Accommodating informative dropout and death: a joint modelling approach for longitudinal and semi-competing risks data

Qiuju Li

MRC Biostatistics Unit, Cambridge, UK

qiuju.li@mrc-bsu.cam.ac.uk

Joint work with Dr. Li Su

Workshop on Flexible Models for Longitudinal and Survival Data with Applications in Biostatistics (Warwick, 27th-29th July, 2015)
Introduction

Joint modelling of longitudinal and semi-competing risks data

Application: HERS data analysis

Conclusions

Outline

1. Introduction

2. Joint modelling of longitudinal and semi-competing risks data

3. Application: HERS data analysis

4. Conclusions
longitudinal and semi-competing risks data, e.g., CD4 counts, dropout and HIV-related death in the HIV epidemiology research study (HERS).

Measurements of CD4 counts over time for different subjects:
- Complete data
- Dropout
- Death
- Dropout & Death
Joint modelling of longitudinal and semi-competing risks data

Application: HERS data analysis

Conclusions

Some concepts

- mortal cohort;
- immortal cohort;
  
  (Aalen and Gunnes, 2010)

longitudinal profile models:

- unconditional models, e.g., random-effects models \( f(Y_i(t)) \);
- fully conditional models, e.g., \( f(Y_i(t) | S_i = s), s > t \);
- partly conditional models, e.g., \( f(Y_i(t) | S_i > t) \) (Kurland and Heagerty, 2005; Kurland et al., 2009);
  
  - GEE approaches;
  - a likelihood-based joint modelling approach proposed subsequently.
Notation

- scheduled repeated measurements of a longitudinal outcome $Y_i = (Y_{i1}, \ldots, Y_{iM})'$, taken at visits $1, \ldots, M$, e.g., $M = 12$ for the HERS data;
- informative dropout and death
  - dropout time denoted by $D_i$, observed data $D_i^* = \min(D_i, S_i, C_i)$, $\delta_i^D = I(D_i \leq S_i, D_i \leq C_i)$;
  - death time denoted by $S_i$, observed data $S_i^* = \min(S_i, C_i)$, $\delta_i^S = I(S_i \leq C_i)$;
  - $C_i$ denotes non-informative censoring, e.g., end of study;
- covariates $X_i$, e.g., sex, treatment arm;
Time-to-event processes

Time-to-event data,
- time to dropout: last visit of follow-up;
- time to death:

\[ 0 \] \[ T_i \] \[ \tau \]

\( \tau \): the end of study.
Discrete time-to-event data (*mathematical attractiveness*),

- time to dropout: last visit of follow-up;
- time to death:

\[
S_i = \tau \quad \text{or} \quad S_i = r 
\]

\(
\tau: \text{the end of study.} 
\)

The discrete death time \( S_i = r \) (*Barrett et al, 2015, JRSSB*).
Joint models

Joint models for the longitudinal and semi-competing risks data,

\[
\begin{align*}
Y_{ij} &= x_{ij}^T \beta + z_{ij}^T b_i + \epsilon_{ij} \\
Pr(D_i = r | D_i \geq r) &= 1 - \Phi(x_{D,i}^T \alpha_D + \gamma_{D,r}^T W_{D,i} b_i) \\
Pr(S_i = r | S_i \geq r) &= 1 - \Phi(x_{S,i}^T \alpha_S + \gamma_{S,r}^T W_{S,i} b_i)
\end{align*}
\]

- \(\Phi(\cdot)\) standard normal cdf;
- \(\beta, \alpha_D, \alpha_S\) regression coefficients; \(\gamma_{D,r}, \gamma_{S,r}\) association effects;
- random effects \(b_i \sim N(0, \Sigma_b)\);
- \(\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)\);
- \(W_{D,i} b_i, W_{S,i} b_i\) vectors of linear combinations of random effects, e.g., \(W_{D,i} b_i = (b_{i0}, b_{i1})^T\);
- conditional independence assumption given random effects \(b_i\);
Likelihood function,

\[
\prod_i L_i(\theta; y_i, D_i^*, \delta_i^D, S_i^*, \delta_i^S) = \prod_i \int_{-\infty}^{\infty} f(\text{longitudinal data}|\theta, b_i) \times Pr(\text{dropout data}|\theta, b_i) \times Pr(\text{death data}|\theta, b_i) \times f(b_i|\theta) \, db_i.
\]
Four possible scenarios of the observed time-to-event data,

(1) neither dropout nor death:
$$D_i^* = d, S_i^* = s, (\delta_i^D, \delta_i^S) = (0, 0);$$

(2) dropout only:
$$D_i^* = d, S_i^* = s, (\delta_i^D, \delta_i^S) = (1, 0);$$

(3) death only:
$$D_i^* = d, S_i^* = s, (\delta_i^D, \delta_i^S) = (0, 1);$$

(4) both dropout and death:
$$D_i^* = d, S_i^* = s, (\delta_i^D, \delta_i^S) = (1, 1);$$
For the scenario (1), the likelihood function of observed data \( \{y_i = (y_{i1}, \ldots, y_{in_i})', D_i^* = d, \delta_i^D = 0, S_i^* = s, \delta_i^S = 0\} \),

\[
L_i(\theta; y_i, D_i^*, \delta_i^D, S_i^*, \delta_i^S) = \int_{-\infty}^{\infty} \phi(y_i; x_i\beta + z_i b_i, \sigma^2 I_{n_i}) \prod_{k=1}^{d} \Phi(x_{D,ir}^T \alpha_D + \gamma_{D,r}^T W_{D,ir} b_i) \prod_{\ell=1}^{s} \Phi(x_{S,ir}^T \alpha_S + \gamma_{S,r}^T W_{S,ir} b_i) \phi(b_i; 0, \Sigma_b) db_i
\]

\[= L_i(\cdot \backslash b_i) \Phi^{d+s}(A_{ds} + B_{ds} h_i; 0, I_{d+s} + B_{ds} H_i^{-1} B_{ds}^T) \]

- **closed-form likelihood** (skewed normal distribution, Arnold 2009);
- \( L_i(\cdot \backslash b_i), h_i, H_i, A_{ds}, B_{ds} \) function/vectors/matrices free of \( b_i \);
- \( \phi(\cdot; \mu, \Sigma) \) and \( \Phi^{d+s}(\cdot; \mu, \Sigma) \) denote multivariate normal pdf/cdf.
Marginal mean profile conditional on being alive

- Unconditional population mean profile for an immortal cohort
  \[ E(Y_{ij}|x_{ij}) = x_{ij}^T \beta; \]

- Conditional mean profile given being alive for a mortal cohort, we can compute
  \[ E(Y_{ij}|x_{ij}, S_i \geq j) = x_{ij}^T \beta + z_{ij}^T E(b_i|S_i \geq j). \]

Analogously, \( f(b_i|S_i \geq j) \) is a multivariate skew-normal distribution,

\[
 f(b_i|S_i \geq j) = f(b_i|S_i > (j - 1)) = \frac{Pr(S_j > (j - 1)|b_i) f(b_i)}{Pr(S_i > (j - 1))},
\]

closed form of its expectation can be obtained.
Statistical inference

1. Maximum likelihood-based approach (*exact likelihood*);
   - R software utilising *nlminb* or *optim*.
2. Bayesian approach;
   - implemented using WinBUGS.
HERS data description

HIV epidemiology research study (HERS) (Smith et al., 1997)

<table>
<thead>
<tr>
<th># of subjects: 850 (HIV positive at baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD4 counts reviewed every 6 months up to 12 visits</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>time-to-event data</th>
<th># of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>scenario (1): $(\delta_i^D, \delta_i^S) = (0, 0)$</td>
<td>374</td>
</tr>
<tr>
<td>scenario (2): $(\delta_i^D, \delta_i^S) = (1, 0)$</td>
<td>352</td>
</tr>
<tr>
<td>scenario (3): $(\delta_i^D, \delta_i^S) = (0, 1)$</td>
<td>23</td>
</tr>
<tr>
<td>scenario (4): $(\delta_i^D, \delta_i^S) = (1, 1)$</td>
<td>78</td>
</tr>
</tbody>
</table>

Objective: study the role of baseline patient characteristics (i.e., viral load, antiretroviral therapy (art), # of symptoms) on variation in longitudinal CD4 counts.
Models proposed for the HERS data

\[
Y_{ij} = \beta_0 + \beta_1 \text{visit} + \beta_2 \sim \text{viral load} + \beta_5 \text{symptoms} + \beta_6 \text{art} \\
+ \beta_7 \sim \text{visit} \ast \text{viral load} + \beta_{10} \text{visit} \ast \text{symptoms} + \beta_{11} \text{visit} \ast \text{art} + b_{i0} + b_{i1} + \epsilon_{ij}
\]

\[
Pr(D_i = r|D_i \geq r) = 1 - \Phi(\alpha_{D,i0} + \alpha_{D,i1} \sim \text{viral load} + \alpha_{D,i4} \text{symptoms} \\
+ \alpha_{D,i5} \text{art} + \alpha_{D,i6} r + \alpha_{D,i7} r^2 + \gamma_{D,0} b_{i0} + \gamma_{D,1} b_{i1})
\]

\[
Pr(S_i = r|S_i \geq r) = 1 - \Phi(\alpha_{S,i0} + \alpha_{S,i1} \sim \text{viral load} + \alpha_{S,i4} \text{symptoms} \\
+ \alpha_{S,i5} \text{art} + \alpha_{S,i6} r + \alpha_{S,i7} r^2 + \gamma_{S,0} b_{i0} + \gamma_{S,1} b_{i1})
\]
### WinBUGS results

#### Longitudinal process (CD4 counts)

<table>
<thead>
<tr>
<th></th>
<th>Joint modelling (WinBUGS)</th>
<th>Linear mixed effects (LME)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>sd</td>
</tr>
<tr>
<td>intercept</td>
<td>15.11</td>
<td>0.71</td>
</tr>
<tr>
<td>visit</td>
<td>-0.87</td>
<td>0.13</td>
</tr>
<tr>
<td>viral load [1] (0-500)</td>
<td>10.00</td>
<td>0.79</td>
</tr>
<tr>
<td>viral load [2] (500-5k)</td>
<td>6.61</td>
<td>0.72</td>
</tr>
<tr>
<td>viral load [3] (5k-30k)</td>
<td>2.94</td>
<td>0.82</td>
</tr>
<tr>
<td>symptoms</td>
<td>-0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>art at baseline</td>
<td>-4.66</td>
<td>0.43</td>
</tr>
<tr>
<td>visit*viral load [1]</td>
<td>0.47</td>
<td>0.14</td>
</tr>
<tr>
<td>visit*viral load [2]</td>
<td>0.44</td>
<td>0.13</td>
</tr>
<tr>
<td>visit*viral load [3]</td>
<td>0.28</td>
<td>0.14</td>
</tr>
<tr>
<td>visit*symptoms</td>
<td>-0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>visit*art</td>
<td>0.12</td>
<td>0.06</td>
</tr>
<tr>
<td>$\gamma_{D,0}$</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>$\gamma_{D,1}$</td>
<td>0.44</td>
<td>0.05</td>
</tr>
<tr>
<td>$\gamma_{S,0}$</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>$\gamma_{S,1}$</td>
<td>1.23</td>
<td>0.17</td>
</tr>
</tbody>
</table>
Introduction

Joint modelling of longitudinal and semi-competing risks data

Application: HERS data analysis

Conclusions

Marginal mean profiles

Subjects: viral load [1], # of symptoms=1, antiretroviral therapy (art) at baseline;

\[
E(Y_{ij} | x_{ij}) = x_{ij}^T \beta
\]

naive LME: \(E(Y_{ij} | x_{ij}) = x_{ij}^T \beta\)

joint model: \(E(Y_{ij} | x_{ij}) = x_{ij}^T \beta\)

\(E(Y_{ij} | x_{ij}, S_i \geq j)\)
Conclusions

1. a likelihood-based approach to capture the partly conditional mean profiles, accommodating both informative dropout and death;
2. offer inference for both mortal and immortal cohort;
3. a new model for semi-competing risks data in the joint modelling framework;
4. approach demonstrated by an analysis of the HERS data.
Key references


Thank You!

Authors:
Qiuju Li & Li Su
MRC Biostatistics Unit, Cambridge, UK