

Modèles semi-markoviens : Application à l'analyse de l'évolution de pathologies chroniques

Yohann Foucher

Thèse sous la direction de
Jean-Pierre Daurès
et de
Jean-Paul Soulillou

Institut Universitaire de Recherche Clinique
Université Montpellier 1

04 Octobre 2007

Plan

- 1 Introduction
- 2 Définition des états de gravité
- 3 Modélisation multi-états
- 4 Amélioration de la structure multi-états
- 5 Discussion et perspectives

Plan

- 1 Introduction
- 2 Définition des états de gravité
- 3 Modélisation multi-états
- 4 Amélioration de la structure multi-états
- 5 Discussion et perspectives

Problématique clinique (1)

Transplantation rénale

- Evènement d'intérêt : Décès du patient ou retour en dialyse
- Nombre d'articles référencés dans PubMed :
 - Mots-clef= $\{\text{survival, renal, transplantation}\} \longleftrightarrow 14844$
 - Mots-clef= $\{\text{cox, renal, transplantation}\} \longleftrightarrow 622$



Modèle de Cox

Problématique clinique (1)

Transplantation rénale

- Evènement d'intérêt : Décès du patient ou retour en dialyse
- Nombre d'articles référencés dans PubMed :
 - Mots-clef={survival, renal, transplantation} \longleftrightarrow 14844
 - Mots-clef={cox, renal, transplantation} \longleftrightarrow 622



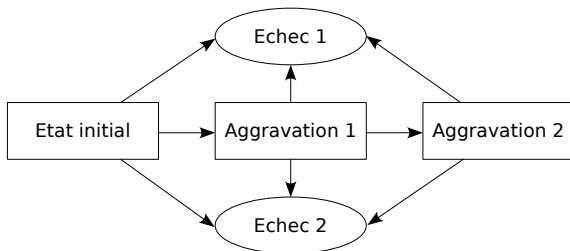
Modèle de Cox

Limites du modèle

- Un seul évènement
- Aucun processus d'aggravation
- Aucun marqueur précoce d'aggravation

Problématique clinique (2)

Apport de la modélisation multi-états [1]



Structure aggravation/échecs

- Plusieurs évènements terminaux
- Prise en compte d'un processus d'aggravation
- Identification de facteurs d'échec précoce
- Identification de facteurs d'aggravation

Problématique clinique (3)

Données disponibles

- Cohorte multicentrique DIVAT (Données Informatisées et Validées en Transplantation) : plus de 7000 greffes depuis 1990.
- Patients suivis au CHU de Nantes : plus de 1000 greffes
- Données recueillies
 - au moment de la greffe : âge et sexe du donneur, nb. incompatibilités, immunisation, âge et sexe du receveur, ischémie froide, comorbités, etc.
 - au cours du suivi : délai de reprise au démarrage, traitement, marqueurs biologiques, rejets aigus, complications, date de retour en dialyse ou de décès, etc.
- Marqueur d'aggravation : clairance de la créatinine (CL)

Problématique clinique (4)

Critères d'inclusion

- Patients greffés à partir de 1996
 - Meilleur renseignement du suivi de la CL
 - Changement des traitements d'induction
- Patients majeurs au moment de la greffe
- Aucune transplantation de pancréas associée
- Centre Hospitalier Universitaire de Nantes

Population d'étude

- 819 patients âgés en moyenne de 48,5 ans
- Donneurs âgés en moyenne de 46,9 ans
- Environ 60% des receveurs et des donneurs sont des hommes

Problématique statistique (1)

Approche markovienne : la plus répandue

- Markov à temps discret
 - inadapté à l'évolution continue
- Markov à temps continu et homogène :
 - Distributions des temps de séjour exponentielles
 - Forces de transition constantes (processus sans mémoire)
 - Hypothèse a priori trop forte dans l'étude du vivant
- Markov à temps continu et non-homogène :
 - Les probabilités de transition dépendent du temps chronologique

Problématique statistique (1)

Approche markovienne : la plus répandue

- Markov à temps discret
 - inadapté à l'évolution continue
- Markov à temps continu et homogène :
 - Distributions des temps de séjour exponentielles
 - Forces de transition constantes (processus sans mémoire)
 - Hypothèse a priori trop forte dans l'étude du vivant
- Markov à temps continu et non-homogène :
 - Les probabilités de transition dépendent du temps chronologique

Intérêt du semi-markov

- L'évolution d'un patient dépend principalement du temps passé dans l'état ^[4]

Problématique statistique (2)

Difficultés méthodologiques liées au semi-markov

- Définition des distributions des temps d'attente
- Introduction des covariables
- Censure par intervalle des transitions
- Prise en compte d'une dépendance des observations
- Test d'adéquation du modèle

Problématique statistique (2)

Difficultés méthodologiques liées au semi-markov

- Définition des distributions des temps d'attente
- Introduction des covariables
- Censure par intervalle des transitions
- Prise en compte d'une dépendance des observations
- Test d'adéquation du modèle

Choix de la structure multi-états

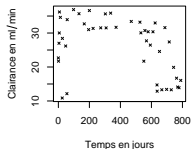
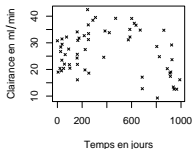
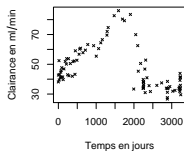
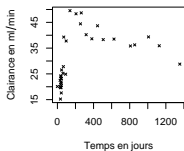
- Définition des états de gravité à partir de la CL
 - Hypothèse de travail à partir d'experts médicaux ?
 - Calcul statistique des seuils de décision ?

Plan

- 1 Introduction
- 2 Définition des états de gravité
- 3 Modélisation multi-états
- 4 Amélioration de la structure multi-états
- 5 Discussion et perspectives

Méthode (1)

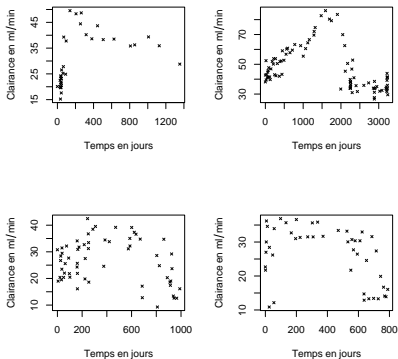
Principe de la classification



Evolution de CL pour les retours en dialyse

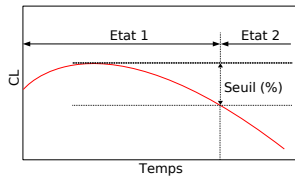
Méthode (1)

Principe de la classification



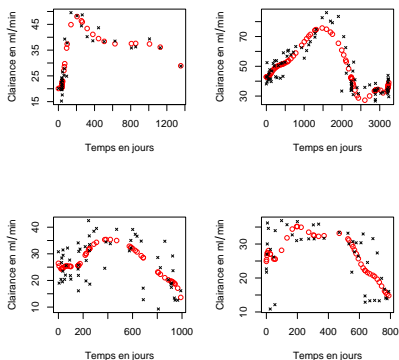
Evolution de CL pour les retours en dialyse

- Détermination de la valeur du seuil : s .



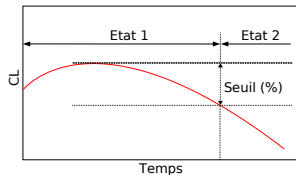
Méthode (1)

Principe de la classification



Evolution de CL pour les retours en dialyse

- Détermination de la valeur du seuil : s .



- Détermination du nombre de noeuds intérieurs (fixes) de la B-spline cubique : d .

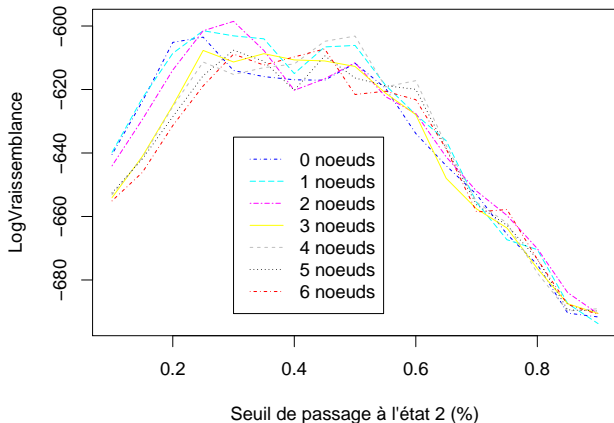
Méthode (2)

Soit un échantillon composé de N patients ($h = 1, \dots, N$). Pour le h ème sujet, on note $y_h = (y_{h,0}, y_{h,1}, \dots, y_{h,n_h})$ les n_h mesures de CL aux temps $\{v_{h,0}, v_{h,1}, \dots, v_{h,n_h}\}$.

Estimation de s et d (1)

- 1 A partir d'une valeur du nombre de noeuds $d \rightarrow$ Lissage des mesures y_h de la CL : $\tilde{y}_h = (\tilde{y}_{h,0}, \tilde{y}_{h,1}, \dots, \tilde{y}_{h,n_h})$
- 2 A partir d'une valeur du seuil $s \rightarrow$ Détermination des états de gravité correspondants à $\tilde{y}_h : z_h = (z_{h,0}, z_{h,1}, \dots, z_{h,n_h})$
- 3 Estimation de la vraisemblance partielle, \mathcal{VP} , du modèle de Cox où la variable z_h est considérée comme temps-dépendante
- 4 Répétition des étapes 1 à 3 pour la grille de valeurs $d = 0, \dots, 10$ et $s = 10\%, \dots, 90\%$
- 5 Le couple (\hat{d}, \hat{s}) est celui qui maximise $\mathcal{VP} : (\hat{d}, \hat{s}) = \text{Argmax}_{(k,s)} \{\mathcal{VP}\}$

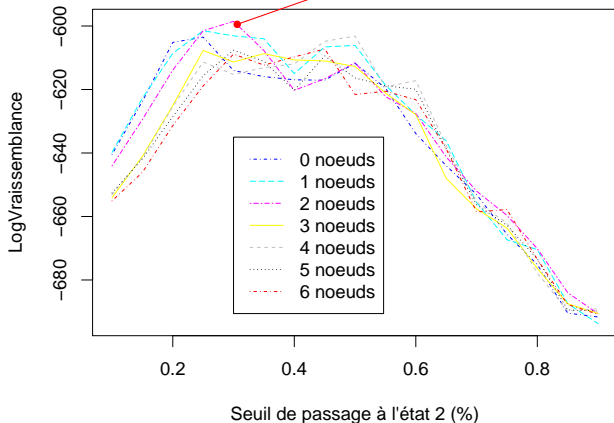
Résultats (1)



Estimation du degré, d , et du seuil de diminution de la clairance de la créatinine, s .

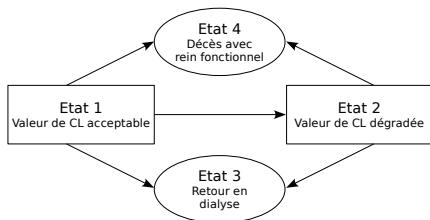
Résultats (1)

$\hat{d} = 2$ et $\hat{s} = 30\%$



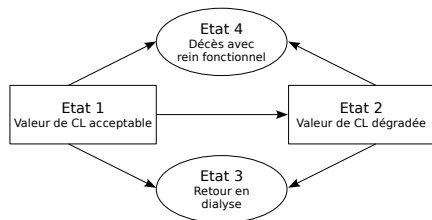
Estimation du degré, d , et du seuil de diminution de la clairance de la créatinine, s .

Résultats (2)



Structure multi-états

Résultats (2)



Structure multi-états

Trajectoires	N	Pourcent.
1	537	64,0 %
1; 2	190	22,7 %
1; 2; 3	61	7,3 %
1; 2; 4	18	2,1 %
1; 3	16	1,9 %
1; 4	17	2,0 %
total	839	100,0 %

Répartition des patients selon leur trajectoire

Plan

- 1 Introduction
- 2 Définition des états de gravité
- 3 Modélisation multi-états**
- 4 Amélioration de la structure multi-états
- 5 Discussion et perspectives

Modèle semi-markovien (1)

Définitions (1)

- Echantillon constitué de N sujets, $h = 1, \dots, N$.
- Le sujet h occupe l'état $W_{h,r}$ au temps $v_{h,r}$, $r = 1, \dots, n_h$.
 - n_h est le nombre de visites pour le sujet h .
 - Tous les patients débutent dans l'état 1, $W_{h,1} = 1$.
 - Si l'évènement final est le retour en dialyse, $W_{h,n_h} = 3$.
 - Si l'évènement final est le décès, $W_{h,n_h} = 4$.
 - Pour toutes les autres visites $\{W_{h,r}, r = 1, \dots, n_h\} \in \{1, 2\}$.
- Soit $\{X_{h,r}, r = 0, \dots, m_h\}$, la séquence d'états observés et distincts.
 - m_h est le nombre de transitions observées chez le sujet h .
- $D_{h,r}$ est la durée passée dans l'état $X_{h,r}$.

Modèle semi-markovien (2)

Définitions (2)

$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{h,0}, D_{h,0}, \dots, X_{r,h} = i)$$

Modèle semi-markovien (2)

Définitions (2)

$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{h,0}, D_{h,0}, \dots, X_{r,h} = i)$$



Propriété semi-markovienne

$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{r,h} = i)$$

Modèle semi-markovien (2)

Définitions (2)

$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{h,0}, D_{h,0}, \dots, X_{r,h} = i)$$



Propriété semi-markovienne

$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{r,h} = i)$$



Théorème de Bayes
 $P(A, B) = P(A|B)P(B)$

$$P(X_{h,r+1} = j | X_{r,h} = i) \times P(D_{h,r} \leq x | X_{h,r+1} = j, X_{r,h} = i)$$

Modèle semi-markovien (2)

Définitions (2)

$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{h,0}, D_{h,0}, \dots, X_{r,h} = i)$$

Propriété semi-markovienne

$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{r,h} = i)$$

Théorème de Bayes
 $P(A, B) = P(A|B)P(B)$

$$P(X_{h,r+1} = j | X_{r,h} = i) \times P(D_{h,r} \leq x | X_{h,r+1} = j, X_{r,h} = i)$$

P_{ij} : Trajectoires

Modèle semi-markovien (2)

Définitions (2)

$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{h,0}, D_{h,0}, \dots, X_{r,h} = i)$$

Propriété semi-markovienne

$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{r,h} = i)$$

Théorème de Bayes
 $P(A, B) = P(A|B)P(B)$

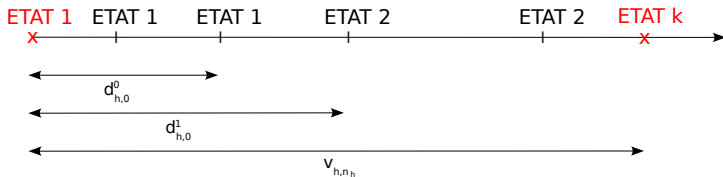
$$P(X_{h,r+1} = j | X_{r,h} = i) \times P(D_{h,r} \leq x | X_{h,r+1} = j, X_{r,h} = i)$$

P_{ij} : Trajectoires

$F_{ij}(x)$: Temps d'attente

Modèle semi-markovien (3)

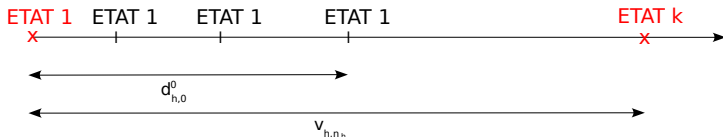
Construction de vraisemblance (1)



$$C_{h,1} = P_{12}P_{2k} \int_{d_{h,0}^0}^{d_{h,0}^1} f_{12}(u)f_{2k}(v_{h,n_h} - u)du$$

Modèle semi-markovien (4)

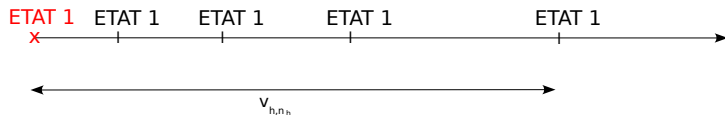
Construction de vraisemblance (2)



$$C_{h,2} = P_{1k} f_{1k}(v_{h,n_h}) + P_{12} P_{2k} \int_{d_{h,0}^0}^{v_{h,n_h}} f_{12}(u) f_{2k}(v_{h,n_h} - u) du$$

Modèle semi-markovien (5)

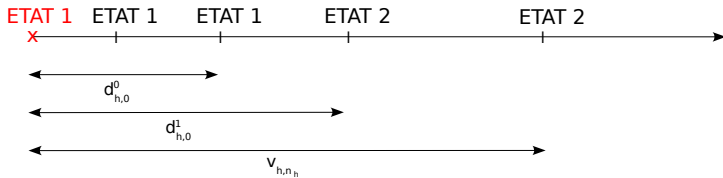
Construction de vraisemblance (3)



$$C_{h,3} = \sum_{j \neq 1} P_{1j} S_{1j}(v_{h,n_h})$$

Modèle semi-markovien (6)

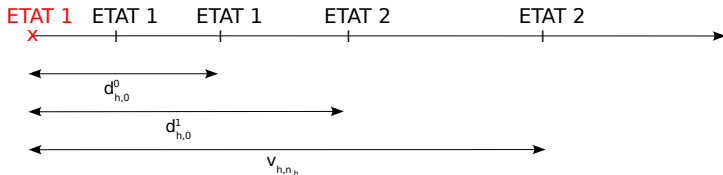
Construction de vraisemblance (4)



$$C_{h,4} = P_{12} \int_{d_{h,0}^0}^{d_{h,0}^1} f_{12}(u) \left\{ \sum_{j=3}^4 P_{2j} S_{2j}(v_{h,n_h} - u) \right\} du$$

Modèle semi-markovien (6)

Construction de vraisemblance (4)



$$C_{h,4} = P_{12} \int_{d_{h,0}^0}^{d_{h,0}^1} f_{12}(u) \left\{ \sum_{j=3}^4 P_{2j} S_{2j}(v_{h,n_h} - u) \right\} du$$

$$\mathcal{V} = \prod_{h=1}^N \prod_{i=1}^4 (C_{h,i})^{\delta_{h,i}}$$

Modèle semi-markovien (7)

Modélisation des facteurs liés à la durée dans l'état i avant de transiter vers l'état j

- Hypothèse de proportionnalité des risques (PH).
 - $\lambda_{ij}(x, \eta_{h,ij}(x)) = \lambda_{0,ij}(x) \exp(\beta_{ij} z_{h,ij})$
- Si hypothèse PH non-respectée ($\log(-\log(S_{ij}(t)))$ dans chaque groupe) :
 - $\beta_{ij}(x) = \beta_{ij}^{(1)} + \beta_{ij}^{(2)} x + \beta_{ij}^{(3)} x^2$

Modélisation des facteurs liés à la trajectoire $i \rightarrow j$

- Principe de régression logistique multinomiale
- $P_{ij}(y_{h,ik}, k \neq i) = \exp(\gamma_{ij} + \beta_{ij} y_{h,ij}) / \sum_{k \neq i} \exp(\gamma_{ik} + \beta_{ik} y_{h,ik})$
 - $\sum_{j \neq i} P_{ij}(y_{h,ik}, k \neq i) = 1$
 - Transitions de référence : $\gamma_{14} = \beta_{14} = 0$ et $\gamma_{24} = \beta_{24} = 0$.

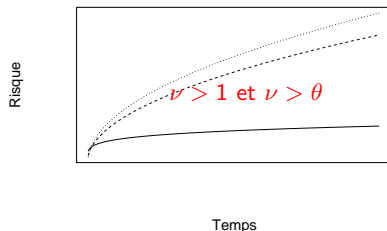
Modèle semi-markovien (8)

Modélisation de la fonction de risque de base

- Choix d'une approche paramétrique
- Utilisation d'une distribution de Weibull généralisée [5] :

- Le risque est non-monotone au cours du temps

- $\lambda_{ij}(t) = \frac{1}{\theta_{ij}} \left(1 + \left(\frac{t}{\sigma_{ij}}\right)^{\nu_{ij}}\right)^{\frac{1}{\theta_{ij}} - 1} \frac{\nu_{ij}}{\sigma_{ij}} \left(\frac{t}{\sigma_{ij}}\right)^{\nu_{ij} - 1} \quad \forall \nu_{ij}, \sigma_{ij}, \theta_{ij} > 0$



Fonctions de risque ajustables par une loi $WG(\sigma, \nu, \theta)$

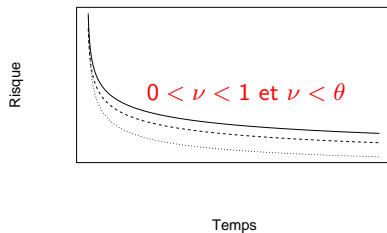
Modèle semi-markovien (8)

Modélisation de la fonction de risque de base

- Choix d'une approche paramétrique
- Utilisation d'une distribution de Weibull généralisée [5] :

- Le risque est non-monotone au cours du temps

- $\lambda_{ij}(t) = \frac{1}{\theta_{ij}} \left(1 + \left(\frac{t}{\sigma_{ij}}\right)^{\nu_{ij}}\right)^{\frac{1}{\theta_{ij}} - 1} \frac{\nu_{ij}}{\sigma_{ij}} \left(\frac{t}{\sigma_{ij}}\right)^{\nu_{ij} - 1} \quad \forall \nu_{ij}, \sigma_{ij}, \theta_{ij} > 0$



Fonctions de risque ajustables par une loi $WG(\sigma, \nu, \theta)$

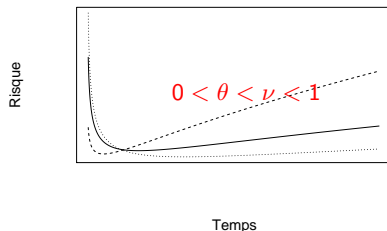
Modèle semi-markovien (8)

Modélisation de la fonction de risque de base

- Choix d'une approche paramétrique
- Utilisation d'une distribution de Weibull généralisée [5] :

- Le risque est non-monotone au cours du temps

$$\lambda_{ij}(t) = \frac{1}{\theta_{ij}} \left(1 + \left(\frac{t}{\sigma_{ij}}\right)^{\nu_{ij}}\right)^{\frac{1}{\theta_{ij}} - 1} \frac{\nu_{ij}}{\sigma_{ij}} \left(\frac{t}{\sigma_{ij}}\right)^{\nu_{ij} - 1} \quad \forall \nu_{ij}, \sigma_{ij}, \theta_{ij} > 0$$



Fonctions de risque ajustables par une loi $WG(\sigma, \nu, \theta)$

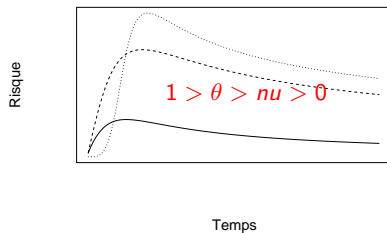
Modèle semi-markovien (8)

Modélisation de la fonction de risque de base

- Choix d'une approche paramétrique
- Utilisation d'une distribution de Weibull généralisée [5] :

- Le risque est non-monotone au cours du temps

$$\lambda_{ij}(t) = \frac{1}{\theta_{ij}} \left(1 + \left(\frac{t}{\sigma_{ij}}\right)^{\nu_{ij}}\right)^{\frac{1}{\theta_{ij}} - 1} \frac{\nu_{ij}}{\sigma_{ij}} \left(\frac{t}{\sigma_{ij}}\right)^{\nu_{ij} - 1} \quad \forall \nu_{ij}, \sigma_{ij}, \theta_{ij} > 0$$



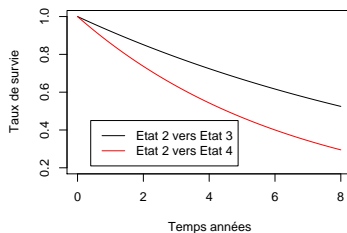
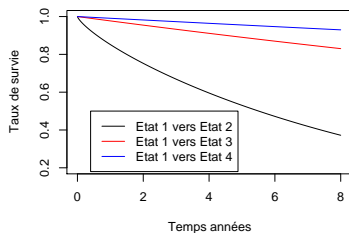
Fonctions de risque ajustables par une loi $WG(\sigma, \nu, \theta)$

Résultats (1)

Distributions de base des durées dans les états

Transition	σ_{ij}		ν_{ij}		θ_{ij}	
	Estim.	ET	Estim.	ET	Estim.	ET
1 \rightarrow 2	68,84	82,00	0,77	0,05	0,25	0,21
1 \rightarrow 3	42,91	47,44	1	.	1	.
1 \rightarrow 4	109,80	73,31	1	.	1	.
2 \rightarrow 3	12,96	3,34	1	.	1	.
2 \rightarrow 4	6,30	3,93	1	.	1	.

Paramètres des fonctions de risque de base



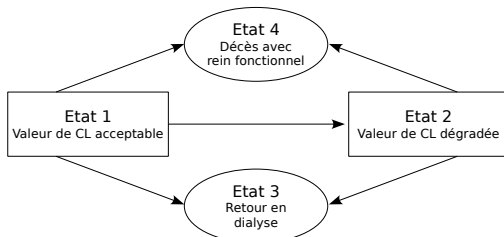
Fonctions de survie de base des temps d'attente

Résultats (2)

Covariables associées aux trajectoires (P_{ij})

Transition	Variable	Coef.	ET	p-value
1 → 2	Intercept	1,88	0,68	0,0063
1 → 2	Délai de reprise	0,73	0,34	0,0306
1 → 2	Age du donneur	-2,04	0,70	0,0034
1 → 3	Intercept	-2,81	0,53	0,0001
2 → 3	Intercept	1,14	0,34	0,0007
2 → 3	Incompatibilités A+B+DR	0,93	0,47	0,0483

Paramètres de la chaîne de Markov



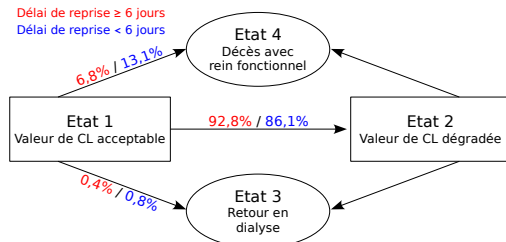
Probabilités relatives aux séquences d'états

Résultats (2)

Covariables associées aux trajectoires (P_{ij})

Transition	Variable	Coef.	ET	p-value
1 → 2	Intercept	1,88	0,68	0,0063
1 → 2	Délai de reprise	0,73	0,34	0,0306
1 → 2	Age du donneur	-2,04	0,70	0,0034
1 → 3	Intercept	-2,81	0,53	0,0001
2 → 3	Intercept	1,14	0,34	0,0007
2 → 3	Incompatibilités A+B+DR	0,93	0,47	0,0483

Paramètres de la chaîne de Markov



Probabilités relatives aux séquences d'états

Résultats (3)

Covariables associées aux temps d'attente ($S_{ij}(\cdot)$)

Transition	Variable	Coef.	ET	exp(Coef.)	p-value
1 → 2	Traitement d'induction	0,36	0,14	1,43	0,0063
1 → 2	Sexe du receveur	-0,26	0,13	0,77	0,0541
1 → 2	Age du donneur	0,96	0,23	2,61	0,0001
1 → 3	Ischémie froide	5,02	1,20	151,41	0,0001
2 → 3	Incompatibilité A+B+DR	0,90	0,29	2,46	0,0017
2 → 3	PRA	1,10	0,35	3,00	0,0016
2 → 3	PRA × durée	-0,47	0,22	0,63	0,0309
2 → 4	Délai de reprise	2,01	0,60	7,46	0,0008
2 → 4	Sexe du receveur	1,52	0,64	4,57	0,0174
2 → 4	Sexe du receveur × durée	-4,31	1,19	0,01	0,0003
2 → 4	Sexe du receveur × durée ²	1,30	0,32	3,67	0,0001

Effets des covariables sur les vitesses de transition

Résultats (3)

Covariables associées aux temps d'attente ($S_{ij}(\cdot)$)

Transition	Variable	Coef.	ET	exp(Coef.)	p-value
1 → 2	Traitement d'induction	0,36	0,14	1,43	0,0063
1 → 2	Sexe du receveur	-0,26	0,13	0,77	0,0541
1 → 2	Age du donneur	0,96	0,23	2,61	0,0001
1 → 3	Ischémie froide	5,02	1,20	151,41	0,0001
2 → 3	Incompatibilité A+B+DR	0,90	0,29	2,46	0,0017
2 → 3	PRA	1,10	0,35	3,00	0,0016
2 → 3	PRA × durée	-0,47	0,22	0,63	0,0309
2 → 4	Délai de reprise	2,01	0,60	7,46	0,0008
2 → 4	Sexe du receveur	1,52	0,64	4,57	0,0174
2 → 4	Sexe du receveur × durée	-4,31	1,19	0,01	0,0003
2 → 4	Sexe du receveur × durée ²	1,30	0,32	3,67	0,0001

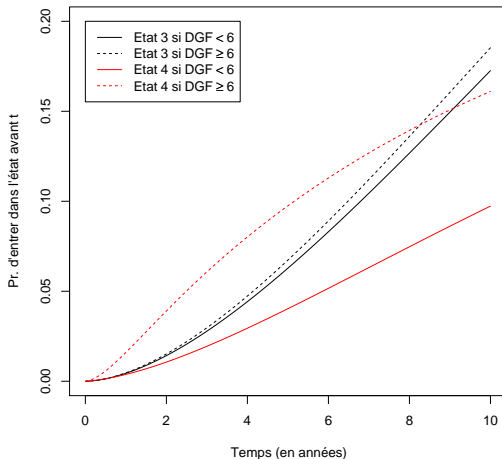
Effets des covariables sur les vitesses de transition

Calcul de la proportion attendue d'un évènement terminal k au temps t ($Q_k(t)$)

$$Q_k(t) = \int_0^t \left\{ P_{1k} f_{1k}(x) + P_{12} P_{2k} \int_0^x f_{12}(u) f_{13}(x-u) du \right\} dx$$

Densité de probabilité d'observer l'état k au temps x

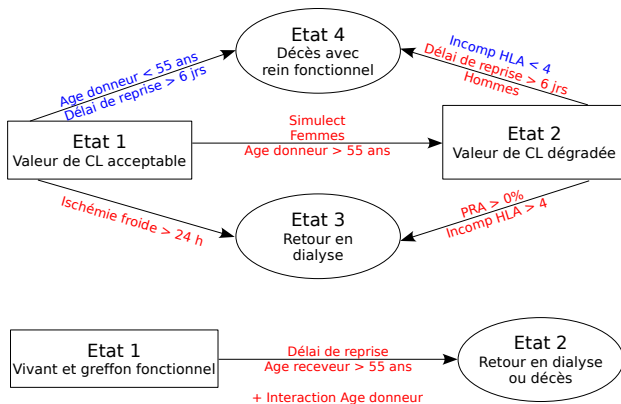
Résultats (4)



$Q_k(t)$ en fonction du nombre d'incompatibilités ($k = 3, 4$)

Résultats (5)

Comparaison des résultats avec ceux d'un modèle de Cox

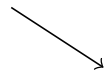


Hypothèse de stationnarité

$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{h,0}, D_{h,0}, \dots, X_{r,h} = i)$$



$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{r,h} = i)$$



$$P(X_{h,r+1} = j | X_{r,h} = i) \times P(D_{h,r} \leq x | X_{h,r+1} = j, X_{r,h} = i)$$

P_{ij} : Trajectoires

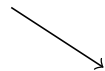
$F_{ij}(x)$: Temps d'attente

Hypothèse de stationnarité

$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{h,0}, D_{h,0}, \dots, X_{r,h} = i)$$



$$P(D_{h,r} \leq x, X_{h,r+1} = j | X_{r,h} = i)$$



$$P(X_{h,r+1} = j | X_{r,h} = i) \times P(D_{h,r} \leq x | X_{h,r+1} = j, X_{r,h} = i)$$



P_{ij} : Trajectoires

$F_{ij}(x)$: Temps d'attente



Quantités indépendantes du temps chronologique

Hypothèse de stationnarité (1)

Définition de la statistique

- Tableau de contingence composé de $L \times K$ cellules
 - L intervalles de temps depuis l'origine ($l = 1, 2, \dots, L$)
 - K types de transition ($k = 1, 2, \dots, K$)
- Statistique d'adéquation de type Pearson

$$G = \sum_{l=1}^L \sum_{k=1}^K \left\{ (n_{l,k} - e_{l,k})^2 / e_{l,k} \right\}$$

- $e_{l,k}$ est le nombre de transitions attendues dans la cellule (l, k)
- $n_{l,k}$ est le nombre de transitions observées dans la cellule (l, k)
- Seuls les temps de transition vers un état terminal sont exactement renseignés

Hypothèse de stationnarité (2)

Calcul de l'effectif attendu dans la cellule (l, k)

- La probabilité qu'un individu h transite vers l'état k ($k = 3, 4$) au temps t depuis la greffe est égale à :

$$P_{1k}f_{1k}(t) + P_{12}P_{2k} \int_0^t f_{12}(u)f_{2k}(t-u)du$$

Hypothèse de stationnarité (2)

Calcul de l'effectif attendu dans la cellule (l, k)

- La probabilité qu'un individu h transite vers l'état k ($k = 3, 4$) au temps t depuis la greffe est égale à :

$$P_{1k} f_{1k}(t) + P_{12} P_{2k} \int_0^t f_{12}(u) f_{2k}(t-u) du$$

- Par intégration, la probabilité que cet évènement k se produise entre t_{l-1} et t_l ($l = 1, \dots, L$) est égale à :

$$P_{1k} \int_{t_{l-1}}^{t_l} f_{1k}(t) dt + P_{12} P_{2k} \int_{t_{l-1}}^{t_l} \int_0^t f_{12}(u) f_{2k}(t-u) dudt$$

Hypothèse de stationnarité (2)

Calcul de l'effectif attendu dans la cellule (l, k)

- La probabilité qu'un individu h transite vers l'état k ($k = 3, 4$) au temps t depuis la greffe est égale à :

$$P_{1k} f_{1k}(t) + P_{12} P_{2k} \int_0^t f_{12}(u) f_{2k}(t-u) du$$

- Par intégration, la probabilité que cet évènement k se produise entre t_{l-1} et t_l ($l = 1, \dots, L$) est égale à :

$$P_{1k} \int_{t_{l-1}}^{t_l} f_{1k}(t) dt + P_{12} P_{2k} \int_{t_{l-1}}^{t_l} \int_0^t f_{12}(u) f_{2k}(t-u) dudt$$

- Avec la censure, l'effectif attendu s'écrit :

$$e_{l,k} = \left\{ \sum_{R(t_{l-1})} P_{1k} \int_{t_{l-1}}^{\min(v_h, n_h, t_l)} f_{1k}(t) dt + P_{12} P_{2k} \int_{t_{l-1}}^{\min(v_h, n_h, t_l)} \int_0^t f_{12}(u) f_{2k}(t-u) dudt \right\}$$

Hypothèse de stationnarité (3)

Calcul de la statistique

- 5 intervalles de temps définis par les quantiles : $L = 5$
- 2 évènements étudiés (transitions exactement observées) : $k = 3, 4$.

Temps (en années)		Transitions		Contributions	
		$e \rightarrow 3$	$e \rightarrow 4$	$e \rightarrow 3$	$e \rightarrow 4$
[0, 011; 0, 689]	Observé	12	8	5,19%	10,12%
	Attendu	9,38	5,25		
]0, 689; 2, 168]	Observé	13	8	21,21%	2,89%
	Attendu	20,91	10,02		
]2, 168; 3, 826]	Observé	16	5	15,73%	11,73%
	Attendu	23,17	8,81		
]3, 826; 5, 213]	Observé	17	4	2,14%	4,45%
	Attendu	14,87	6,18		
]5, 213; 9, 158]	Observé	14	7	25,31%	0,24%
	Attendu	23,08	7,51		

- La statistique de test $G = 14,12$
- Problème : Obtenir la distribution de G sous l'hypothèse nulle de stationnarité (H_0)

Hypothèse de stationnarité (4)

Bootstrap semi-paramétrique [3]

- 1 Génération d'un échantillon de Bootstrap de taille N . Chaque individu h^* de cet échantillon est observé aux temps $\{v_{h^*,0}^*, v_{h^*,1}^*, \dots, v_{h^*,n_{h^*}}^*\}$.
- 2 Simulation des séquences d'états à partir des $P_{ij}(y_{h^*,ik}^*, k \neq i)$
 - lois multinomiales ou binomiales
- 3 Simulation des durées dans les états en fonction des séquences d'états et des $F_{ij}(x, \eta_{h,ij}(x))$
 - Génération d'une variable aléatoire uniforme U
 - Calcul de la fonction de répartition inverse de U
- 4 Construction des contributions à la vraisemblance en fonction des trajectoires (étapes 2 et 3) et des temps d'observation (étape 1)
- 5 Estimation du modèle semi-markovien
- 6 Calcul de la statistique G^*

Hypothèse de stationnarité (5)

Calcul de la p-value

- Répétition B fois des étapes 1 à 6 $\Rightarrow G_b^*$ ($b = 1, \dots, B$)
- La p-value du test

$$\begin{cases} H_0 : \text{Le modèle est stationnaire} \\ H_1 : \text{Le modèle n'est pas stationnaire} \end{cases}$$

est alors égale à :

$$p = B^{-1} \sum_{b=1}^B \mathbb{1}_{\{G_b^* \geq G\}}$$

Hypothèse de stationnarité (5)

Calcul de la p-value

- Répétition B fois des étapes 1 à 6 $\Rightarrow G_b^*$ ($b = 1, \dots, B$)
- La p-value du test

$$\begin{cases} H_0 : \text{Le modèle est stationnaire} \\ H_1 : \text{Le modèle n'est pas stationnaire} \end{cases}$$

est alors égale à :

$$p = B^{-1} \sum_{b=1}^B \mathbb{1}_{\{G_b^* \geq G\}}$$

Application

- $B = 400$ échantillons de Bootstrap (temps de calcul)
- L'hypothèse de stationnarité n'est pas rejetée ($p = 0,3975$)

Plan

- 1 Introduction
- 2 Définition des états de gravité
- 3 Modélisation multi-états
- 4 Amélioration de la structure multi-états**
- 5 Discussion et perspectives

Amélioration de la structure multi-états (1)

Rappels

- Mesures répétées d'un marqueur pronostique.
- Calcul d'un seuil qui discrimine les états de bon et de mauvais pronostic
- Problèmes de la méthode de maximisation de la vraisemblance partielle d'un modèle de Cox :
 - La vraisemblance n'est pas dérivable au seuil → utilisation d'une grille de valeurs
 - Que représente la maximisation de la vraisemblance en pratique ?
 - Nécessité de distinguer le pouvoir pronostique des échecs par rapport au pouvoir pronostique des succès
- **Théorie des tests diagnostiques (courbes ROC)**

Amélioration de la structure multi-états (2)

Quelques définitions

- Soit X_d la variable aléatoire de pronostic mesurée au temps d
- Soit T le temps de survie avant l'échec (décès ou retour en dialyse)
- Un patient est à risque au temps d si le test $\{X_d > c\}$ est positif [2]
- On note $n_{FN}(c, t, d)$ et $n_{FP}(c, t, d)$ les effectifs de faux négatifs et de faux positifs issus du test réalisé au temps d pour un pronostic au temps t ($t > d$)
- Soient C_{FN} et C_{FP} les coûts respectivement associés à un faux négatif et à un faux positif

Amélioration de la structure multi-états (3)

Fonction de coût

- Fonction de coût relative à la décision issue du test $\{X_d > c\}$:

$$C(c, t, d) = C_{FP}n_{FP}(c, t, d) + C_{FN}n_{FN}(c, t, d)$$

- Si $k = C_{FN}/C_{FP}$:

$$C(c, t, d) \propto n_{FP}(c, t, d) + k \times n_{FN}(c, t, d)$$

- On montre que :

$$\begin{aligned} C(c, t, d) &\propto \int_c^\infty \exp(-\Lambda(t|x_d)) f_{X|d}(x_d) dx_d \\ &+ k \times \int_{-\infty}^c [1 - \exp(-\Lambda(t|x_d))] f_{X|d}(x_d) dx_d \end{aligned}$$

Amélioration de la structure multi-états (3)

Fonction de coût

- Fonction de coût relative à la décision issue du test $\{X_d > c\}$:

$$C(c, t, d) = C_{FP}n_{FP}(c, t, d) + C_{FN}n_{FN}(c, t, d)$$

- Si $k = C_{FN}/C_{FP}$:

$$C(c, t, d) \propto n_{FP}(c, t, d) + k \times n_{FN}(c, t, d)$$

- On montre que :

$$\begin{aligned} C(c, t, d) &\propto \int_c^\infty \exp(-\Lambda(t|x_d)) f_{X|d}(x_d) dx_d \\ &+ k \times \int_{-\infty}^c [1 - \exp(-\Lambda(t|x_d))] f_{X|d}(x_d) dx_d \end{aligned}$$

Amélioration de la structure multi-états (3)

Fonction de coût

- Fonction de coût relative à la décision issue du test $\{X_d > c\}$:

$$C(c, t, d) = C_{FP}n_{FP}(c, t, d) + C_{FN}n_{FN}(c, t, d)$$

- Si $k = C_{FN}/C_{FP}$:

$$C(c, t, d) \propto n_{FP}(c, t, d) + k \times n_{FN}(c, t, d)$$

- On montre que :

$$\begin{aligned} C(c, t, d) &\propto \int_c^\infty \exp(-\Lambda(t|x_d)) f_{X|d}(x_d) dx_d \\ &+ k \times \int_{-\infty}^c [1 - \exp(-\Lambda(t|x_d))] f_{X|d}(x_d) dx_d \end{aligned}$$

Amélioration de la structure multi-états (4)

Définition du seuil de décision optimal

- Soit \tilde{c} la valeur du seuil qui minimise $C(c, t, d)$:

$$\left. \frac{\partial C(c, t, d)}{\partial c} \right|_{\tilde{c}} = 0$$

- \tilde{c} satisfait l'équation suivante :

$$\Lambda(t|\tilde{c}) = \log((1 + k))$$

Amélioration de la structure multi-états (5)

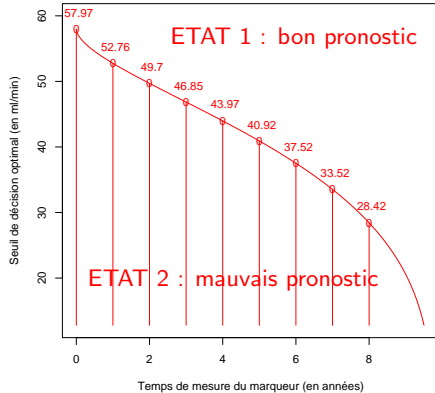
Application

- Hypothèse : prédire un bon pronostic à tort est plus grave que prédire un échec à tort

$$k = C_{FN}/C_{FP} = 10$$

- Pronostic d'un événement avant le 10^{ième} anniversaire de la greffe

$$t = 10 \text{ ans}$$



Seuils optimaux de décision ($k = 0, 1$ et $t = 10$ ans)

Plan

- 1 Introduction
- 2 Définition des états de gravité
- 3 Modélisation multi-états
- 4 Amélioration de la structure multi-états
- 5 Discussion et perspectives**

Discussion et perspectives (1)

Synthèse des développements

- Construction structure aggravation/échecs
 - Définition des états d'aggravation à partir d'un marqueur précoce de pronostic
 - Méthode basée sur le maximum de vraisemblance
 - Développement de la théorie des tests pronostiques
- Réflexions autour de l'introduction de facteurs explicatifs
- Utilisation de distributions générales des temps de survie
- Proposition d'un test d'adéquation de la stationnarité
- Introduction d'une structure de dépendance des observations

Discussion et perspectives (2)

Perspectives

- Prise en compte de retours entre états transitoires :
 - Processus de “guérison” du patient ?
 - Difficulté liée à la censure par intervalle
- Prise en compte d'une transition de l'état dialyse à l'état décès
- Nécessité d'approfondir la théorie et l'application de la théorie des tests pronostiques
 - Variables d'ajustement
 - Construction d'un score de pronostic temps-dépendant
- Reprendre l'estimation du modèle semi-Markovien à partir des nouveaux états de gravité
- Travail d'illustration et d'interprétation des résultats d'un point de vue clinique

Références

- [1] Commenges D.
Multi-state models in epidemiology.
Lifetime data analysis, 5:309–321, 1999.
- [2] Heagerty PJ, Lumley T, and Pepe SP.
Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker.
Biometrics, 56:337–344, 2000.
- [3] Aguirre-Hernandez R and Farewell VT.
A Pearson-type goodness-of-fit test for stationary and time-continuous Markov regression models.
Statistics in Medicine, 21:1899–1911, 2002.
- [4] Perez-Ocon R and Ruiz-Castro JE.
Semi-Markov Models and Applications, chapter 14, pages 229–238.
Kluwer Academic Publishers, 1999.
- [5] Bagdonavicius V and Nikulin M.
Accelerated Life Models.
Chapman and Hall/CRC, 2002.