_k Z) \quad k = 1, 2 \tag{4}$$

where $\lambda_{k0}(t)$ are unspecified continuous positive functions and Z is a binary covariate representing the exposure. We wished to fit a proportional subdistribution hazards model for failure of cause 1:

$$\alpha_1(t;Z) = \alpha_{10}(t)\exp(\gamma Z).$$

We examined the properties of the estimator of γ under the condition given above. Because of the nice structure of the partial likelihood for the subdistribution, results in Solomon (1984) are somewhat straightforward. Let us recall the particularity, in terms of counting process, of the Fine and Gray model, in the presence of administrative censoring, *i.e.*, when the potential censoring time is observed on all individuals (referred as "censoring complete data" in Fine and Gray, 1999).

Suppose a sample of size n and define the process $N_i(t) = \mathbf{1}_{(T_i \leq t, \varepsilon_i = 1)}$, which takes value zero until individual i fails from cause 1. Let $Y_i(t) = 1 - N_i(t-)$, be the indicator of individual i being at risk of failure before time t. We suppose that C_i is independent of (N_i, Y_i, Z_i), where the subscript i corresponds to the individual, and that (N_i, Y_i, Z_i, C_i) are independent and identically distributed replicates of (N, Y, Z, C). In case of censoring, $Y_i^*(t) = \{1 - N_i(t-)\}\mathbf{1}_{(C_i \geq t)} = Y_i(t-)\mathbf{1}_{(C_i \geq t)}$.

Let $\hat{\gamma}$ be the MPLE of γ. Using the theorem of Struthers and Kalbfleisch (1986, pp 365), $\hat{\gamma}$ is a consistent estimator of γ^*, where γ^* is the solution of

$$\int_0^\infty \frac{1}{n}\sum_{i=1}^{n} E\left\{ Y_i^*(t)\alpha_1(t;Z_i)\left[Z_i - \frac{\sum_{j=1}^n E[Z_j Y_j^*(t)\exp(\gamma Z_j)]}{\sum_{j=1}^n E[Y_j^*(t)\exp(\gamma Z_j)]} \right] \right\} dt = 0$$

or equivalently

$$I(\gamma, \beta_1) = \int_0^\infty E\left\{ f_1(t;Z)\mathbf{1}_{(C\geq t)}\left[Z - \frac{E[Z\exp(\gamma Z)(1 - F_1(t;Z))\mathbf{1}_{(C\geq t)}]}{E[\exp(\gamma Z)(1 - F_1(t;Z))\mathbf{1}_{(C\geq t)}]} \right] \right\} dt = 0,$$

where $f_1(t;Z) = dF_1(t;Z)/dt$.

No explicit solution for γ^* is available, but a Taylor expansion around $(0,0)$ provides the following approximation:

$$\gamma^* \simeq \frac{-1}{\left[\frac{\partial}{\partial\gamma}I(\gamma,\beta_1)\right]_{(0,0)}}\left\{I(0,0)+\beta_1\left[\frac{\partial}{\partial\beta_1}I(\gamma,\beta_1)\right]_{(0,0)}\right\} \tag{5}$$

with $I(0,0)$ depending on β_2, the covariate distribution, and the censoring distribution.

4. Illustration: Absolutely continuous bivariate exponential model

To illustrate the application of formula (5), we exemplified these results using a parametric model, for which explicit solutions for γ^* are obtained. Additionally, a simulation study was performed to investigate the relevance of the approximation in small samples.

4.1 *Parametric setting*

For illustration, we used a parametric latent failure times model. In such a model, each possible cause of failure is represented by a latent failure time, T_k, $k = 1, 2$, while the observable variables introduced above are defined as $T = \min(T_1, T_2)$ and $\epsilon = 2 - \mathbf{1}_{(T_1 \leq T_2)}$.

We considered the case where (T_1, T_2) has an absolutely continuous bivariate exponential distribution as introduced by Block and Basu (1974), denoted $(T_1, T_2) \sim ACBVE(a_1, a_2, a_{12})$, where a_1, a_2, and a_{12} are the distribution parameters. Accordingly, the joint survival function of (T_1,T_2) is given by:

$$\begin{aligned} S(t_1,t_2) = \Pr(T_1 > t_1, T_2 > t_2) = \; & a/(a_1+a_2)\exp[-a_1t_1 - a_2t_2 - a_{12}\max(t_1,t_2)] \\ & -a_{12}/(a_1+a_2)\exp[-a\max(t_1,t_2)], \end{aligned}$$

where $a = a_1 + a_2 + a_{12}$, with the cause-specific and subdistribution hazards for failure of cause 1:

$$\lambda_{10}(t) = \lambda_{10} = \frac{aa_1}{a_1+a_2}$$

and

$$\alpha_{10}(t) = \frac{aa_1}{a_1 + a_2 \exp(at)},$$

respectively.

Such a choice allows to consider both independent ($a_{12} = 0$) and dependent ($a_{12} \neq 0$) latent failure times. Moreover, when $a_{12} = 0$, both T_1 and T_2 have a marginal exponential distribution of parameter a_1 and a_2, respectively. Thus, this distribution can also accomodate with the classical independent exponential latent failure times approach.

Let Z be a binary exposure of interest with $p = \Pr(Z = 1)$. According to the regression model (4), the cause-specific hazard, the subdistribution function and the subdistribution hazard function of failure of cause 1 express as:

$$\lambda_1(t; Z) = \frac{aa_1}{a_1 + a_2} \exp(\beta_1 Z),$$

$$F_1(t; Z) = \frac{a_1 \exp(\beta_1 Z)}{a_1 \exp(\beta_1 Z) + a_2 \exp(\beta_2 Z)}[1 - \exp[-\Psi(Z)t]],$$

and

$$\alpha_1(t; Z) = \frac{a_1 \exp(\beta_1 Z)\Psi(Z) \exp\{-\Psi(Z)t\}}{a_2 \exp(\beta_2 Z) + a_1 \exp(\beta_1 Z) \exp\{-\Psi(Z)t\}},$$

respectively, where $\Psi(Z) = \lambda_{10} \exp(\beta_1 Z) + \lambda_{20} \exp(\beta_2 Z)$.

In case of no censoring (referred as "complete data"), equation (5) reduces to:

$$\gamma* \simeq c_0 + c_1\beta_1 \tag{6}$$

where $c_0 = c_0(\beta_2, Var(Z))$ and $c_1 = c_1(\beta_2, p)$. Calculation details are given in the Appendix.

4.2 *Numerical Studies*

Validity of formula (6) was investigated by performing a series of simulation studies. Each simulated data set of size $n = 400$ consisted of two balanced groups (exposed or

unexposed) of observations, each corresponding to a different ACBVE distribution. For the unexposed group ($Z = 0$), the (a_1, a_2) parameters were set to $\{(0.5, 1), (1, 0.5)\}$, while a_{12} ranged in $\{0, 0.5, 1, 1.5, 2, 2.5\}$. For the exposed group ($Z = 1$), an ACBVE distribution was used with parameter (a'_1, a'_2, a'_{12}). To reach proportional cause-specific hazards (4), the following ACBVE parameters a'_1, a'_2 were derived:

$$\begin{cases} a'_2 = a_{12} \left(-A + \frac{a a_2 \exp(\beta_1)(A+1)}{(a_1+a_2)\exp(\beta_1-\beta_2)} - 1 \right)^{-1} \\ a'_1 = a'_2 A \end{cases} \tag{7}$$

where $A = \frac{a_1}{a_2} \exp(\beta_1 - \beta_2) + 1$, while $a'_{12} = a_{12}$ for simplicity.

ACBVE failure times were generated using the method proposed in Friday and Patil (1977). For any data set, $\widehat{\gamma}$ was computed by using the `cmprsk` package of R (R Development Core Team, 2004), while γ^* was computed from equation (5), with numerical integrations based on the `integrate` function of R. For each configuration of the simulation parameters, $N = 5000$ independent data sets were generated to estimate the mean (and standard error) of the N values of $\widehat{\gamma}$. Table (2) reports the results of the simulations.

[Table 2 about here.]

Expectedly, formula (6) provides a good approximation of $\hat{\gamma}$ for small values of β_1 in reasonable sample sizes. However, departures of β_1 from zero led to a biased value of approximation of γ^*. By contrast, dependence between latent failure times had no effect on both γ^* and $\hat{\gamma}$. This is likely due to the $\exp(-at)$ term, which is negligible in the evaluation of the integral (see Appendix).

Additionally, we studied the influence of β_2 and p on γ^*. Results are displayed in Figure (3).

[Figure 3 about here.]

It seems that c_1 has a little influence on the value of γ^*, whereas the shape of the calculated regression coefficient is rather driven by c_0. This comment applied in both settings, *i.e.*, when β_2, the regression coefficient for the competing failure cause, varied from -1 to 1, and for value of p varying from 10 % to 90%.

5. Concluding Remarks

In a competing risks setting, when estimating the effect of an exposure on a specific failure time, there is still an open choice (dilemma) between the cause-specific and cumulative incidence approaches. The two models share the same proportional hazards assumptions but for two quantities that differ (greatly). Therefore, this can lead to model misspecification.

Actually, we showed that, when an exposure acts on a failure cause of interest through a multiplicative effect on the cause-specific hazard, analysis based on a proportional hazards model for the subdistribution hazard achieved a different effect, depending on the cause-specific effect on the failure cause of interest, but also on the cause-specific effect on the competing failure causes. This was exemplified on a real data set in MGUS, where a protective effect of advanced age on the subdistribution hazard of plasma cell malignancy was found, without any cause-specific effect. Of note, fitting a parametric exponential model to the MGUS data (as a_{12} is of little influence on γ^*), we found, under the assumption of $\beta_1 = 0$ and $\beta_2 = -1.4$ that γ^* was -1.10, which is not too far from the estimated value -0.84 (se 0.278).

An explicit relationship between both effects was derived assuming a parametric ACBVE model of the latent failure times. This relates closely to papers from Solomon (1984) and Struthers and Kalbfleisch (1986). They considered misspecification with

regards to missing covariates, functional form misspecification for the covariates and accelerated failure times model. By contrast, we considered misspecification with regards to the the hazard function for the failure cause of interest. Through simulation studies, we found that this approximate works well for small values of β_1, whatever the dependence between the latent failure times. The converse analysis was not conducted, because it appeared less plausible.

References

Andersen, P. K., Abildstrom, S. Z. and Rosthoj, S. (2002). Competing risks as a multi-state model. *Statistical Methods in Medical Research* **11**, 203–215.

Block, H. W. and Basu, A. P. (1974). A continuous bivariate exponential extension. *Journal of the American Statistical Association* **69**, 1031–1037.

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

Fine, J. P. (2001). Regression modelling of competing crude failure probabilities. *Biostatistics* **2**, 85–97.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for subdistribution of a competing risk. *Journal of the American Statistical Association* **94**, 496–509.

Friday, D. and Patil, G. (1977). A bivariate exponential model with applications to reliability and computer generation of random variables. In Tsokos, C. and Shimi, I., editors, *The Theory and Applications of Reliability*, pages 527–549. Academic Press, New York.

Gray, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of

a competing risk. *The Annals of Statistics* **116**, 1141–1154.

Kyle, R. (1993). "benign" monoclonal gammopathy - after 20 to 35 years of follow-up. In *Mayo Clinic Proceedings*, volume 68, pages 26–36.

Pepe, M. S. (1991). Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association* **86**, 770–778.

Pepe, M. S. and Mori, M. (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine* **12**, 737–751.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541–554.

R Development Core Team (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.

Sampford, M. R. (1952). The estimation of response time distributions. II. Multistimulus distributions. *Biometrics* **8**, 307–353.

Solomon, P. J. (1984). Effect of misspecification of regression models in the analysis of survival data. *Biometrika* **71**, 291–298.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73**, 363–369.

Therneau, T.M. Grambsch, P. (2000). *Modeling Survival Data. Extending the Cox Model*. Springer, New York.

Tsiatis, A. A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceeding of the National Academy of Sciences* **72**, 20–22.

Tsiatis, A. A. (1998). Competing risks. In Armitage, P. and Colton, T., editors,

Appendix A

Expections are taken relative to the covariate Z. The following development heavily relies on Solomon (1984) and Struthers and Kalbfleisch (1986) apart from the counting process involved in the partial likelihood.

A Taylor expension around the neigbourhood of $(\gamma, \beta_1) = (0, 0)$ leads to:

$$\begin{aligned}
\gamma* &\simeq \frac{\Gamma^{-1}}{Var(Z)} \int_0^\infty E\left\{(a_1 + a_2)\exp(-\Psi_0(Z)t)\left[Z - \frac{p[l_2(1) + a_1\exp(-\Psi_0(1)t)]}{(a_1 + l_2(1))E[(1 - F_1(t;Z))_{|(0,0)}]}\right]\right\} dt \\
&+ \beta_1\Gamma^{-1}\int_0^\infty \frac{(1-\lambda_1 t)[a_2 + a_1\exp(-at)]\exp(-\Psi_0(1)t)}{E[(1 - F_1(t;Z))_{|(0,0)}]} \\
&+ \frac{E[\exp(-\Psi_0(Z)t)]a_1[a_2 + a_1\exp(-at)]}{(a_1 + l_2(1))E^2[(1 - F_1(t;Z))_{|(0,0)}]}\{l_2(1) + [a_1\lambda_1 t + (\lambda_1 t - 1)l_2(1)]\exp(-\Psi_0(1)t)\}dt,
\end{aligned}$$

where $\Psi_0(Z)) = \Psi(Z)_{|(\beta_1=0)}$ and $l_i(Z) = a_i\exp(\beta_i Z)$, $i \in (1, 2)$

$$\Gamma = \int_0^\infty \frac{E[\exp(-\Psi_0(Z)t)]}{E^2[(1 - F_1(t;Z))_{|(0,0)}]} \times \frac{[a_2\exp(\beta_2) + a_1\exp(-\Psi_0(1)t)][a_2 + a_1\exp(-at)]}{a_1 + a_2\exp(\beta_2)}dt$$

$$E[\exp(-\Psi_0(Z)t)] = p\exp[-(\lambda_{10} + \lambda_{20}\exp(\beta_2))t] + (1-p)\exp(-at)$$

$$\begin{aligned}
E[(1 - F_1(t;Z))_{|(0,0)}] &= [(pa_1 + a_2)l_2(1) + (1-p)a_1(l_2(1) + a_1)\exp(-at) \\
&+ pa_1(a_1 + a_2)\exp(-\Psi_0(1)t) + (1-p)a_1a_2] \\
&/ (a_1 + a_2)(a_1 + a_2\exp(\beta_2))
\end{aligned}$$

and $p = \Pr(Z = 1)$.

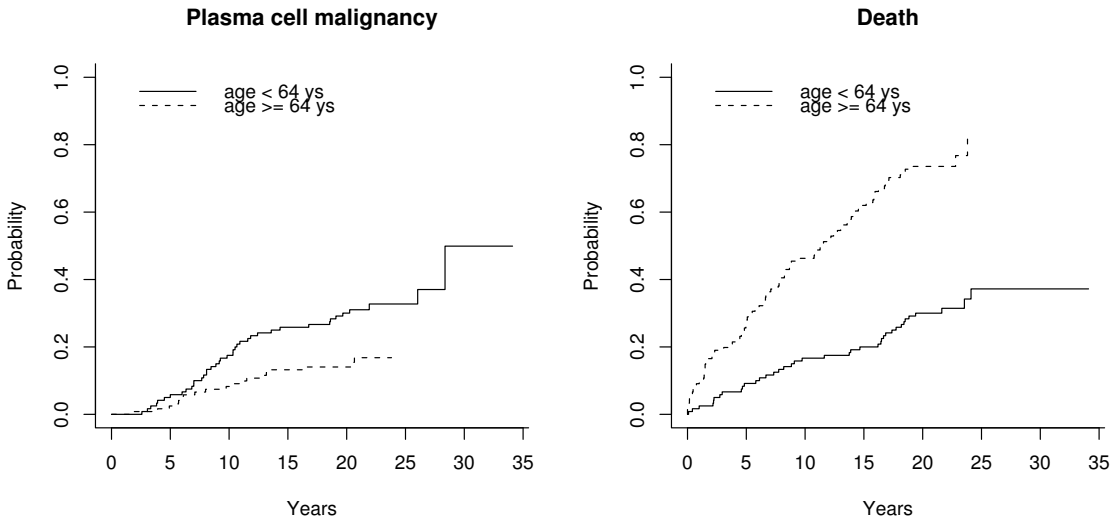


Figure 1. Cumulative incidence of first death or plasma cells malignancy according to age category for the MGUS data.

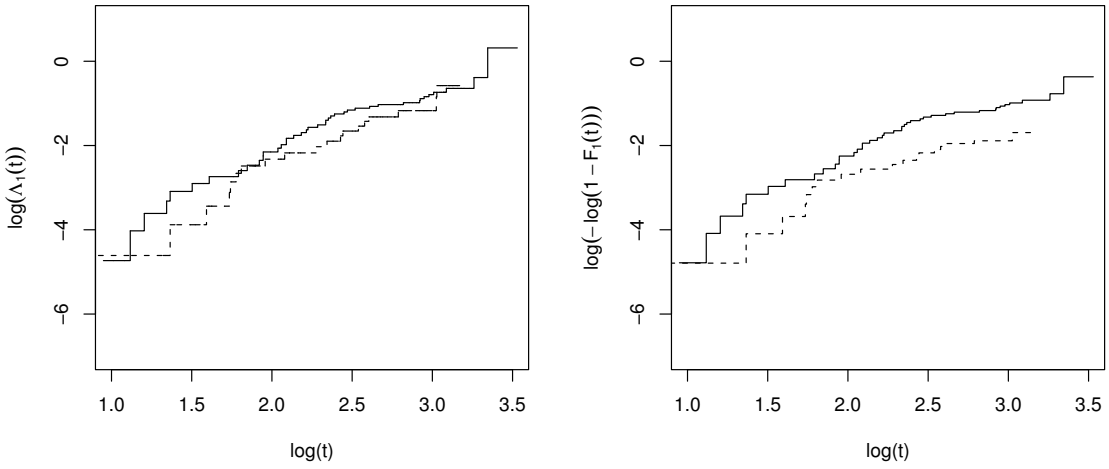


Figure 2. Cumulative hazard of occurrence of a plasma cell malignancy in MGUS data, expressed either as the cumulative cause-specific hazard, $\log \Lambda_1(t)$, or the cumulative subdistribution hazard, $\log[-\log(1 - F_1(t))]$.

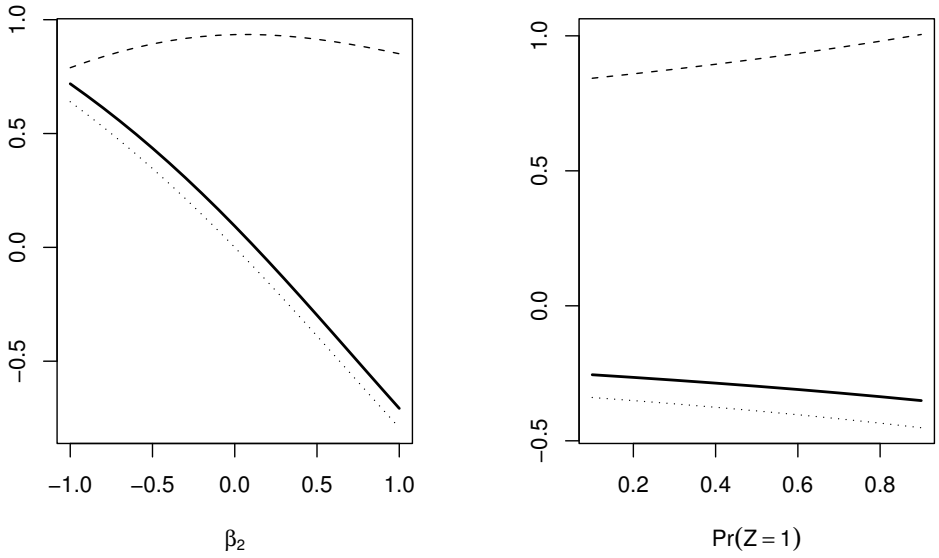


Figure 3. c_0 (dotted line), c_1 (dashed line), γ^* (solid line), while β_2 varying with $(\beta_1, p, a_1, a_2, a_{12}) = (0.1, 0.5, 0.5, 1, 0)$(left plot), while p varying (right plot) with $(\beta_1, \beta_2, a_1, a_2, a_{12}) = (0.1, 0.5, 0.5, 1, 0)$

Table 1

Estimated effects of the covariate "age $\geq$ 64 ys" on the cause-specific and subdistribution hazards of developping a plasma cell malignancy and dying without developping a malignancy.

Effet of age	Estimated regression parameter (Standard Error)	
Failure cause	Cause-specific hazard	Subdistribution hazard
Plasma cell malignacy	-0.225 (0.285)	-0.842 (0.278)
Death as first event	1.39 (0.198)	1.31 (0.192)

Table 2

Simulation Results

a_1	a_2	a_{12}	β_1	β_2	γ^*	Mean$(\hat{\gamma})(se)$
0.5	1	0	0.1	0.5	-0.296	-0.301 (0.186)
0.5	1	0	0.2	0.5	-0.205	-0.213 (0.181)
0.5	1	0	0.3	0.5	-0.114	-0.128 (0.183)
0.5	1	0	0.4	0.5	-0.022	-0.038 (0.177)
0.5	1	0	0.5	0.5	0.068	0.049 (0.174)
0.5	1	0.5	0.1	0.5	-0.296	-0.308 (0.188)
0.5	1	1	0.1	0.5	-0.296	-0.306 (0.189)
0.5	1	1.5	0.1	0.5	-0.296	-0.307 (0.189)
0.5	1	2	0.1	0.5	-0.296	-0.300 (0.189)
0.5	1	2.5	0.1	0.5	-0.296	-0.304 (0.186)
1	0.5	0.5	0.1	0.5	-0.152	-0.174 (0.128)
1	0.5	1	0.1	0.5	-0.152	-0.173 (0.129)
1	0.5	1.5	0.1	0.5	-0.152	-0.173 (0.129)
1	0.5	2	0.1	0.5	-0.152	-0.173 (0.128)
1	0.5	2.5	0.1	0.5	-0.152	-0.179 (0.129)

Chapitre 5

Validité du modèle de Fine & Gray quand la covariable dépend du temps

5.1 Objectifs

Les développements précédents ont supposé que les variables explicatives étaient constantes au cours du temps. En termes cliniques, ces covariables ne peuvent donc modéliser que des quantités mesurées à l'inclusion dans l'étude, qu'il s'agisse de caractéristiques démographiques, liées à l'état du malade ou de son traitement. Cependant, en recherche clinique, on est souvent confronté à des marqueurs cliniques ou biologiques dont la mesure est répétée au cours du temps (par exemple, le taux de CD4 chez les patients atteints de SIDA, le taux d'immunoglobuline monoclonale chez les patients atteints de Myélome multiple, ...), dont on souhaite étudier l'influence sur la survenue d'un événement d'intérêt. On peut les modéliser par des covariables dépendant du temps, en distinguant (Kalbfleisch and Prentice, 1980) :

- les covariables dites *internes*, qui correspondent à des mesures réalisées au cours du temps chez l'individu, observables jusqu'à l'interruption de son suivi (par exemple, un taux de globules blancs, la présence de complications infectieuses, ...) ; elles permettent de modéliser la survenue d'un événement au cours du temps par un processus de saut au moment de l'événement.
- les covariables dites *externes*, qui comprennent les covariables *définies*, déterminées à l'avance pour un individu (par exemple, l'âge d'un individu dans un essai clinique de longue durée), et les covariables *ancillaires*, dont la mesure ne dépend pas intrinsèquement de l'individu (comme la pollution de l'air).

L'inclusion de covariables dépendant du temps de type *externe* ne pose pas de problème majeur d'estimation, car toutes les quantités précédemment définies (conditionnelllement aux covariables) restent identifiables. Fine et Gray ont ainsi autorisé l'inclusion de covariables dépendant du temps de la forme "temps par covariable" (*i.e.*, de covariable *définies* selon la terminologie précédente). Cette famille de covariables est cependant extrêmement réductrice car même un simple processus de saut n'y appartient pas.

Ainsi, il convient de distinguer plusieurs situations où l'inclusion de covariables dépendant du temps de type *interne* entraine des biais d'estimation. Dans ce travail, nous nous sommes donc intéressés aux problèmes d'inférence liés à l'inclusion de covariables dépendant du temps de type *interne* dans le modèle de régression de Fine & Gray.

5.2 Méthodes

Par construction, la fonction de sous répartition utilise une filtration atypique. La structure de la pseudo variable aléatoire T^* (2.10) entraîne que le processus associé à une covariable *interne* n'est pas adapté à la filtration (naturelle) engendrée par les observations. Ceci a été illustré dans le contexte de la greffe de moelle allogénique, où le décès en rémission est un événement en compétition de la rechute. La modélisation de l'effet de la survenue d'une maladie du greffon contre l'hôte (GvHD) sur l'incidence cumulée de rechute revient à introduire dans le modèle de Fine & Gray une covariable *interne*. Il s'agit d'une covariable binaire dépendant du temps qui est nulle avant la survenue de la GvHD et qui saute d'une unité après. Une étude de simulation a été conduite afin d'estimer, dans cette situation, le biais d'estimation sur le paramètre de régression. Pour ce faire, une méthode de simulation de temps d'événements, avec des covariables internes provenant du modèle de Fine et Gray (à risques proportionnels), a été proposée.

5.3 "A note on including time-dependent covariate in regression model for competing risks data"

Cet article a été soumis pour publication à *Biometrical Journal*.

A note on including time-dependent covariate in regression model for competing risks data

A. Latouche[*1], **R. Porcher**[1], and **S. Chevret**[1]

[1] Département de Biostatistique et Informatique Médicale

Hôpital Saint-Louis, AP-HP,

Université Paris 7 and INSERM ERM 0321

F-75010, Paris, France

Summary

Recently, regression analysis of the cumulative incidence function has gained interest in competing risks data analysis, through the model proposed by Fine and Gray (JASA 1999;94:496–509). In this note, we point out that inclusion of time-dependent covariates in this model can lead to serious bias. We illustrate the problems arising in such a context, using bone marrow transplant data as a working example and numerical simulations. Practical advices are given, preventing the misuse of this model.

Key words: time-dependent covariate, subdistribution, competing risk

1 Introduction

In longitudinal cohort studies, competing risks failure time data are commonly encountered. For instance, after allogeneic bone-marrow transplantation (aBMT) for leukemic patients in complete remission, deaths in remission compete with relapse. This will be our working example. To isolate the effect of covariates

* Corresponding author: e-mail: aurelien.latouche@chu-stlouis.fr, Phone: +33 142 499 742, Fax: +33 142 499 745

on these risks, several regression models can be used. Actually, regression analysis of competing risks failure time can be performed either by modelling the cause specific hazard function or the cumulative incidence function (also known as the subdistribution function). The former approach is commonly used in this setting (Rosenberg et al., 2004; Cornelissen et al., 2001). However, the instantaneous risk of specific failure cause is sometimes of less interest than the overall probability of this specific failure. In our working example, actually, the overall probability of death in remission, often referred as "treatment related mortality" appears more interesting than the instantaneous risk of dying in remission. Otherwise, the overall probability of relapse is also of interest to quantify the outcomes in the population of transplanted patients.

Such a probability of failure could be formulated as either the marginal distribution (of the specific failure cause), that is the probability of this failure cause in a population where only this failure cause acts, or the cumulative incidence function, i.e., the overall probability of the specific cause of failure in the presence of the competing failure causes. However, the marginal distribution is not identifiable from available data without additional assumptions, such as independence between competing failure causes (Tsiatis, 1998). Therefore, cumulative incidence functions may appear more relevant than marginal probabilities (Pepe and Mori, 1993; Korn and Dorey, 1992; Gaynor et al., 1993).

To assess the effect of a covariate on the cumulative incidence of a competing risk, Fine and Gray (1999) proposed a regression model. It has been recently used to model clinical data in cancer (Colleoni et al., 2000; Robson et al., 2004) or hematology (Rocha et al., 2001, 2002). It allows to estimate the effect of constant (time-fixed) covariates on the subdistribution hazard of specific failure causes. Time-by-covariate interaction is handled by this model, but most of time-dependent covariates such as "one time jumps" (taking 0 value unless the outcome of interest is observed, and 1 thereafter) are not. For instance, in the context of our working example, patients with leukemia frequently develop after aBMT acute graft versus host disease (aGvHD) wherin the transplanted immune cells attack the host tissues. Some evidence exists

to consider that occurrence of aGvHD modifies patients' outcome as it increases risk of mortality but decreases risk of relapse. One could be interested in estimating the effect of such a time-dependent covariate (taking zero values unless the aGvHD is observed and 1 thereafter) on the occurrence of failures of interest (death or relapse).

We show that inclusion of such internal time-dependent covariate is not relevant when modelling the subdistribution hazards as it implies conditioning on the future. This article should be considered as a guideline for preventing the misuse of the model in this setting. In Section 2, we present the Fine and Gray regression model. A real data example is proposed in Section 3. In Section 4, we present a Monte Carlo simulation study to assess the resulting bias in estimating the effect of a time-dependent covariate using the Fine and Gray model. Concluding remarks are presented in Section 5.

2 Models

Let T be the failure time, ϵ the cause of failure, where $\epsilon = 1$ denotes the cause of interest and $\epsilon = 2$ the competing cause (considering, without loss of generality, a single competing failure cause), and $F_i = \Pr(T \leq t, \epsilon = i)$ the cumulative incidence function of failure from the cause i $(= 1, 2)$. Gray (1988) defined the subdistribution hazard for cause i as:

$$\lambda_i(t) = \lim_{dt \to 0} \frac{1}{dt} \Pr\left\{t \leq T \leq t + dt, \epsilon = i | T \geq t \cup (T \leq t \cap \epsilon \neq i)\right\},$$

by contrast to the cause-specific hazard:

$$\alpha_i(t) = \lim_{dt \to 0} \frac{1}{dt} \Pr\left\{t \leq T \leq t + dt, \epsilon = i | T \geq t\right\}.$$

Similarly to the Cox model for the cause-specific hazard, $\alpha_i(t; X(t)) = \alpha_{i0}(t) \exp\{b_i X(t)\}$, where $\alpha_{i0}(t)$ is a non specified baseline hazard function, and b_i is the regression parameter, Fine and Gray (1999) proposed a regression model for the subdistribution hazard: $\lambda_i(t; X(t)) = \lambda_{i0}(t) \exp\{\beta_i X(t)\}$. By construction, the subdistribution hazard is explicitely related to the cumulative incidence function of failure from

cause i, by $\lambda_i(t) = -d\log\{1-F_i(t)\}/dt$, while the relation between the cause-specific hazard and the cumulative incidence function is less straightforward, and involves the cause-specific hazard of failure from other causes.

For inference in this model from $j = 1, \ldots, N$ individuals, the risk set at time t expresses as $\mathcal{R}(t) = \{j : (t \leq T_j) \cup (T_j \leq t \cap \epsilon_j \neq i)\}$. This includes individuals who have not failed from any cause by t, like in the Cox model for the cause specific hazard (with risk set at time t defined by $\{j : t \leq T_j)\}$), and, in addition, those who have previously failed from the competing cause before t.

Let T be the time of failure of the individual, and Z be the time to occurrence of any event of interest. Suppose that we wish to estimate the effect of $X(t) = 1_{\{Z \leq t\}}$ on the subdistribution hazard $\lambda_1(t)$ at a particular time τ. If $T \leq \tau$ and the cause of failure is not that of interest ($\epsilon = 2$), the risk set comprises individuals who have not experienced any failure, and those who have previously failed from the competing cause. Moreover, in the case of an absorbing competing cause of failure such as death, the covariate value of a patient who dies cannot be observed anymore while the patient is still considered to be at risk until the maximum observation time of the cohort. This is illustrated in Figure 1, in absence of censoring.

[Fig. 1 about here.]

Let "*non-identifiable path*" denote further those observations, in opposition to "*identifiable path*" where the occurrence of the competing cause of failure does not avoid the observation of $X(t)$.

3 A clinical example

We illustrated estimation of the effect of such a time-dependent covariate on a specific failure cause on real data. Data consist in a sample of 180 children with acute leukemia who underwent aBMT between 1994 and 1998 (Rocha et al., 2001). Of these 180 patients, 34 developed aGvHD followed by either relapse for 6 patients or death in remission for 22. Among the 146 patients who did not experience aGvHD, there were 60 relapses and 22 deaths in remission (Figure 2). No patient was loss to follow-up. We were concerned by estimating the effect of aGvHD on the occurrence of relapse ($\epsilon = 1$).

[Fig. 2 about here.]

Estimation of β_1 was carried out using the `survival` package of R Team (2004) with competing failure observations censored at their follow-up time (difference between the reference date and the entry date), as censoring only resulted from administrative loss to follow-up. Estimation of b_1 was performed by using a standard Cox model, where deaths in remission were censored at the time of death. The time-dependent covariate, aGvHD, was considered as a one-time jump, taking the value 0 unless aGvHD is observed. Of note, for the Fine and Gray model, the last value of the jump was carried out forward after the competing failure time of death in remission.

The estimated effect of aGvHD on the hazard of relapse, with death in remission defining the competing cause of failure, was statistically significant, with $\hat{\beta}_1 = -0.975$ ($SE = 0.429$, $p = 0.023$). By contrast, the effect of aGvHD on the cause-specific hazard of relapse was not, with $\hat{b}_1 = -0.404$ ($SE = 0.43$, $p = 0.322$).

In the next Section, a simulation study will exhibit the fact that the former estimate have no sense as we are obviously in a "*non-identifiable path*" setting.

4 Simulation

We conducted a simulation study to numerically illustrate probles arising when using the Fine and Gray model to estimate the effect of a time-dependent covariate on the subdistribution hazard of failure. Specifically, we were interested in examining the bias in estimating β_1 when the competing cause of failure is either non absorbing or absorbing for the covariate process. For the time dependent covariate, we considered a one jump process as defined by $X(t) = 1_{\{Z \leq t\}}$, where Z is the time to occurrence of some event that could be related to the outcome. We attempted to mimic the data example exposed above.

All simulations were based on 1,000 realizations with sample sizes of 250. For simplicity, we supposed the absence of right censoring. The occurrence of jump in the covariate process was generated from a Bernoulli distribution with parameter $q = 0.6$.

Then, the time to occurrence of aGvHD, Z, was chosen to reach a probability near 1 of aGvHD at time 100 (days), as aGvHD is defined only within the first 100 days post-transplant, with a shape similar to that observed on real data sample. Thus, the individual times Z were generated from a random variable $40 \times W$, where W has Weibull distribution with shape parameter of 2 and scale parameter of 1.

Generating failure times was complicated by the presence of the time-dependent covariate. It was based on inversion of the cumulative subdistribution hazard functions, adapting the method proposed by Leemis et al. (1990) in the case of survival data. Lifetime data from the cause of interest were generated as described by Fine and Gray (1999). Details of the failure times generation is presented in the Appendix. Simulation codes are available upon request to the corresponding author.

We simulated two types of covariate paths: (i) identifiable paths, when the covariate process of patients who experienced the competing failure cause can still be observed, and (ii) non identifiable paths, when the time-dependent covariate $X(t)$ cannot be observed after occurrence of the competing failure cause. In this latter case we used the value of $X(t)$ at the time of failure throughout the risk set. Parameters (β_1, β_2) were set at $(0.5, 0.5)$ in (i), and at $(-0.5, 0.5)$ in (ii).

From the 1,000 simulations, we computed the mean estimate of β_1 ($E(\hat{\beta}_1)$) and of the proportion γ of patients who experienced the competing cause of failure before any jump of $X(t)$, for values of p ranging from 0.1 to 1 and values of $K = r_2/r_1$ ranging from 0.1 to 2.

We begin by presenting simulation results from model with identifiable paths. Figure 3 displays the mean estimate of $\hat{\beta}_1$ against K (Figure 3a) and p (Figure 3c). Whatever the value of K and of p, $E(\hat{\beta}_1)$ was close to its nominal value. This exemplifies the ability of the model to estimate the regression coefficient when the entire covariate path is known.

[Fig. 3 about here.]

Figure 4 displays simulations results when the occurrence of the competing cause of failure avoids the observation of the jump process (non identifiable paths). Contrarily to the previous observable case, $\hat{\beta}_1$ was systematically biased, with bias increasing with K (Figure 4a). Interestingly, the shape of the estimated

β_1 against K was very similar to that of γ (Figure 4b). Next, for $K = 1$, we computed $E(\hat{\beta}_1)$ for values of parameter p from 0.1 to 1 (Figure 4c). It appears that the estimates $\hat{\beta}_1$ are biased, except in the case of $p = 1$, *i.e.*, when all individuals fail from the cause of interest. In this case, γ is obviously null, as shown on Figure 4d. When p is close to zero, $F_1(t) \approx 0$, and the model is "ill-posed", so that computing $\hat{\beta}_1$ does not make any sense. Similar shapes were observed for values of $r_1 = 0.005, 0.01, 0.02$, with an increase in the bias of $E(\hat{\beta}_1)$ as r_1 increases (or equivalently an increase in γ as shown on Figure 4b). Of note, a linear decrease of γ with p was observed (Figure 4d), whereas such pattern was not found between $E(\hat{\beta}_1)$ and p.

[Fig. 4 about here.]

Moreover in our simulation setting, one can show that: $\gamma = (1-p)\{q + (1-q) \times \mathbf{C}\}$, where $\mathbf{C}$ is the probability of jump after failure, conditional on failure from competing cause, and is therefore independent of p and q. As a result, γ is indeed a decreasing linear function of p as shown in Figures 3d and 4d.

5 Discussion

In this paper, we showed, on the basis of a working example and a simulation study, that the Fine and Gray model is not appropriate for estimating the effect of any time-dependent covariate unless the entire covariate path is observable. Otherwise, *i.e.*, in the case of so-called "internal" time-dependent covariate using the terminology of Kalbfleisch and Prentice (1980), the use of the Fine and Gray model can lead to a serious bias in estimate, even in the simple studied case of a one time jump process, which is actually often observed in clinical epidemiology data.

Since the Fine and Gray model can only be used if the entire path of the time-dependent covariate is known, obviously, this prohibits the introduction of any time-dependent covariate in the model when death is a competing cause of failure. For instance, in our working example, no valid estimation of the effect of aGvHD on the subdistribution hazard of relapse could be obtained, due to deaths in remission.

Nevertheless, besides death, non fatal competing events could also be considered similarly, unless checking carefully that the observation period does not end with the occurrence of the competing event.

Our main concern was to prevent the misuse of the Fine and Gray model with time-dependent explanatory variables. Our simulation studies also provide a better understanding of the structure of the "unnatural" risk set of the Fine and Gray model, pointing out that competing failures stay in the risk set until censoring time.

To cope with estimation of the effect of time-dependent covariates, other statistical models should be proposed. Multistate models with cause-specific transition rate have already been used (Andersen et al., 2002; Hougaard, 1999). Further work is needeed to estimate time-dependent transition (non-homogeneous markov process) in this setting.

Appendix

Briefly, the subdistribution of failure times from the cause of interest ($\epsilon = 1$) is given by $F_1(t, X(t)) = 1 - [1 - p\{1 - \exp(-r_1 t)\}]^{\exp(\beta_1 X(t))}$, which is a unit exponential mixture with mass $1 - p$ at $+\infty$, where p is the proportion of failures from the cause of interest, and uses the proportional subdistribution hazards model to obtain the subdistribution for nonzero covariate values. Let $\psi(X(t))$ be the link function relating the covariate process to the subdistribution hazard function, and $\Psi(.)$ the cumulative link function *i.e.* $\Psi(t) = \int_0^t \psi(X(u))du$. Let $\Lambda_1(.)$ be the cumulative subdistribution hazard function, $\Lambda_1(t) = \int_0^t \lambda_1(u)du$. As a result, $\Lambda_1(t) = -\log S_{10}(t)$ for $t \leq \tau$ and $\Lambda_1(t) = -\log S_{10}(\tau) + \exp(\beta_1) \times \{\log S_{10}(\tau) - \log S_{10}(t)\}$ otherwise, where $S_{10}(t) = 1 - F_1(t, X(t) = 0)$. Failure times from the cause of interest were thus generated through, $t \leftarrow \Psi^{-1}[\Lambda_1^{-1}\{-\log(1-u)\}]$, where u is taken from a uniform distribution on $[0, 1]$ and $\psi(X(t)) = \exp\{\beta_1 X(t)\}$.

Since the subdistribution for the competing failure cause was considered exponentially distributed with rate r_2, we directly used the non-modified algorithm of Leemis et al. (1990) to generate corresponding competing failure times, with the link function $\psi(X(t)) = \exp\{\beta_2 X(t)\}$.

References

P. K. Andersen, S.Z. Abildstrom, and S. Rosthoj. Competing risks as a multi-state model. *Statistical Methods in Medical Research*, 11(2):203–215, 2002.

Marco Colleoni, Anne O'Neill, Aron Goldhirsch, Richard D. Gelber, Marco Bonetti, Beat Thrlimann, Karen N. Price, Monica Castiglione-Gertsch, Alan S. Coates, Jurij Lindtner, John Collins, Hans-Jrg Senn, Franco Cavalli, John Forbes, Anne Gudgeon, Edda Simoncini, Hernan Cortes-Funes, Andrea Veronesi, Martin Fey, and Carl-Magnus Rudenstam. Identifying breast cancer patients at high risk for bone metastases. *Journal of Clinical Oncology*, 18(23):3925–3935, 2000.

Jan J. Cornelissen, Michael Carston, Craig Kollman, Roberta King, Adriaan W. Dekker, Bob Lwenberg, and Claudio Anasetti. Unrelated marrow transplantation for adult patients with poor-risk acute lymphoblastic leukemia: strong graft-versus-leukemia effect and risk factors determining outcome. *Blood*, 97(6):1572–1577, March 2001.

J. P. Fine and R. J. Gray. A proportional hazards model for subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.

J. J. Gaynor, E. J. Feuer, C. C. Tan, D. H. Wu, D. J. Straus, B.D. Clarkson, MF Brennan, and CR Little. On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. *Journal of the American Statistical Association*, 88:602–607, 1993.

R. J. Gray. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, 116:1141–1154, 1988.

P. Hougaard. Multi-state models: A review. *Lifetime Data Analysis*, 5:239–264, 1999.

J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York, 1980.

E.L. Korn and F.J. Dorey. Applications of crude incidence curves. *Statistics in Medicine*, 11(6):813–829, 1992.

L. Leemis, L. Shih, and K. Reynertson. Variate generation for the accelerated life and proportional hazards models with time dependent covariates. *Statistics and Probability Letters*, 10(1):335–339, 1990.

M.S. Pepe and M. Mori. Kaplan-meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*, 12(8):737–751, 1993.

ME Robson, PO Chappuis, J Satagopan, N Wong, J Boyd, JR Goffin, C Hudis, D Roberge, L Norton, LR Begin, K Offit, and WD Foulkes. A combined analysis of outcome following breast cancer: differences in survival based on BRCA1/BRCA2 mutation status and administration of adjuvant treatment. *Breast Cancer Research*, 6(1), 2004.

V. Rocha, R. F. Franco, R. Porcher, H. Bittencourt, W. A. Jr Silva, A. Latouche, A. Devergie, H. Esperou, P. Ribaud, G. Socie, M.A. Zago, and E. Gluckman. Host defense and inflammatory gene polymorphisms are associated with outcomes after HLA-identical sibling bone marrow transplantation. *Blood*, 100(12):3908–3918, Dec 2002.

Vanderson Rocha, Jacqueline Cornish, Eric L. Sievers, Alexandra Filipovich, Franco Locatelli, Cristina Peters, Mats Remberger, G. Michel, William Arcese, Sandro Dallorso, Karin Tiedemann, Alessandro Busca, Ka-Wah Chan, Shunichi Kato, Juan Ortega, Marcus Vowels, Axel Zander, G. Souillet, Anthony Oakill, Ann Woolfrey, Andrea L. Pay, Ann Green, Federico Garnier, Irina Ionescu, Peter Wernet, Girolamo Sirchia, Pablo Rubinstein, Sylvie Chevret, and Eliane Gluckman. Comparison of outcomes of unrelated bone marrow and umbilical cord blood transplants in children with acute leukemia. *Blood*, 97:2962 – 2971, May 2001.

Philip S Rosenberg, Yi Huang, and Blanche P Alter. Individualized risks of first adverse events in patients with fanconi anemia. *Blood*, April 2004.

R Development Core Team. *R: A language and environment for statistical computing statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2004. ISBN 3-900051-00-3.

A. Tsiatis. *Encyclopedia of Biostatistics*, chapter Competing risks, pages 824–834. John Wiley & Sons, New York, 1998.

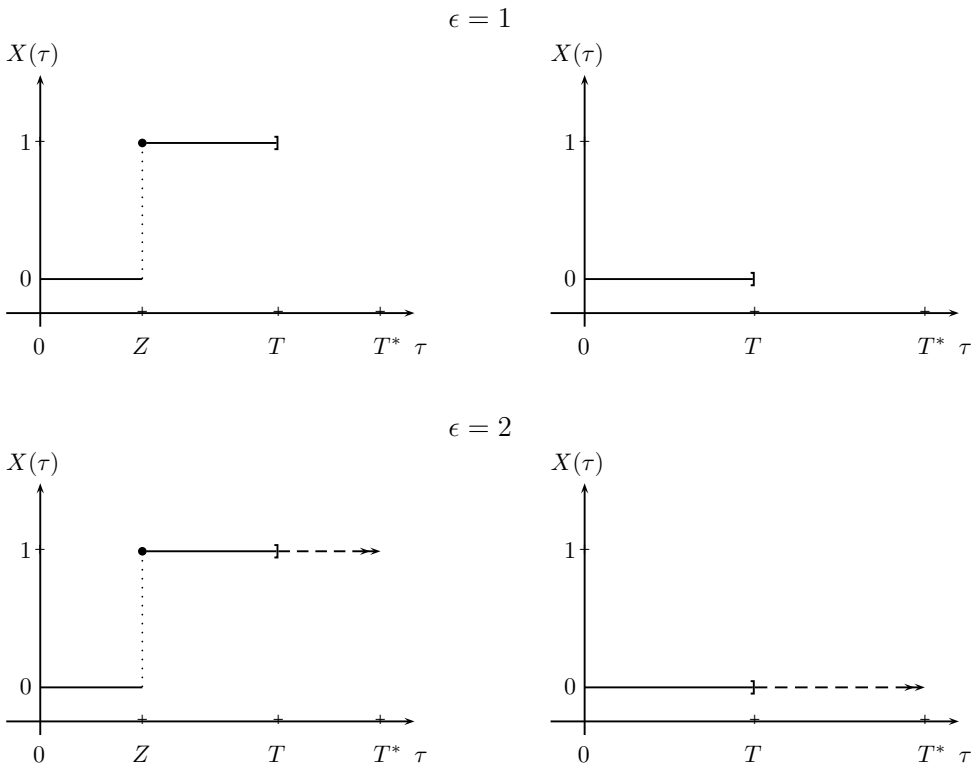


Fig. 1 Illustration of the covariate path, $X(\tau)$ overtime τ according to the experienced events. T denotes the failure time and Z denotes the time to occurrence of the event of interest. Upper plots concern patients who failed from the cause of interest ($\epsilon = 1$) while lower plots concern patients who failed from the competing failure cause ($\epsilon = 2$). Left plots concerns patients who experienced the event of interest, and right plots concern patients who did not. T^* denotes the maximal failure time among the sample.

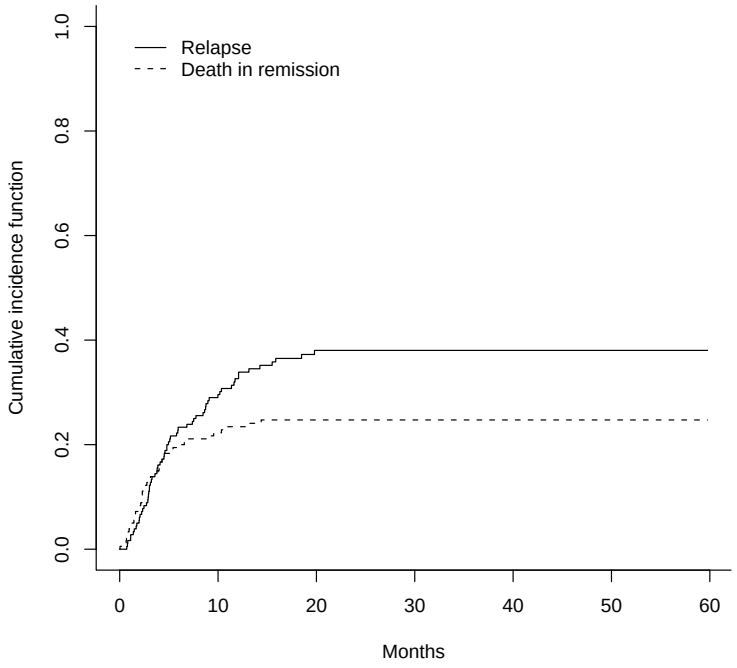


Fig. 2 Estimated cumulative incidences of relapse and death in remission

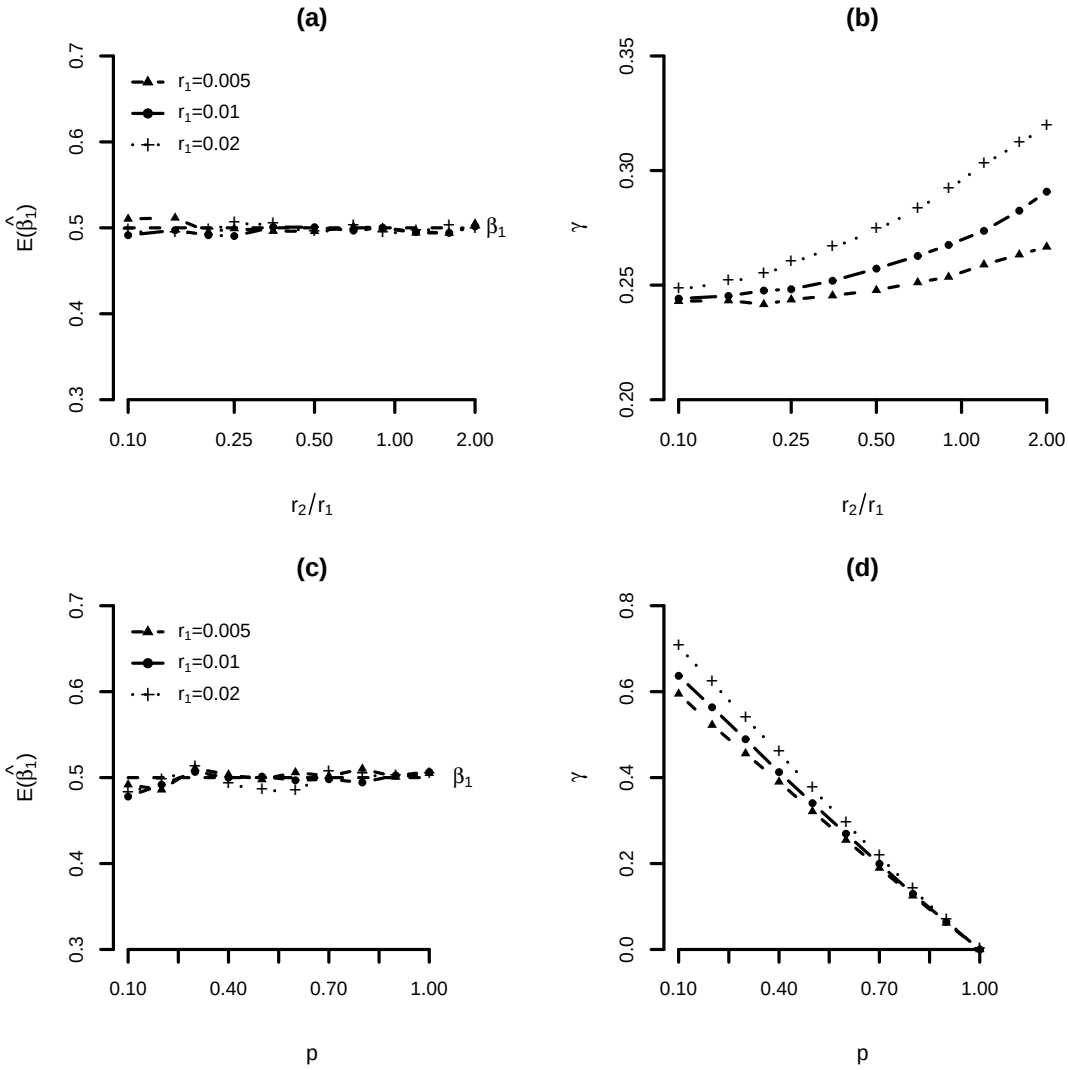


Fig. 3 Simulation results in the case of identifiable path: Mean value of $\hat{\beta}_1$ (a) and the proportion γ of patients who experience the competing failure cause before any jump (b) against the ratio r_2/r_1 of the rates of failures from cause 2 and 1. Mean value of $\hat{\beta}_1$ (c) and γ (d) against the proportion p of failure from cause 1.

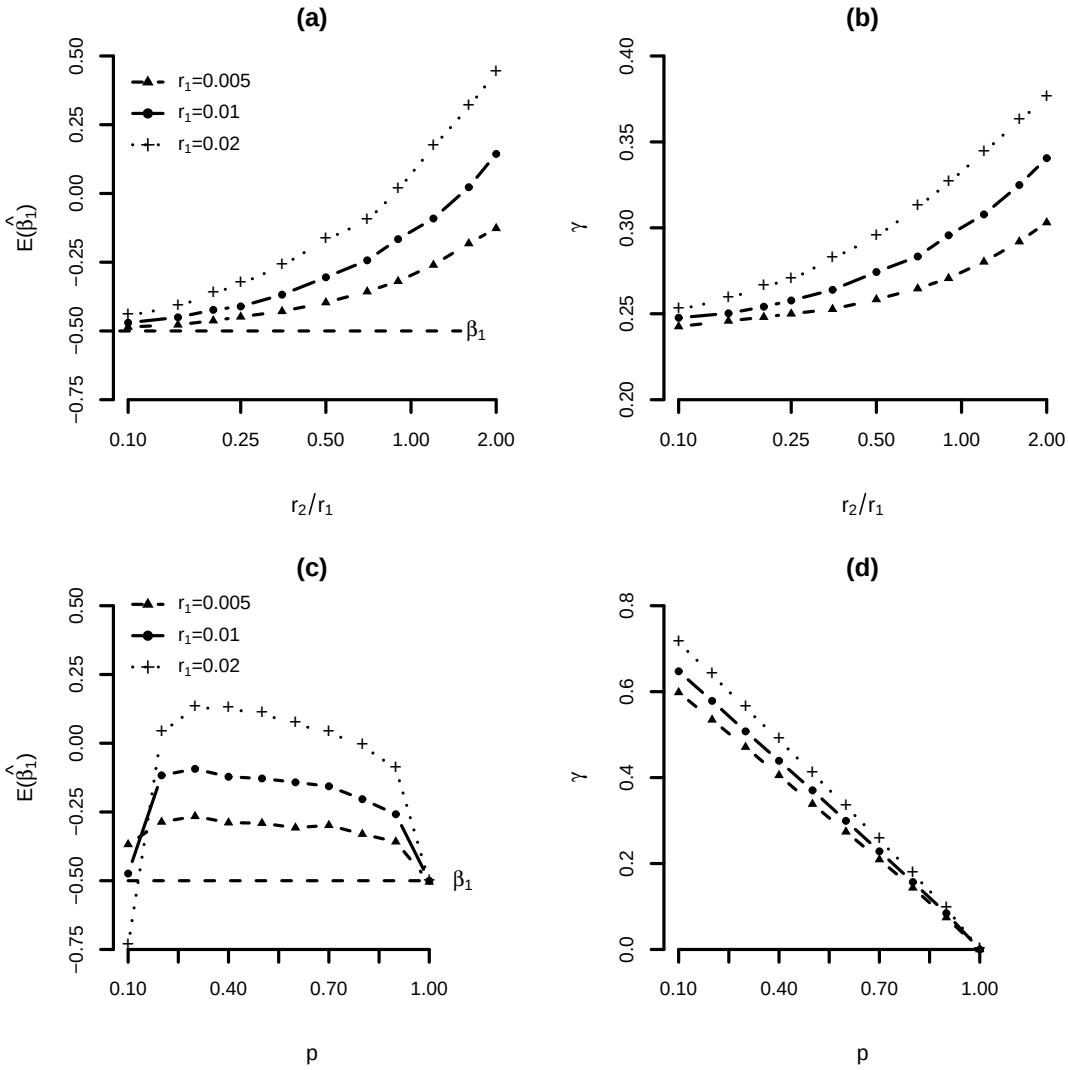


Fig. 4 Simulation results in the case of non-identifiable path: Mean value of $\hat{\beta}_1$ (a) and the proportion γ of patients who experience the competing failure cause before any jump (b) against the ratio r_2/r_1 of the rates of failures from cause 2 and 1. Mean value of $\hat{\beta}_1$ (c) and γ (d) against the proportion p of failure from cause 1.

Chapitre 6

Stratégie de modélisation en présence de compétition

6.1 Objectifs

Nous avons vu précédemment que le choix en situation de compétition entre une modélisation de la fonction de risque cause-spécifique et une modélisation de la fonction de risque de sous-répartition dépendait d'abord de l'interprétation que l'on souhaite donner à la notion de facteur de risque, et notamment de l'hypothèse sous jacente sur le mécanisme d'action de ce facteur. En pratique, lors de l'analyse de données de recherche clinique, il n'est pas toujours aisé de déterminer quelle définition du risque d'intérêt (risque instantané cause-spécifique ou probabilité d'événement) est la mieux adaptée à la question posée. Certains choix semblent assez simples au premier abord. Par exemple, on s'intéressera à la probabilité de maladie aiguë du greffon contre l'hôte (survenant dans les 100 jours post-greffe) dans un modèle où, après greffe de moelle osseuse, les patients peuvent expérimenter cette maladie, rechuter ou mourir, dans la mesure où l'événement d'intérêt est limité dans le temps. Au contraire, la fonction de risque cause-spécifique de décès cardiovasculaire semble bien adaptée à l'évaluation d'un agent anti-hypertenseur, dans la mesure où c'est le risque instantané de cet événement qui est l'objet d'intérêt. Le plus souvent, néanmoins, une compréhension approfondie des implications du choix de l'un ou l'autre des deux modèles est nécessaire, d'autant plus que les cliniciens sont de plus en plus confrontés à ces modèles dans la littérature médicale. Dans ce chapitre, nous présentons donc un travail à but didactique, illustrant les différences entre les deux modèles considérés en les mettant en parallèle plutôt qu'en les opposant.

6.2 Méthodes

Nous avons proposé en premier lieu une représentation graphique schématique de l'ensemble des individus à risque au cours du temps dans chacun des deux modèles, qui permet une compréhension intuitive et sans formulation mathématique de ces ensembles qui interviennent dans l'estimation des paramètres des modèles de régression.

Nous avons illustré cette représentation graphique sur un exemple fictif où 16 patients sont randomisés entre deux bras de traitement (A et B). Dans un souci de simplification, seules deux causes d'événement ont été considérées, en l'absence de censure. Cet exemple comprend suffisamment peu de sujets pour que les rangs d'événement et leur cause pour chaque individu puissent être rapportés dans un tableau. Les graphiques représentant l'ensemble des individus à risque pour chacun des modèles peuvent donc être facilement retrouvés.

Dans un second temps, nous avons appliqué cette représentation aux données d'un essai thérapeutique conduit dans la leucémie myéloïde aiguë (Castaigne et al., 2004). Dans cet essai, 592 patients ont été randomisés entre trois bras de traitement, le délai sans rechute étant le critère de jugement principal. Par souci de simplification, nous n'avons retenu que la comparaison des bras A (standard) et C (induction séquentielle) pour les individus de moins de 50 ans, soit 190 sujets. Nous avons montré que le choix d'un modèle de Cox cause-spécifique conduisait à une conclusion différente de celle obtenue à l'aide d'un modèle de Fine & Gray quant au bénéfice ajusté du bras C sur le bras A.

6.3 "How to improve our understanding of competing risks analysis ?"

Cet article a été soumis pour publication à *Journal Of Clinical Oncology*.

How to improve our understanding of competing risks analysis?

Aurélien Latouche, Matthieu Resche-Rigon, Sylvie Chevret, Raphaël Porcher

Affiliation**:** Département de Biostatistique et Informatique Médicale, ERM 0321-INSERM, Hôpital Saint Louis, Université Paris 7, AP-HP, Paris

Acknowledgments of research support: We wish to acknowledge Pr Sylvie Castaigne and Pr Hervé Dombret for the availability of the ALFA 9000 trial data

Corresponding author : Aurélien Latouche, Département de Biostatistique et Informatique Médicale, Hôpital Saint Louis. 1, avenue Claude Vellefaux, 75475 Paris cedex 10, France.

Phone number : 33 1 42 49 97 42

Fax number : 33 1 42 49 97 45

Email address: aurelien.latouche@chu-stlouis.fr

Running head: Understanding competing risks

Key Words : competing risks; cumulative incidence; survival models

Abstract (234 words)

- PURPOSE: In the analysis of competing risks outcomes, there is still an open debate with regards to the use of a regression model, either the cause-specific Cox model or the model proposed by Fine and Gray for the hazard associated with the cumulative incidence function. Our purpose was to further detail the implications of either choice when estimating the benefit of a new treatment on delaying the outcome of interest in this competing risks setting.
- PATIENTS AND METHODS: Differences are exemplified first through the risk set associated with each model, using an easily understandable graphical representation. Then, differences are illustrated using a real data sample from the randomized clinical trial ALFA 9000 conducted in acute myeloid leukemia to assess the benefit on relapse-free interval of timed-sequential induction over the standard treatment.
- RESULTS: Estimated adjusted treatment benefit could be modified by the handling of competing risks outcomes, with disappearance of statistical significance when using a cause-specific model (p= 0.13) as compared to the use of Fine and Gray model (p= 0.05).
- CONCLUSION: Differences in quantities of interest could help in selecting the regression model to be used. If effect on instantaneous hazard is that of interest, the cause-specific model is to be used, while if we are only interested in modeling the effect of covariate on the probability of a particular outcome, the Fine and Gray model is that of choice.

Introduction

In evaluating the clinical benefit of treatments in oncology, the gold standard is to demonstrate a benefit in terms of overall survival (OS). Nevertheless, a recent analysis of oncology drugs approved in the USA through the regular process in the last 13 years showed that two thirds of them were based on other endpoints than survival[1]. This is also true with regards to evaluation of new treatment strategies in malignant hematological diseases. In this setting, indeed, expected benefits in survival are likely to be small, requiring the conduct of large randomized clinical trials[2]. Therefore, when such large trials appear unrealistic, alternative end points are more and more used, also justified by the complex course of these patients. Actually, since these patients can fail from many causes such as treatment-related complications, relapse, or death, we are also interested in estimating the effect of the new treatment on either cause of failure. Accordingly, besides overall survival, one of the most used end points in evaluating new therapies in malignant hematological diseases is time to relapse, also called the relapse-free interval (RFI). Nevertheless, statistical analysis of RFI should be handled differently from that of OS.

The paper is organized as follows. First, we present a rapid overview of competing risks analysis, exemplified on the analysis of RFI by contrast to OS. Then, in order to understand the fundamental differences between available regression models, we present a graphical representation of the risk set associated with each model. Third, we illustrated, using a real sample data, the differences reached by using both approaches in the estimation of the benefit on RFI of timed-sequential induction in acute myeloid leukemia. Finally, we provide some help for practitioners in choosing either modeling approach.

An overview of methods for competing risks

There are several reasons why the statistical analyses of OS and of RFI are actually different. First, while the observation process of death can be disturbed only by the interruption of follow-up (generating "censored observations"), the occurrence of relapse itself, and not only the observation process, can be also suppressed by the previous occurrence of death in remission. In other words, in the analysis of OS, there is only one risk (death) acting on the population, whereas in the analysis of RFI, there is a so-called "competing risks" setting, defined by the simultaneous action of two risks, namely death in remission and relapse.

Secondly, when reporting analysis of the potential benefit of treatment on the outcome, a graphical display of the probabilities of failure causes against time is useful. Therefore, to display cumulative probability of the outcome over time from the sample of enrolled patients, nonparametric methods are commonly used in both settings. Nevertheless, while the worldwide known Kaplan-Meier method is used to estimate the OS curve, the literature has emphasized the inappropriateness of this method when estimating the cumulative probability of relapse[3-6]. Actually, the cumulative incidence function (CIF) of relapse appears more attractive, in the sense that it takes into account the competing risk of death in remission. In other words, it is interpretable as the "crude" probability of relapse when at least two competing risks act on the population (without further assumptions about the dependence among the failure times), while the Kaplan-Meier approach achieves an estimate of the "net" probability of relapse, i.e., in the hypothetical setting where the risk of relapse is the only one acting on the population, assuming independent competing risks[7]. Indeed, without such an assumption, the Kaplan-Meier estimate has no probability interpretation[8].

Finally, when evaluating the treatment benefit adjusting for potential confounders, regression models have to be used. Despite numerous available models, the Cox semi-parametric proportional hazards model is the quasi-unique way to perform prognostic analyses for OS,

whatever the causes, in routine practice[9]. It expresses the probability of death in an infinitesimal interval of time, conditionally on being alive at the onset of the time interval, as a multiplicative function of covariates. By contrast, in the competing risks setting provided by the analysis of RFI, two major tools are available. The cause-specific hazard function for relapse is the instantaneous risk of occurrence of relapse, conditional on being alive and free of relapse. It can be directly modeled, as previously, by using a Cox model, where the deaths in remission are censored at the time of their death[10].

Otherwise, some authors have proposed to use regression models for the hazard associated with the CIF[6], such as the semi-parametric model developed by Fine and Gray[11]. The effect of the covariate in such a model can be directly translated in terms of CIF. It has been yet used in the prognostic evaluation on neutrophils recovery after bone marrow transplantation[12], breast cancer death[13,14] or metastasis[15], and even RFI[16,17]. The general interest in these quantities as found in papers published in the *Journal of Clinical Oncology* until November 2004, is summarized in Table 1, suggesting a clear preference for cause-specific or Cox model over the Fine and Gray model, though the parallel use of cumulative incidence functions is frequently reported. Of note, most of the review papers on practical use of competing risks analysis have emphasized the use of cumulative incidence functions, while overlooking regression models[3,6,18-20]. Nevertheless, before claiming that Cox model is misused in this context and that regression models in competing risks must be based on regression model for the risk associated with CIF, we aim at further detailing the implications of either choice when estimating the benefit of a new treatment on delaying relapse.

Illustrative example : Representation of the Risk Sets

Let us consider an illustrative example, where N=16 patients are randomized between two treatment arms (A and B). Suppose that the main outcome measure is a failure time, while

patients are subject to another competing risks outcome that prevents the observation of the former. For simplicity, assume that there is no censoring, that is, all patients were followed up until either outcome happens.

Table 2 displays the observed ranks of time to outcomes after randomization (either primary or competing risks), that are ordered increasingly irrespective of the treatment arm. Regression analysis is based on iterative computations over time, involving the risk set at each ordered outcome of interest. Such a risk set is defined at any particular time *t*, by the set of individuals who are still exposed to the outcome of interest at *t*. Nevertheless, there are some discrepancies between the risk sets computed in the Cox model and the CIF based model: In the Cox cause-specific hazard model, an individual is at risk at time *t* only if he has not experienced any outcome yet. In the CIF-based model, an individual is at risk at time *t* if he has not experienced the outcome of interest before *t*, which also includes individuals having already experienced the competing risks outcome before *t*. Despite this unnatural definition, this risk set accommodates with valid statistical properties[11].

The Figure 1 represents the risk sets associated with either regression approach. It is a abacus-like plot of the Table 2. Actually, it exhibits that, when using the Cox cause-specific model, the risk set decreases more rapidly over time due to the disappearance of outcomes of interest and competing risks outcomes, than when using the CIF based approach, that is only affected by the occurrence of the outcomes of interest. This could result in differences in estimates of treatment effect, due to the fact that patients who experienced the competing risks outcome will belong to all risk sets of the Fine and Gray regression analysis while they will be excluded from the risk set at the time they fail in the Cox model. In case of distinct effect of the treatment under study on the two competing risks, for instance, one could expect a benefit of the treatment through the use of Cox model while no benefit will be shown through the use of the CIF based approach, and conversely.

Of note, when analyzing the outcome of interest, results are similar whether the competing causes are distinguished or mixed altogether.

A real data set example: the ALFA 9000 trial

We reanalyzed the previously cited randomized clinical trial[16], conducted by the Acute Leukemia French Association (ALFA) cooperative group, where 592 patients with either de novo or secondary acute myeloid leukemia (AML) were randomized between one of the three following induction arms: reinforced standard "3+7" induction (arm A, 197 patients including 110 aged less than 50 years); double induction (arm B, 198 patients including 114 aged less than 50 years); or timed-sequential induction (arm C, 197 patients including 121 aged less than 50 years). The main end point was RFI, calculated from the date of the first CR to the date of the first relapse. In the primary analysis, authors accounted for competing risks deaths and allogeneic bone-marrow transplantations in first complete remission (CR) using cumulative incidence curves, then compared by the Gray test while the Fine and Gray model was used to estimate CIF associated-hazards ratio (or sub-distribution hazard ratio, SHR). Overall differences in RFI were not significant among the 3 randomization arms (P=0.39 and 0.15 when arm B and C were compared to arm A, respectively, using the Gray test) ; by contrast, in patients aged 50 years or less, RFI was significantly improved from the arm C as compared to the control arm A (P=0.038, using the Gray test)[16].

We decided to perform separate analyses using either the Cox cause-specific hazards model, or the Fine and Gray model for the CIF associated hazards. For simplicity, we focused on the comparison in the AML patients aged less than 50 years who were randomly allocated in the reinforced standard "3+7" induction (A) and the timed-sequential induction (C) arms.

Of the 231 patients aged 50 years or less allocated in arm A or arm C, 41 (17 in arm A and 24 in arm C) did not achieve CR, due to either resistant disease or induction death. Further

analyses will deal with the remaining 190 patients. After CR, 90 relapsed (52 in arm A and 38 in arm C), 35 received allogeneic bone marrow transplantation in first CR (16 in arm A and 19 in arm C) and 16 died in first CR (8 in either arm A or C). Figure 2 displays the schematic representation of the risks sets according to the model. At the beginning, there are 190 patients in both risk sets.

Table 3 displays the estimation of the estimated hazard ratio of treatment arm and potential confounders (namely, sex, age, karyotypic abnormalities and de novo AML) using both regression models. We note that benefit of the timed-sequential induction arm (arm C) is similarly estimated in both approaches, while discrepancies in estimates were found for the effect of age and unfavorable karyotype. When multivariable models were fit, results differed according to the approach: The Cox cause-specific model only retained unfavorable karyotype as associated with RFI while the CIF based Fine and Gray model retained both treatment arm and age as associated with RFI (Table 3). This could rely on the differences in competing risks outcomes according to patient subsets. In fact, while treatment arm has no significant effect on the CIF of BMT (p= 0.67) nor on the CIF of death in CR (p= 0.95), there was a significant decreased incidence of death in CR in the subset of patients age 26 or less (p= 0.03) (Figure 3) as well as non significant trends towards increase in CIF or both BMT (p= 0.18) and death in CR (p= 0.30) in the subset of patients with unfavorable karyotype. In other words, a decrease (resp., increase) in the cumulative incidence function of relapse in one subset of patients could be due either to a physiological effect of the treatment on that risk, or to an increase (resp., decrease) in the probability of the competing risks outcomes.

Discussion

We have attempted to clarify the issues raised by the use of regression models in the setting of competing risks. Actually, while there is a general consensus over the use of cumulative

incidence function estimates when displaying outcomes over patients subsets, the choice of regression modeling in order to estimate the benefit of treatment or prognostic covariates in such a setting is still an open issue.

In the analysis of competing risks data, the first question should be to address the competing risks setting itself. Such a setting involves the simultaneous exposure to more than one risk. In fact, when one exposure is clearly delayed over the other, for instance if no relapse is observed within the first 100 days while allografted patients are exposed to acute graft versus host disease, such a competition is questionable. Nevertheless, it is clear from Figure 1 that, in this case, risks sets will be similar so that both regression analyses will achieve similar results.

Secondly, the competing risks outcomes should be clearly discussed with regards to the independence assumption with the process of interest. For instance, when assessing survival of hepato-cellular carcinoma, hepatitis transplantation could appear as likely related with the risk of death, so that ignoring the competition and treating them as censored observations is misleading.

Third, in case of more than one simultaneous risks acting on the population, and contrary to what is commonly stated, it appears that both approaches can be valuable, since they focus on distinct quantities of interest. If we are interested in estimating the effect of the treatment on the hazard of a specific outcome (such as relapse), conditionally on being alive at the time, the cause-specific approach appears of prime interest. By contrast, if we focus on estimating the effect of the treatment on the crude probability of experiencing that outcome while other risks act on the population, the Fine and Gray approach should be used. Differences in risk sets reached by the use of both models are exemplified on a small data set.

Nevertheless, some could favor the cause-specific approach due to the following points. First, the cause-specific hazard can be expressed using the Cox model, which is widely used in the

medical literature. Moreover, it appears more clinically understandable to assume a multiplicative effect of the treatment (or any covariate) on the cause-specific hazard, i.e. on the instantaneous risk of relapse, conditional on being alive and free of relapse. In other words, it seems natural that the physiological effect of a treatment or any prognostic exposure would be to reduce or increase the probability of the outcome of interest at any time, conditionally on being still alive at that time.

On the contrary, a decrease (resp., increase) in the cumulative incidence function could be due either to a physiological effect of the exposure or to an increase (resp., decrease) in the probability of the competing failure causes. As illustrated in the AML90 trial analyses, the overall effect of unfavorable karyotype on CIF associated risk could be "masked" by the simultaneous increase in competing risks outcomes: Patients with unfavorable karyotype were more likely to die in CR or being transplanted, so that the resulting effect is to erase the increase in CIF of relapse. By contrast, young patients were poorly exposed to death, so that resulting effect on CIF of relapse could be artificially increased. Therefore, to better interpret and analyze the effect of treatment (or any other covariate) on the cumulative incidence of relapse using the Fine and Gray model, it is mandatory to display concomitantly the estimated effects of that covariate on either competing risks outcomes, as pointed out by Pepe[3].

Finally, the choice of either approach could rely on the main question of interest. Sometimes, our primary concern is to model the instantaneous risk of developing the outcome of interest. For instance, in the previous example, estimating the influence of treatment on the instantaneous risk of developing relapse could be considered as the most important question. In this case, one should use the Cox cause-specific model, that allows to estimate the hazard ratio of covariates as a measure of association between this risk and the covariate. By contrast, our primary concern can be to model the prevalence of the outcome of interest, as epidemiologists do. This is notably the case when the occurrence of the outcome of interest is

restricted over a short time period (for instance, acute graft versus host disease following allogeneic bone marrow transplantation) or on a particular exposure (for instance, death in intensive care unit). In these cases, the model of interest is clearly the CIF-associated model, such as that proposed by Fine and Gray.

References

1. Marwick C: Can disease-free survival be a surrogate for overall survival for drug approvals? J Natl Cancer Inst 96:173, 2004

2. Yusuf S, Collins R, Peto R: Why do we need some large, simple randomized trials? Stat Med 3:409-22, 1984

3. Pepe MS, Mori M: Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? Stat Med 12:737-51, 1993

4. Bradburn MJ, Clark TG, Love SB, et al: Survival analysis Part III: multivariate data analysis -- choosing a model and assessing its adequacy and fit. Br J Cancer 89:605-11, 2003

5. Bradburn MJ, Clark TG, Love SB, et al: Survival analysis part II: multivariate data analysis--an introduction to concepts and methods. Br J Cancer 89:431-6, 2003

6. Satagopan JM, Ben-Porat L, Berwick M, et al: A note on competing risks in survival data analysis. Br J Cancer 91:1229-35, 2004

7. Tsiatis AA: Competing risks, in Armitage P, Colton T (eds): Encyclopedia of Biostatistics. New York, John Wiley & Sons, 1998, pp 824-834

8. Andersen PK, Abildstrom SZ, Rosthoj S: Competing risks as a multi-state model. Stat Methods Med Res 11:203-15, 2002

9. Cox DR: Regression models and life tables. J Roy Stat Assoc B 34:187-220, 1972

10. Prentice RL, Kalbfleisch JD, Peterson AV, Jr., et al: The analysis of failure times in the presence of competing risks. Biometrics 34:541-54, 1978

11. Fine JP, Gray RJ: A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 94:496-509, 1999

12. Shpall EJ, Quinones R, Giller R, et al: Transplantation of ex vivo expanded cord blood. Biol Blood Marrow Transplant 8:368-76, 2002

13. Martelli G, Miceli R, De Palo G, et al: Is axillary lymph node dissection necessary in elderly patients with breast carcinoma who have a clinically uninvolved axilla? Cancer 97:1156-63, 2003

14. Robson ME, Chappuis PO, Satagopan J, et al: A combined analysis of outcome following breast cancer: differences in survival based on BRCA1/BRCA2 mutation status and administration of adjuvant treatment. Breast Cancer Res 6:R8-R17, 2004

15. Colleoni M, O'Neill A, Goldhirsch A, et al: Identifying breast cancer patients at high risk for bone metastases. J Clin Oncol 18:3925-35, 2000

16. Castaigne S, Chevret S, Archimbaud E, et al: Randomized comparison of double induction and timed-sequential induction to a "3 + 7" induction in adults with AML: long-term analysis of the Acute Leukemia French Association (ALFA) 9000 study. Blood 104:2467-74, 2004

17. de Botton S, Coiteux V, Chevret S, et al: Outcome of childhood acute promyelocytic leukemia with all-trans-retinoic acid and chemotherapy. J Clin Oncol 22:1404-12, 2004

18. Alberti C, Metivier F, Landais P, et al: Improving estimates of event incidence over time in populations exposed to other events: application to three large databases. J Clin Epidemiol 56:536-45, 2003

19. Gooley TA, Leisenring W, Crowley J, et al: Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Stat Med 18:695-706, 1999

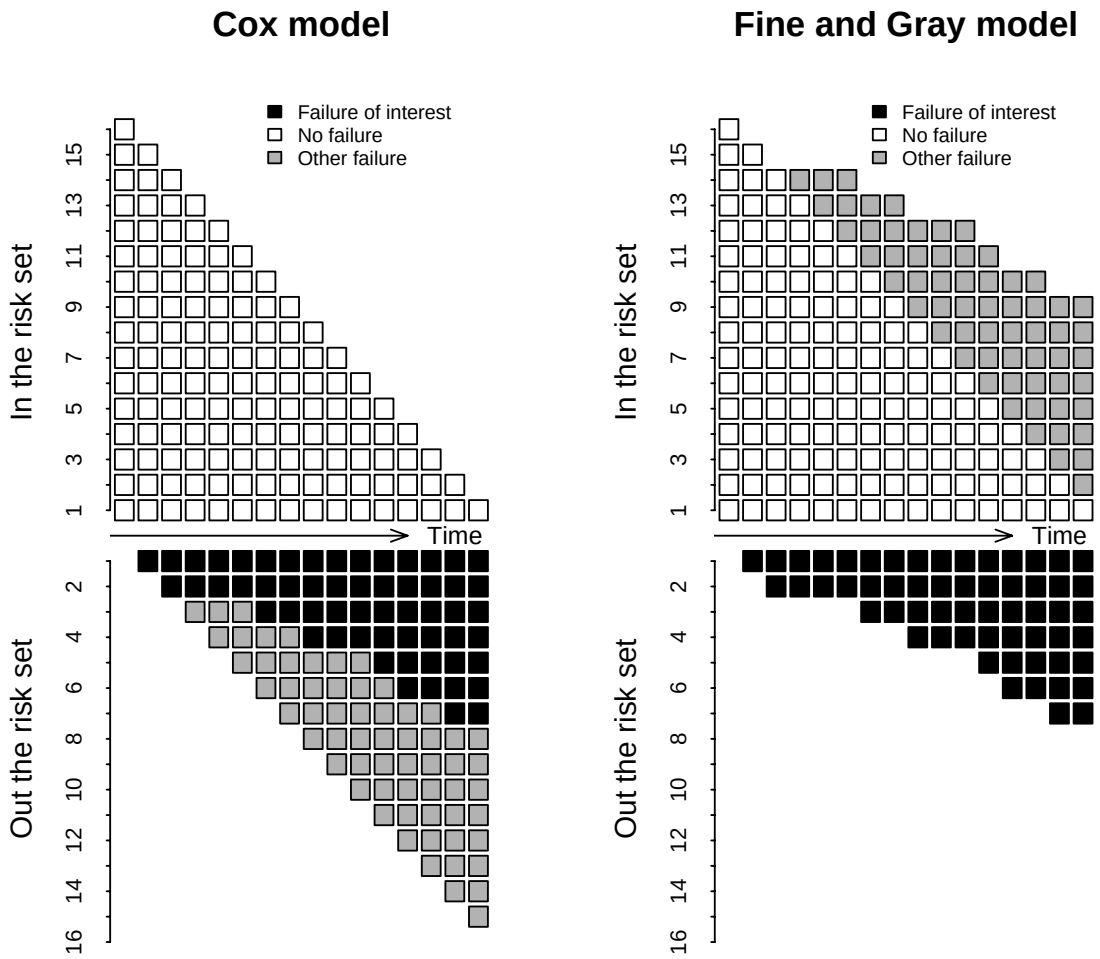
20. Machin D: On the evolution of statistical methods as applied to clinical trials. J Intern Med 255:521-8, 2004

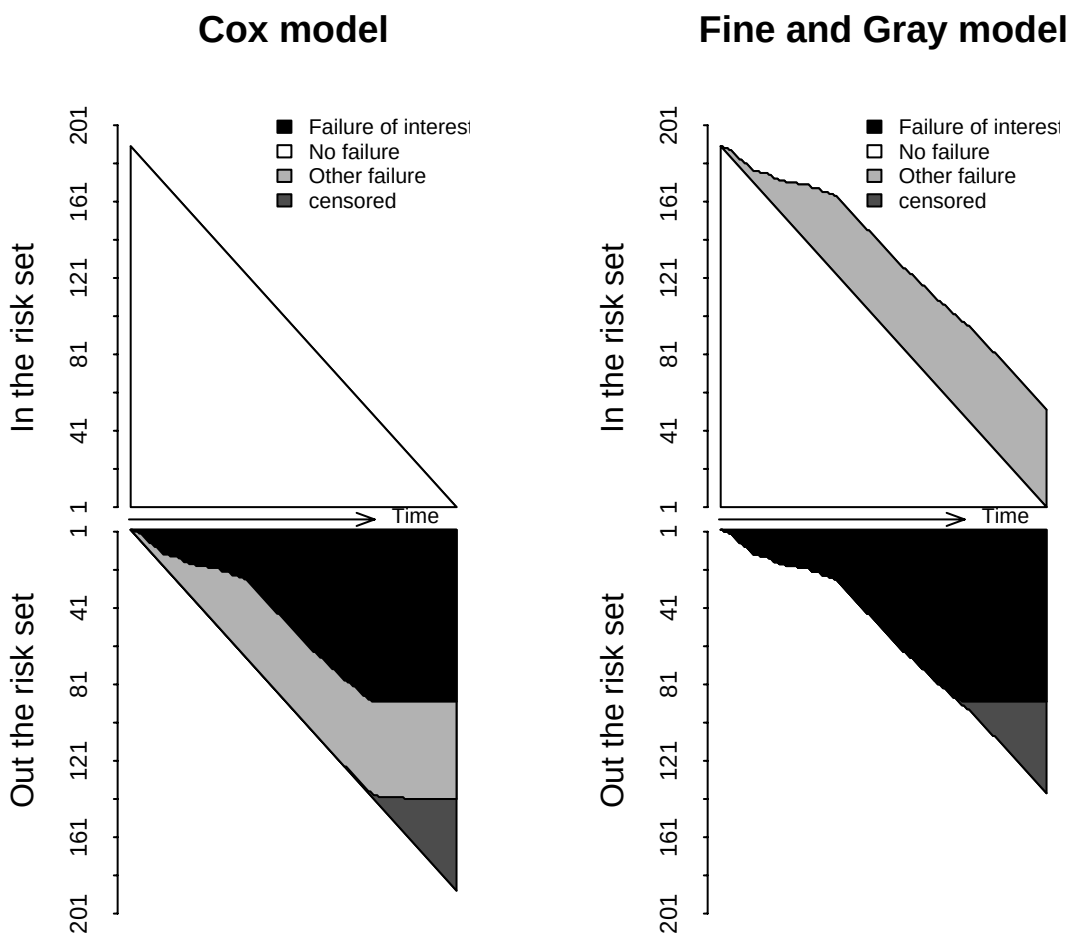
Figure legends

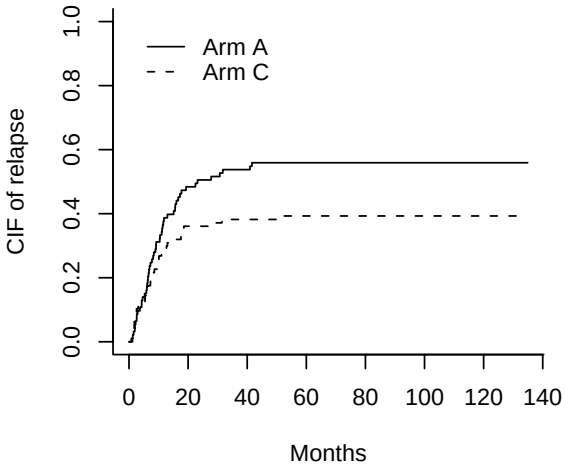
Figure 1: Fictive data set- Schematic representation of the risk sets of both regression models

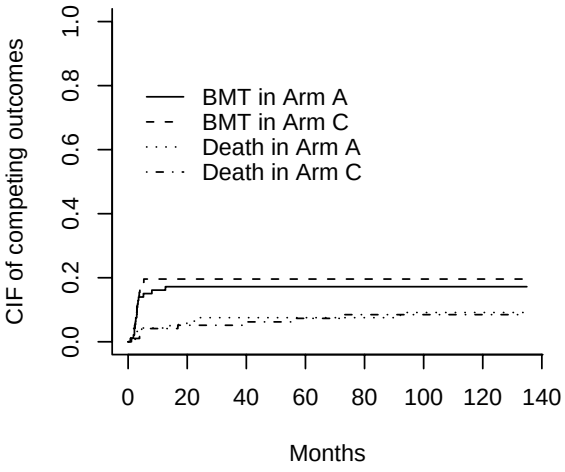
Figure 2: ALFA 9000 Trial- Schematic representation of the risk sets of both regression models

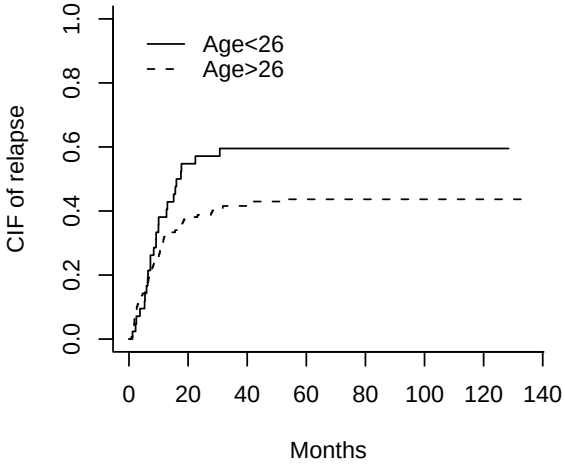
Figure 3: ALFA 9000 Trial- Estimated cumulative incidence functions of relapse and competing risks outcomes according to arm (C vs. A), and age (<26 vs. > 26)


Cox model
Fine and Gray model
Failure of interest
No failure
Other failure
In the risk set
Out the risk set
Time
1
3
5
7
9
11
13
15
2
4
6
8
10
12
14
16


Cox model
Fine and Gray model
Failure of interest
No failure
Other failure
censored
In the risk set
Out the risk set
Time
1
41
81
121
161
201


CIF of relapse
0.0
0.2
0.4
0.6
0.8
1.0
Arm A
Arm C
0
20
40
60
80
100
120
140
Months


CIF of competing outcomes
0.0
0.2
0.4
0.6
0.8
1.0
BMT in Arm A
BMT in Arm C
Death in Arm A
Death in Arm C
0
20
40
60
80
100
120
140
Months


CIF of relapse
0.0
0.2
0.4
0.6
0.8
1.0
Age<26
Age>26
0
20
40
60
80
100
120
140
Months

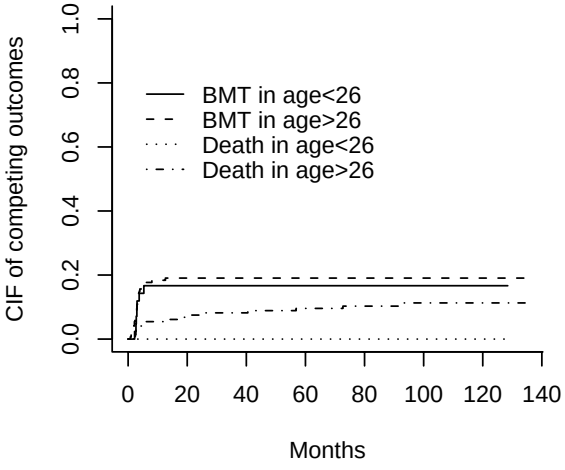

CIF of competing outcomes
0.0
0.2
0.4
0.6
0.8
1.0
BMT in age<26
BMT in age>26
Death in age<26
Death in age>26
0
20
40
60
80
100
120
140
Months

Table 1: Journal of Clinical Oncology online search in text, abstract, or title, from January 1983 up to November, 2004

Query	Number of references found
"Competing risk(s)"	111
"Competing risk(s)" AND "cause specific"	30
"Competing risk(s)" AND "Cox"	76
"Competing risk(s)" AND "cumulative incidence"	72
"Competing risk(s)" AND "Fine and Gray"	8
"Competing risk(s)" AND "cumulative incidence" AND "Cox"	53

Table 2: Illustrative example : Fictive data set

Rank of failure	Type of outcome 1: Outcome of interest 2: Competing risks outcome	Treatment arm
1	1	B
2	1	B
3	2	B
4	2	B
5	2	A
6	1	A
7	2	B
8	1	B
9	2	A
10	2	A
11	1	A
12	1	B
13	2	B
14	1	A
15	2	A
16	1	A

Table 3: ALFA 9000 Trial Data. Estimation of the benefit of time-sequential induction on RFI according to the regression model

Measure of treatment benefit (95%CI)*	Cox cause-specific hazard model	CIF based hazard Fine and Gray model**
Univariable models		
Treatment arm C	0.80 (0.64-0.98); p=0.032	0.80 (0.65-0.99); p=0.038
Age ≥ 26 (Q1)	0.72 (0.46-1.15) ; p= 0.17	0.66 (0.42-1.04); p=0.08
De novo AML	0.85 (0.27-2.70); p=0.79	1.74 (0.52-5.83); p=0.37
Female gender	1.07 (0.71-1.62); p=0.75	1.00 (0.66-1.51); p=0.98
Unfavorable karyotype	2.43 (1.26-4.69); p= 0.008	1.43 (0.72-2.85); p= 0.30
Multivariable model		
Treatment arm C	0.83 (0.64-1.06); p=0.13	0.79 (0.60-1.00); p=0.05
Age ≥ 26 (Q1)	0.63 (0.36-1.10); p=0.10	0.56 (0.33-0.95); p=0.032
Unfavorable karyotype	2.36 (1.21-4.57); p= 0.01	1.57 (0.81-3.04); p= 0.18

* either cause-specific hazard ratio (HR) or CIF-associated hazard ratio (also called the subdistribution hazard ratio, SHR)

** with death or BMT in CR as competing risks outcomes

Chapitre 7

Conclusions et Perspectives

Dans ce travail, nous nous sommes intéressés aux modèles de régression en présence de compétition. Deux approches ont été envisagées, l'une basée sur la fonction de risque cause-spécifique et l'autre sur la fonction de risque associée à la fonction d'incidence cumulée (ou fonction de risque de sous-répartition). Dans les deux cas, nous avons considéré une formulation à risques proportionnels. Les quantités modélisées sont cependant fondamentalement différentes. La première représente un risque instantané (ou une intensité dans le cadre des modèles multi-états) et alors que la seconde est reliée de façon univoque à la fonction d'incidence cumulée. Nous avons insisté sur le fait que l'effet d'une covariable ne se traduit pas de manière similaire sur ces deux quantités, et que l'intérêt des modèles de régression basés sur la fonction d'incidence cumulée réside dans le fait que l'effet d'une covariable sur la fonction de risque de sous répartition, se traduit directement sur la fonction d'incidence cumulée.

Afin que de tels modèles soient couramment utilisés, le paquet `cmprsk`[1] implémente le modèle de Fine & Gray pour le logiciel R (R Development Core Team, 2004) et permet son utilisation en recherche clinique. Pepe and Mori (1993) ont insisté sur l'importance de tels logiciels, qui rendent possible une utilisation plus large de nouvelles méthodologies, et ainsi permettent d'utiliser le modèle le plus adapté. Cependant, lorsque de tels programmes sont disponibles, il est tentant d'utiliser le modèle le plus récent en oubliant que les autres modèles (ici, le modèle de Cox) peuvent toujours être appropriés à certaines situations.

L'opposition qui a été implicitement introduite entre les deux modèles repose notamment sur la confusion entre une modélisation brute, pour laquelle le modèle de Fine & Gray serait préférable parce que l'on ne peut pas utiliser l'estimateur de Kaplan-Meier pour estimer la

[1] `http://biowww.dfci.harvard.edu/~gray`

probabilité d'événement, et une modélisation nette, pour laquelle le modèle de Cox devrait être utilisé. Or, dans le premier cas, les quantités modélisées par le modèle de Cox et le modèle de Fine & Gray sont toutes les deux identifiables à partir des observations, sans hypothèses autres que leur relation avec les covariables. L'argument que l'hypothèse de risques latents indépendants n'est pas statistiquement vérifiable ne peut donc être utilisé ici pour choisir l'un des deux modèles. L'interprétation du risque cause-spécifique, en présence de toutes les causes ou en supposant qu'une seule cause agit, est ainsi indépendante de la modélisation utilisée (temps latents ou multi-états) et il convient juste de préter attention à la signification de cette intensité (en terme de probabilité nette) (Andersen et al., 2002; Hougaard, 2000).

L'implication d'un choix de modélisation a été étudiée dans le cadre de la planification d'un essai clinique ou d'une étude pronostique. Nous avons developpé une formule de calcul du nombre de sujet nécessaire en situation de compétition pour le modèle de Fine & Gray.

L'étude des conséquences d'un modèle mal spécifié pour la fonction de risque de sous répartition a également été conduite, en étudiant l'influence sur l'estimation du paramètre de régression dans le modèle de Fine & Gray quand le vrai modèle est un modèle à risques proportionnels pour la fonction de risque cause-spécifique. Nous avons ensuite étudié les implications de l'inclusion de covariables dépendantes du temps dans le modèle de Fine & Gray. Notre dernier article présente une synthèse didactique sur la stratégie d'utilisation des approches présentées dans ce travail de thèse.

Pour décrire une situation de risques compétitifs, plusieurs approches peuvent donc être utilisées de manières complémentaires plutôt qu'exclusives. Il semble important, comme le soulignent Fine and Gray (1999), de développer un modèle joint pour les fonctions d'incidence cumulée de toutes les causes d'événement en compétition. Dans ce cas, de part la relation univoque entre l'ensemble des fonctions de risque cause-spécifique et l'ensemble des fonctions d'incidence cumulée, la notion d'un choix entre les deux points de vue revient à un simple problème de paramétrisation de la vraisemblance des obervations.

Dans tous les cas où l'on modélise des quantités brutes, il est de plus préférable de toujours présenter les résultats obtenus pour toutes les causes d'événement en compétition, afin de décrire l'ensemble de l'information contenue dans l'échantillon. Ne présenter de résultats que pour l'une

des causes d'événements revient à ne permettre qu'une interprétation partielle de quantités qui ne sont définies que lorsque toutes les causes d'événement agissent (fonction d'incidence cumulée ou risque de sous-répartition) ou dont l'interprétation en termes de probabilités nettes nécessite des hypothèses supplémentaires (risque cause-spécifique).

Nous pensons aussi que le développement d'une modèle de régression pour la probabilité conditionnelle (Pepe and Mori, 1993) permettrait d'apporter une information complémentaire de celles déja disponibles. En effet la fonction de probabilité conditionnelle pour l'événement de type k représente la probabilité d'avoir expérimenté l'événement de type k avant l'instant t sachant que l'on n'a expérimenté aucun des autres événements en compétition. Il pourrait alors s'agir de la quantité d'intérêt pour les cliniciens.

Enfin, ce travail s'est limité à l'analyse des risques en compétition, c'est à dire d'événements mutuellement exclusifs. Cependant, on est souvent confronté à une situation de risques dits semi-compétitifs (Fine et al., 2001), où la survenue d'un événement terminal empêche celle des autres, mais pas l'inverse. Par exemple, dans le suivi des patients allogreffés, le décès empêche la rechute ou la maladie du greffon contre l'hôte (GvHD) et la rechute empêche la GvHD. Au contraire, un patient qui développe une GvHD peut avoir un risque accru de décès et un risque moindre de rechute (effet dit "greffon contre leucémie"). Afin de pouvoir analyser ce type de situation, il semble nécessaire de développer un modèle de régression en présence de risques semi-compétitifs. L'exemple précédent pourrait être modélisé par une approche à temps latents sous la contrainte $T_1 \leq T_2 \ldots \leq T_K$. Les modèles à risques semi-compétitifs apparaissent comme une extension naturelle des modèles à risques compétitifs, ne se limitant pas à la survenue du premier événement, permettant ainsi de modéliser des situations plus complexes mais fréquemment rencontrées en recherche clinique.

Bibliographie

Andersen, P. K., Abildstrom, S. Z., and Rosthoj, S. (2002). Competing risks as a multi-state model. *Statistical Methods in Medical Research*, 11(2) :203–215.

Andersen, P. K., Borgan, ., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Andersen, P. K., Klein, J. P., and Rosthoj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, 90(1) :15–27.

Bernoulli, D. (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir. In *Histoires et Mémoires de l'Académie Royale des Sciences*, pages 1–45. Paris.

Block, H. W. and Basu, A. P. (1974). A continuous bivariate exponential extension. *Journal of the American Statistical Association*, 69(348) :1031–1037.

Castaigne, S., Chevret, S., Archimbaud, E., Fenaux, P., Bordessoule, D., Tilly, H., de Revel, T., Simon, M., Dupriez, B., Renoux, M., Janvier, M., Miclea, J. M., Thomas, X., Bastard, C., Preudhomme, C., Bauters, F., Degos, L., and Dombret, H. (2004). Randomized comparison of double induction and timed-sequential induction to a "3 + 7" induction in adults with aml : long-term analysis of the acute leukemia french association (alfa) 9000 study. *Blood*, 104 :2467–2474.

Commenges, D. (1999). Risques compétitifs et modèles multi-états en épidémiologie. *Revue d'Epidémiologie et Santé Publique*, 47 :605–611.

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

Fine, J., Jiang, H., and Chappell, R. (2001). On semi-competing risks data. *Biometrika*, 88 :907–920.

Fine, J. P. (2001). Regression modelling of competing crude failure probabilities. *Biostatistics*, 2 :85–97.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446) :496–509.

Fleming, T. R. and Harrington, D. (1991). *Counting processes and survival analysis*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York.

Gray, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, 116 :1141–1154.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer-Verlag, New York.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York.

Lachin, J. M. and Foulkes, M. A. (1986). Evaluation of sample size and power for analysis of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics*, 42 :507–519.

Pepe, M. S. and Mori, M. (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data ? *Statistics in Medicine*, 12 :737–751.

Peterson, A. V. (1976). Bounds for a joint distribution function with fixed sub-distribution functions : Application to competing risks. *Proceeding of the National Academy of Sciences*, 73(1) :11–13.

Pintilie, M. (2002). Dealing with competing risks : testing covariates and calculating sample size. *Statistics in Medicine*, 21 :3317–3324.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34 :541–554.

R Development Core Team (2004). *R : A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.

Sampford, M. R. (1952). The estimation of response time distributions. II. Multistimulus distributions. *Biometrics*, 8 :307–353.

Schmoor, C., Sauerbrei, W., and Schumacher, M. (2000). Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine*, 19 :441–452.

Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*, 39 :499–503.

Tsiatis, A. A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceeding of the National Academy of Sciences*, 72(1) :20–22.

Tsiatis, A. A. (1998). Competing risks. In Armitage, P. and Colton, T., editors, *Encyclopedia of Biostatistics*, pages 824–834. John Wiley & Sons, New York.