

MODÉLISATION CONJOINTE DE DONNÉES LONGITUDINALES ET DE TEMPS D'ÉVÉNEMENTS COMPÉTITIFS

Loïc Ferrer ¹ & Cécile Proust-Lima ²

¹ *INSERM U897, ISPED, Équipe Biostatistique, Université de Bordeaux Segalen, 146 rue Léo Saignat, 33076 Bordeaux Cedex, FRANCE – loic.ferrer@inserm.fr*

² *INSERM U897, ISPED, Équipe Biostatistique, Université de Bordeaux Segalen, 146 rue Léo Saignat, 33076 Bordeaux Cedex, FRANCE – cecile.proust-lima@inserm.fr*

Résumé. La modélisation conjointe de données longitudinales et données de survie est de plus en plus utilisée dans les essais cliniques sur le cancer. Dans le cancer de la prostate par exemple, ces modèles permettent d'étudier l'effet des facteurs pronostiques sur le temps de récurrence, et le lien entre mesures longitudinales du niveau de l'antigène spécifique de la prostate (PSA) et temps de récurrence. En pratique, il existe plusieurs types de récurrences. Il peut être intéressant de les distinguer afin de rechercher, par exemple, quelles covariables impactent le risque d'avoir une rechute distante ou un décès du cancer de la prostate, indépendamment de leur effet sur le risque de rechute locale, ou bien de prédire spécifiquement le risque de rechute locale. Nous présentons un modèle conjoint pour un marqueur longitudinal et des événements compétitifs qui se décompose en deux sous-modèles : un sous-modèle linéaire mixte pour les données longitudinales (mesures répétées des PSA), et un sous-modèle à risques proportionnels spécifique à chaque cause pour les temps d'événements compétitifs (rechutes cliniques par type), tous deux liés par des effets aléatoires partagés. Nous utilisons l'approche du maximum de vraisemblance et l'algorithme EM notamment pour estimer les paramètres du modèle. Plusieurs spécifications de ce modèle sont présentées, à l'aide de différentes structures d'association entre processus longitudinal et de survie. Nous appliquons ces modèles à l'aide d'exemples détaillés sur deux cohortes d'hommes suivis pour un cancer de la prostate après une radiothérapie. Nous concluons sur l'utilité d'un modèle conjoint à risques compétitifs dans ce cadre.

Mots-clés. Modèles conjoints, Risques compétitifs, données longitudinales, cancer de la prostate

Abstract. Joint modeling of longitudinal data and survival data has been increasingly used in clinical trials on cancer. In prostate cancer for example, these models permit to study the effect of prognostic factors on the time of relapse, and the link between longitudinal measures of current level of prostate-specific antigen (PSA) and time of relapse. In practice, there are different types of relapse. It might be interesting to distinguish them and to research, for example, which covariates impact the risk to have a distant recurrence or a death from prostate cancer separately of their effect on the risk of local

recurrence, or to predict the risk of local clinical recurrence. We present a joint model for a longitudinal marker and competing events which is divided into two submodels : a linear mixed submodel for longitudinal data (repeated measures of PSA), and a cause-specific proportional hazards submodel for the competing events times (clinical recurrences by type), both linked by shared random effects. We use the maximum likelihood approach and the EM algorithm to estimate model parameters. Several specifications of this model are presented, using different structures of association between longitudinal and survival process. We apply these models using detailed examples on two cohorts of patients followed-up after a radiation therapy for a localized prostate cancer. We conclude on the usefulness of a joint model of competing risks in this context.

Keywords. Joint models, competing risks, longitudinal data, prostate cancer

1 Contexte

Dans les études longitudinales, des données de marqueurs sont collectées à des temps de mesures répétés jusqu'à la survenue d'un événement. Souvent, il est d'intérêt d'étudier le lien entre ces mesures et la survenue de l'événement. Dans le cancer de la prostate par exemple, l'antigène spécifique de la prostate (PSA) est un marqueur couramment utilisé par les cliniciens pour les patients atteints d'un cancer de la prostate localisé et traités par radiothérapie externe. Taylor et al. (2005) ainsi que Proust-Lima et al. (2014) ont montré, à travers divers types de modèles conjoints, que la dynamique du biomarqueur, le niveau PSA pré-traitement, ainsi que le score de Gleason et T-stage, qui mesurent respectivement l'agressivité des cellules cancéreuses et l'étendue de la tumeur, sont des facteurs de risque de la progression, et permettent de prédire le risque de rechute clinique.

Cependant, il n'existe pas un unique type de rechute clinique après un cancer, et la prise en charge des patients varie largement suivant la nature de cette rechute. Il est donc nécessaire de modéliser séparément les types de rechute clinique pour mieux évaluer l'effet des facteurs pronostiques de progression et prédire les risques de rechute en distinguant par exemple les rechutes locales des rechutes distantes ou à métastases.

L'objectif de ce travail est donc de présenter un modèle conjoint permettant de modéliser les données répétées d'un biomarqueur et des événements compétitifs. Il existe différents types de modèles conjoints : les modèles two-stage, qui, à partir des prédictions faites sur le processus longitudinal, étudient leur impact sur le processus de survie ; les modèles conjoints à effets aléatoires partagés, qui lient les deux processus à l'aide d'effets aléatoires ; et les modèles conjoints à classes latentes, plus récents, qui prennent en compte la corrélation entre les processus longitudinal et de survie par l'identification de sous-populations de sujets. Ici, on se place dans le cadre de modèles conjoints à effets aléatoires partagés pour risques compétitifs (Elashoff et al. (2008) et Williamson et al. (2008)). Nous rappelons dans la section suivante la méthodologie des modèles à effets aléatoires partagés standards puis introduisons leur extension à la prise en compte d'événements compétitifs.

Nous décrivons leur estimation par maximum de vraisemblance. Puis nous appliquons cette méthodologie à la progression de cancer de la prostate après une radiothérapie.

2 Modèle conjoint à effets aléatoires partagés

Pour la suite, introduisons quelques notations. On note T_i^* le vrai temps d'événement pour le sujet i , T_i le temps observé, et C_i le temps de censure. On a donc $T_i = \min(T_i^*, C_i)$ et $\delta_i = I(T_i^* \leq C_i)$ l'indicateur d'événement, avec $\delta_i = 1$ si l'événement est observé, $\delta_i = 0$ sinon. Pour chaque sujet i , on observe $Y_i = (Y_i(t_{i1}), \dots, Y_i(t_{in_i}))$ les n_i mesures répétées du biomarqueur, collectées aux temps de mesures $(t_{i1}, \dots, t_{in_i})$ jusqu'au temps observé T_i . On reprend certaines notations de l'article de Sène, Bellera et Proust-lima (2014), avec X les covariables, dont X_{Li} le vecteur des covariables impactant sur le sous-modèle longitudinal et X_{Si} le vecteur des covariables impactant sur le sous-modèle de survie.

Le modèle conjoint est décomposé en deux sous-parties : un sous-modèle longitudinal et un sous modèle de survie, tous deux liés par des effets aléatoires partagés.

2.1 Sous-modèle longitudinal

Pour modéliser les mesures répétées du biomarqueur, on utilise un modèle linéaire mixte. Sous des hypothèses gaussiennes, on suppose que les vraies (non-observées) mesures répétées du biomarqueur $Y_i^*(t)$ sont expliquées par des covariables que l'on souhaite étudier au niveau de la population (avec des effets fixes) et au niveau individuel (avec des effets aléatoires qui prennent en compte la corrélation des mesures répétées chez un même individu) :

$$\begin{aligned} Y_i(t) &= Y_i^*(t) + \epsilon_i(t) \\ &= X_{Li}(t)^T \beta + Z_i(t)^T b_i + \epsilon_i(t) \end{aligned} \tag{1}$$

où $X_{Li}(t)$ est un p -vecteur des covariables (pouvant dépendre du temps) associé au p -vecteur des effets fixes β , et $Z_i(t)$ est un q -vecteur des covariables (pouvant également dépendre du temps) associé au q -vecteur des effets aléatoires gaussiens b_i , de moyenne $\mathbf{0}$ et de matrice de variance-covariance D . Les erreurs $\epsilon_i(t)$ sont supposées indépendantes et de loi gaussienne avec moyenne zéro et variance σ^2 .

2.2 Sous-modèle de survie classique

Pour modéliser le risque d'événement, on utilise un modèle de survie qui prend en compte l'histoire du biomarqueur jusqu'au temps t à travers les effets aléatoires partagés b_i . La fonction de risque instantané prend donc la forme :

$$\begin{aligned}\lambda_i(t|X_{S_i}(t), b_i) &= \lim_{dt \rightarrow 0} \frac{Pr [t \leq T_i^* < t + dt | T_i^* \geq t, X_{S_i}(t), b_i]}{dt} \\ &= \lambda_0(t) \exp\{X_{S_i}^T(t)\gamma + W_i(b_i, t)^T \alpha\}\end{aligned}\quad (2)$$

avec $t > 0$, $\lambda_0(t)$ la fonction de risque de base, $X_{S_i}(t)$ le r-vecteur des covariables explicatives (qui peuvent dépendre du temps) associé à γ le r-vecteur des coefficients qui explique la relation entre les facteurs pronostiques et le temps de survie. $W_i(b_i, t)$ est une fonction multivariée des effets aléatoires b_i issus du sous-modèle longitudinal (1). Le vecteur des coefficients α traduit le lien entre le processus longitudinal et le processus de survie, tandis que $W_i(b_i, t)$ explique la nature de la dépendance entre les deux processus. Différentes structures d'association entre le biomarqueur et le temps de survie peuvent être proposées, comme le niveau courant du biomarqueur ($W_i(b_i, t) = Y_i^*(t)$), la pente courante ($W_i(b_i, t) = \delta Y_i^*(t)/\delta t$), les deux ($W_i(b_i, t) = (Y_i^*(t), \delta Y_i^*(t)/\delta t)^T$), ou bien les déviations individuelles de la trajectoire moyenne du biomarqueur ($W_i(b_i, t) = b_i^T$).

Afin d'approximer au mieux la fonction de risque de base, Rizopoulos (2010) propose de l'approcher par des B-splines :

$$\log \lambda_0(t) = \kappa_0 + \sum_{d=1}^{m+q-1} \kappa_d B_d(t, q)$$

où $\kappa^T = (\kappa_0, \kappa_1, \dots, \kappa_m)$ sont les coefficients des splines, q est le degré des fonctions B-splines $B(\cdot)$, et m est le nombre de nœuds internes.

2.3 Sous-modèle de survie en présence de risques compétitifs

Introduisons quelques notations supplémentaires. On suppose qu'il existe K différents types de rechute. On note $T_{i1}^*, \dots, T_{iK}^*$ les vrais temps de chaque type d'événement pour le sujet i . Soit T_i le temps d'événement observé, et C_i le temps de censure, on a donc $T_i = \min(T_{i1}^*, \dots, T_{iK}^*, C_i)$. Soit $\delta_i \in \{0, 1, \dots, K\}$ l'indicateur d'événement, avec $\delta_i = k$ si l'événement compétitif k est observé, et $\delta_i = 0$ si aucun événement n'est observé, c'est-à-dire en cas de censure.

Pour la cause $k \in (1, \dots, K)$, la fonction de risque instantané cause-spécifique prend la forme :

$$\begin{aligned}\lambda_i^{(k)}(t|X_{S_i}(t), b_i) &= \lim_{dt \rightarrow 0} \frac{Pr [t \leq T_i^* < t + dt, k | T_i^* \geq t, X_{S_i}(t), b_i]}{dt} \\ &= \lambda_0^{(k)}(t) \exp(X_{S_i}^T(t)\gamma^{(k)} + W_i(b_i, t)^T \alpha^{(k)})\end{aligned}\quad (3)$$

3 Estimation

Afin d'estimer les paramètres du modèle conjoint, on utilise l'approche traditionnelle du maximum de vraisemblance. Soit $\theta = (\beta, \sigma, \gamma, \alpha, \kappa, D)$ le vecteur de l'ensemble des

paramètres inclus dans (1) et (3). Le vecteur des effets aléatoires b_i lie les processus longitudinal et de survie, on parle d'effets aléatoires partagés. En conséquence, les deux processus sont indépendants conditionnellement aux effets aléatoires, et la log-vraisemblance des données observées s'écrit :

$$l(\theta) = \sum_{i=1}^N \log \left[\int_{b_i} f_T(T_i, \delta_i | X_{S_i}, b_i; \theta) f_Y(Y_i | X_{L_i}, b_i; \theta) f_b(b_i; \theta) db_i \right] \quad (4)$$

où f_Y est une fonction de densité gaussienne multivariée avec moyenne $X_{L_i}\beta + Z_i b_i$ et matrice de variance-covariance $\Sigma_i = \sigma^2 I_{n_i}$, tandis que f_b est une fonction de densité gaussienne multivariée de moyenne $\mathbf{0}$ et de matrice de variance-covariance D .

Dans le cadre de modèles conjoints standards,

$$f_T(T_i, \delta_i | X_{S_i}, b_i; \theta) = \lambda_i(T_i | X_{S_i}, b_i; \theta)^{\delta_i} S_i(T_i | X_{S_i}, b_i; \theta)$$

avec $\lambda_i(T_i | X_{S_i}, b_i; \theta)$ la fonction de risque instantané définie dans (2) au temps T_i , et $S_i(T_i | X_{S_i}, b_i; \theta) = \exp\left(-\int_0^{T_i} \lambda_i(t | X_{S_i}, b_i; \theta) dt\right)$. Dans le cadre des modèles conjoints à risques compétitifs, les différents types d'événements sont supposés indépendants entre eux si bien que :

$$\begin{aligned} f_T(T_i, \delta_i | X_{S_i}, b_i; \theta) &= \prod_{k=1}^K [f_T(T_i, \delta_i = k | X_{S_i}, b_i; \theta)] \\ &= \prod_{k=1}^K \left\{ \left[\lambda_i^{(k)}(T_i | X_{S_i}, b_i; \theta) \right]^{I(\delta_i=k)} S_i^{(k)}(T_i | X_{S_i}, b_i; \theta) \right\} \end{aligned}$$

où $\lambda_i^{(k)}(T_i | X_{S_i}, b_i)$ est la fonction de risque instantané spécifique à la cause k présentée dans l'équation (3) au temps T_i , et $S_i^{(k)}(T_i | X_{S_i}, b_i; \theta) = \exp\left(-\int_0^{T_i} \lambda_i^{(k)}(t | X_{S_i}, b_i; \theta) dt\right)$ en est la fonction de survie dérivée.

Que l'on se place dans le cas standard ou en présence de risques compétitifs, les modèles conjoints à effets aléatoires partagés peuvent être estimés à l'aide du package R de Rizopoulos JM (2010). La maximisation de la fonction de log-vraisemblance (4) est réalisée à partir de l'algorithme Expectation-Maximisation (EM). De plus, les deux intégrales dans l'équation (4) ne permettent pas d'avoir des solutions analytiques. Ainsi, on approche l'intégrale sur les effets aléatoires à l'aide d'une quadrature Gauss-Hermite, et on approche l'intégrale dans la fonction de survie par une quadrature Gauss-Kronrod.

4 Applications

Nous présentons les résultats des modèles conjoints sur deux bases de données : RTOG 9406, essai clinique multicentrique américain contenant 630 patients inclus entre 1994 et

2012, et une cohorte de Vancouver constituée de 844 patients inclus entre 1994 et 2013. Les patients ont un cancer localisé de la prostate et ont été traités par radiothérapie externe. À travers des modèles conjoints à effets aléatoires partagés, on s'intéresse à l'effet des facteurs pronostiques classiques sur chaque type de rechute clinique (rechute locale, distante, et décès du cancer de la prostate), et à l'effet de la dynamique des PSA (trajectoires du niveau courant et de la pente courante par exemple) sur chaque type de rechute. De plus, les modèles étudiés permettent de faire face à des covariables dépendantes du temps. On étudie alors l'impact de l'introduction d'un nouveau traitement (l'hormonothérapie) sur le risque de chaque type de rechute, et on distingue l'effet de la dynamique des PSA avant et après l'introduction de ce traitement supplémentaire.

Bibliographie

- [1] Diggle, P. J., Sousa, I. et Chetwynd, A. G. (2008), Joint modelling of repeated measurements and time-to-event outcomes : The fourth Armitage lecture, *Statistics in Medicine*, 27(16), 2981–2998.
- [2] Elashoff, R. M., Li, G. et Li, N. (2008), A joint model for longitudinal measurements and survival data in the presence of multiple failure types, *Biometrics*, 64(3), 762–771.
- [3] Proust-Lima, C. et al. (2008), Determinants of change in prostate-specific antigen over time and its association with recurrence after external beam radiation therapy for prostate cancer in five large cohorts, *International Journal of Radiation Oncology Biology Physics*, 72(3), 782–791.
- [4] Proust-Lima, C., Sène, M., Taylor, J. M. et Jacqmin-Gadda, H. (2014), Joint latent class models for longitudinal and time-to-event data : A review, *Statistical methods in medical research*, 23(1), 74–90.
- [5] Rizopoulos, D. (2010), JM : An R package for the joint modelling of longitudinal and time-to-event data, *Journal of Statistical Software*, 35(9), 1–33.
- [6] Sène, M., Bellera, C. et Proust-Lima, C. (2014), Shared random-effect models for the joint analysis of longitudinal and time-to-event data : application to the prediction of prostate cancer recurrence, *Journal de la Société Française de Statistique*, *Accepté*.
- [7] Taylor, J. M., Yu, M. et Sandler, H. M. (2005), Individualized predictions of disease progression following radiation therapy for prostate cancer, *Journal of clinical oncology*, 23(4), 816–825.
- [8] Williamson, P. R., Kolamunnage Dona, R., Philipson, P. et Marson, A. G. (2008), Joint modelling of longitudinal and competing risks data, *Statistics in medicine*, 27(30), 6426–6438.