A SEMI-MARKOV MODEL FOR INTERVAL-CENSORED DATA
ANALYSIS OF THE EVOLUTION OF KIDNEY TRANSPLANT RECIPIENTS

Y. Foucher, M. Giral, JP. Soulillou, JP. Daures

yohann.foucher@iurc.montp.inserm.fr
University of Montpellier 1, France

ISCB - Geneva 2006
Outline

Introduction

Semi-Markov model

Definitions
Loglikelihood
Modelling assumptions

Kidney transplant recipients

Data and model
Results

Concluding remarks
Introduction

- Multistate approaches are becoming increasingly used for the analysis of longitudinal data.
- Semi-Markov models explicitly define distributions of waiting times.
- In the follow-up of patients, transition times are known to have occurred in some interval.
- **Objective**: The development of a flexible semi-Markov model which allow for interval censoring.
Definitions (1)

- **Probability of jumping from the State** $i$ **to the State** $j$.

- **Staying times** $T_{n+1} - T_n \sim F_{ij}(T_{n+1} - T_n)$.
Definitions (2)

Embedded Markov chain

\[ P_{ij} = P(X_{n+1} = j|X_n = i) \]

- If state \( i \) is not persistent then \( P_{ij} \geq 0 \) and \( P_{ii} = 0 \).
- If state \( i \) is persistent then \( P_{ij} = 0 \) and \( P_{ii} = 1 \).
Definitions (2)

Embedded Markov chain

\[ P_{ij} = P(X_{n+1} = j | X_n = i) \]

- If state \( i \) is not persistent then \( P_{ij} \geq 0 \) and \( P_{ii} = 0 \).
- If state \( i \) is persistent then \( P_{ij} = 0 \) and \( P_{ii} = 1 \).

Distribution of waiting times

- \( F_{ij}(x) = P(T_{n+1} - T_n \leq x | X_{n+1} = j, X_n = i) \).
- \( F_{ij}(x) = F^{(ij)}(x, \varphi_{ij}) \).

\[ \implies S_{ij}(x), f_{ij}(x) \text{ et } \lambda_{ij}(x) \]
Loglikelihood (1)

**Contribution of a transition exactly observed** $δ_{E,h,r}^{E}$

Let $d_{h,r} = T_{h,r+1} - T_{h,r}$, the waiting time in the state $X_{h,r}$ before jumping to the state $X_{h,r+1}$.

$$\lim_{\Delta d \to 0^+} \frac{P(d_{h,r} < x < d_{h,r} + \Delta d, X_{h,r+1} = j | X_{h,r} = i)}{\Delta d}$$

$$= \lim_{\Delta d \to 0^+} \frac{P(d_{h,r} < x < d_{h,r} + \Delta d | X_{h,r+1} = j, X_{h,r} = i)}{\Delta d} \times P(X_{h,r+1} = j | X_{h,r} = i) = P_{ij}f_{ij}(d_{h,r})$$
Loglikelihood (2)

Contribution of a right-censored transition $\delta_{h,r}^R$

Let $d_{h,r}^0$ be a value such that if $d_{h,r}^0 < d_{h,r}$ then $d_{h,r}^0$ is observed and $d_{h,r}$ is not.

\[
P(d_{h,r} > d_{h,r}^0 | X_{h,r} = i) = \sum_{j \neq i} P(X_{h,r+1} = j | X_{h,r} = i) P(d_{h,r} > d_{h,r}^0 | X_{h,r+1} = j, X_{h,r} = i)
\]

\[
= \sum_{j \neq i} P_{ij} \int_{d_{h,r}^0}^{\infty} f_{ij}(u) du = \sum_{j \neq i} P_{ij} S_{ij}(d_{h,r}^0)
\]
Loglikelihood (3)

Contribution of a interval-censored transition $\delta_{h,r}^i$

Let $d_{h,r}^i$ be a value such that if $d_{h,r}^i > d_{h,r}$ then $d_{h,r}^i$ is observed and $d_{h,r}$ is not.

$$P(d_{h,r}^0 < x < d_{h,r}^1, X_{h,r+1} = j | X_{h,r} = i)$$

$$= P(X_{h,r+1} = i | X_{h,r} = i) \int_{d_{h,r}^0}^{d_{h,r}^1} f_{ij}(u)du$$

$$= P_{ij}(\int_0^{d_{h,r}^1} f_{ij}(u)du - \int_0^{d_{h,r}^0} f_{ij}(u)du)$$

$$= P_{ij}(F_{ij}(d_{h,r}^1) - F_{ij}(d_{h,r}^0)) = P_{ij}(S_{ij}(d_{h,r}^0) - S_{ij}(d_{h,r}^1))$$
Loglikelihood (4)

Contribution of an initial observation for the subject $h$

By defining $z_{h,0j}$, the vector of covariates associated with the initial state $j$ for the $h^{th}$ subject, the usual multinomial logistic regression can be written as:

$$P(X_{h,1} = j) = \frac{\exp(\gamma_{0j} + \beta_{0j}z_{h,0j})}{\sum_{k=1}^{c} \exp(\gamma_{0k} + \beta_{0k}z_{h,0k})} \quad \text{for } j = 1, \ldots, c$$

with $\gamma_{0c} = 0$ and $\beta_{0c} = 0$, in order to obtain $\sum_{j=1}^{c} \pi_{0j} = 1$. 
Loglikelihood (5)

\[
\ln L = \sum_h \left\{ \gamma_0 x_{h,1} + \beta_0 x_{h,1} z_{h,0} x_{h,1} - \ln \left( \sum_{i=1}^c \exp(\gamma_{0i} + \beta_{0i} z_{h,0} x_{h,1}) \right) \right\}
\]

\[
+ \sum_{ij} \sum_{X_{h,r}=i, X_{h,r+1}=j} \left\{ \delta^E_{h,r} \left[ \ln P_{ij} + \ln S_{ij}(d_{h,r}) + \ln \lambda_{ij}(d_{h,r}) \right] \right\}
\]

\[
+ \delta^I_{h,r} \left[ \ln P_{ij} + \ln(S_{ij}(d^0_{h,r}) - S_{ij}(d^1_{h,r})) \right] \right\}
\]

\[
+ \sum_i \sum_{X_{h,r}=i} \left\{ \delta^R_{h,r} \left[ \ln(\sum_{j \neq i} P_{ij} S_{ij}(d^0_{h,r})) \right] \right\}
\]

where \( \gamma_{0i} = \beta_{0i} = 0 \).
Modelling assumptions (1)

Generalised Weibull distribution \((\nu_{ij}, \sigma_{ij}, \theta_{ij} > 0)\)

- Hazard, \(\lambda_{ij}(x) = \frac{1}{\theta_{ij}} \left( 1 + \left( \frac{x}{\sigma_{ij}} \right)^{\nu_{ij}} \right)^{-\frac{1}{\theta_{ij}}} \left( \frac{x}{\sigma_{ij}} \right)^{\nu_{ij}-1}\)

- Survival, \(S_{ij}(x) = \exp \left( 1 - \left( 1 + \left( \frac{x}{\sigma_{ij}} \right)^{\nu_{ij}} \right)^{-\frac{1}{\theta_{ij}}} \right)\)

![Graphs showing hazard and survival functions for different values of \(\nu\) and \(\sigma\).](image-url)
Modelling assumptions (2)

Incorporation of covariates (PH)

- Proportional Hazard (PH) assumption.
  \[ S_{ij}(x, \eta_{h,ij}) = S_{0,ij}(x)^{\exp(\eta_{h,ij})} \]
  \[ \lambda_{ij}(x, \eta_{h,ij}) = \lambda_{0,ij}(x)^{\exp(\eta_{h,ij})} \]

- Respect of the PH assumption.
  Plotting \( \log(-\log(S_{ij}(x))) \) against the survival time \( x \).
Kidney transplant recipients (1)

Data description

- Prospective study of kidney transplant recipients (DIVAT).
- 997 patients and 1980 exact or censored transitions.
- Data were computerized at each checkup visit.
- 5 explanatory variables have been retained:
  - gender (men = 1; women = 0),
  - cold ischemia time (1 if $\geq 16$ hours and 0 otherwise),
  - year of the transplantation (1 if $< 1998$ and 0 otherwise),
  - recipient age at the time of transplantation (1 if $\geq 55$ years of age and 0 otherwise),
  - delayed graft function (1 if $\geq 6$ days and 0 otherwise).
Kidney transplant recipients (2)

Multistate structure

- 3-gravity states with two markers:
  - Creatinine clearance (CL)
  - Proteinuria (PR)


![Diagram showing the multistate structure with states and transitions]

- State 1: PR < 0.5 gr/day
- State 2: CL -20% in 1 year and/or 0.5 < PR < 1 gr/day
- State 3: CL -30% in 1 year and/or PR > 1 gr/day
- State 4: Definitive rejection
- State 5: Death with kidney function
## Results (1)

### Covariates associated with the initial probabilities

<table>
<thead>
<tr>
<th>Transition</th>
<th>Covariate</th>
<th>Estim.</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 $\rightarrow$ 1</td>
<td>Intercept</td>
<td>2.85</td>
<td>0.19</td>
<td>0.0001</td>
</tr>
<tr>
<td>0 $\rightarrow$ 1</td>
<td>Recipient Gender</td>
<td>-0.39</td>
<td>0.17</td>
<td>0.0226</td>
</tr>
<tr>
<td>0 $\rightarrow$ 1</td>
<td>Delayed graft function</td>
<td>-0.53</td>
<td>0.17</td>
<td>0.0014</td>
</tr>
<tr>
<td>0 $\rightarrow$ 2</td>
<td>Intercept</td>
<td>-0.67</td>
<td>0.44</td>
<td>0.1258</td>
</tr>
<tr>
<td>0 $\rightarrow$ 2</td>
<td>Cold ischemia time</td>
<td>1.13</td>
<td>0.44</td>
<td>0.0092</td>
</tr>
</tbody>
</table>
Covariates associated with the intensities of transition

<table>
<thead>
<tr>
<th>Transition</th>
<th>Covariate</th>
<th>Estim.</th>
<th>SE</th>
<th>RR</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 → 2</td>
<td>Year of transplant</td>
<td>-0.80</td>
<td>0.12</td>
<td>0.45</td>
<td>0.0001</td>
</tr>
<tr>
<td>1 → 3</td>
<td>Recipient Gender</td>
<td>0.29</td>
<td>0.15</td>
<td>1.34</td>
<td>0.0484</td>
</tr>
<tr>
<td>1 → 3</td>
<td>Year of transplant</td>
<td>-1.20</td>
<td>0.21</td>
<td>0.30</td>
<td>0.0001</td>
</tr>
<tr>
<td>2 → 3</td>
<td>Year of transplant</td>
<td>-0.54</td>
<td>0.12</td>
<td>0.59</td>
<td>0.0001</td>
</tr>
<tr>
<td>3 → 5</td>
<td>Recipient age</td>
<td>1.48</td>
<td>0.39</td>
<td>4.41</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
### Results (3)

#### Parameters of the waiting times distributions

<table>
<thead>
<tr>
<th>Transition</th>
<th>$\sigma_{ij}$</th>
<th>ET</th>
<th>$\nu_{ij}$</th>
<th>ET</th>
<th>$\theta_{ij}$</th>
<th>ET</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 $\rightarrow$ 2</td>
<td>36.14</td>
<td>31.97</td>
<td>0.53</td>
<td>0.03</td>
<td>0.24</td>
<td>0.09</td>
</tr>
<tr>
<td>1 $\rightarrow$ 3</td>
<td>34.11</td>
<td>65.20</td>
<td>0.52</td>
<td>0.05</td>
<td>0.19</td>
<td>0.15</td>
</tr>
<tr>
<td>2 $\rightarrow$ 3</td>
<td>33.40</td>
<td>31.34</td>
<td>0.56</td>
<td>0.03</td>
<td>0.30</td>
<td>0.13</td>
</tr>
<tr>
<td>3 $\rightarrow$ 4</td>
<td>10.16</td>
<td>1.56</td>
<td>1.49</td>
<td>0.11</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3 $\rightarrow$ 5</td>
<td>18.48</td>
<td>47.62</td>
<td>1.14</td>
<td>0.23</td>
<td>1.46</td>
<td>3.75</td>
</tr>
</tbody>
</table>
Concluding remarks

Summary of the results

- Multinomial logistic regression useful in order to identify covariates associated with the initial probabilities.

- Parcimony of the generalized Weibull distribution (∪ – or ∩ – shape).
Concluding remarks

Summary of the results

- Multinomial logistic regression useful in order to identify covariates associated with the initial probabilities.
- Parcimony of the generalized Weibull distribution ($\bigcup$ – or $\bigcap$ – shape).

Limits of the model and work in progress

- Delate the transition $1 \rightarrow 3$, even if this transition is informative for clinicians.

$$ P_{12}P_{23} \int_0^{d_{h,r}} f_{12}(x)f_{23}(d_{h,r} - x)dx $$

- Estimate the cut-off of the markers in order to determine the best states of gravity.
- Construction of an hidden semi-Markov model in order to take into account the short-term fluctuation.