Puberty-T1D analyses

Essi Syrjala

December 2, 2020

1 Model

Three-state progressive model for continuous time data with states

- 1 1: Non-diabetic and autoantibody negative ("healthy")
- 2: Autoantibody positive (interval censored)
- 3: Type 1 diabetes ("exact")

will be used to study the effect of pubertal onset on the development of autoantibodies and type 1 diabetes.

Study included regular 3-12-month follow-up for diabetes associated autoantibodies, including ICA, IAA, GADA, and IA-2A. Information on type 1 diabetes diagnosis comes from register.

"Progressive" means that recovery back to previous state is not possible. Then, transition intensity matrix is of the form

$$Q = \begin{pmatrix} -(h_{12}(t) + h_{13}(t)) & h_{12}(t) & h_{13}(t) \\ 0 & -h_{23}(t) & h_{23}(t) \\ 0 & 0 & 1 \end{pmatrix},$$

where $h_{rr'}(t)$ are transition intensities between the states at time t, with r being the previous state and r' the current state.

2 Data simulation from exponential distribution

2.1 Without covariates

Let

- T_{12} be time from state 1 to state 2
- T_{13} time from state 1 to state 3
- T_{23} time from state 2 to state 3

and

$$T_{12} \sim Exp(h_{12}), T_{13} \sim Exp(h_{13}), T_{23} \sim Exp(h_{23}),$$

Let's set

$$h_{12} = \frac{1}{20}, h_{13} = \frac{1}{100}, h_{23} = \frac{1}{10}.$$

which implies that

$$E(T_{12}) = 20, E(T_{13}) = 100, E(T_{23}) = 2.$$

Transition intensity matrix is then

$$Q = \begin{pmatrix} -(\frac{1}{20} + \frac{1}{100}) & \frac{1}{20} & \frac{1}{100} \\ 0 & -\frac{1}{10} & \frac{1}{10} \\ 0 & 0 & 1 \end{pmatrix}.$$

In a time-homogeneous continuous-time Markov model, a single period of occupancy in state r has an exponential distribution, with rate given by h_{rr}, (or mean by $1/h_{rr}$). The remaining elements of the rth row of Q are proportional to the probabilities governing the next state after r to which the individual makes a transition. (Multi-state modelling with R: the msm package)

Simulation was done by simulating "exact" transition times. Interval censoring and follow-up were added afterwards.

Simulation process:

- C is a censoring time
- N is a number of the individuals
- t_{ki} is a kth transition time for individual i and $s_i(t_{ki})$ is a state for individual i at t_{ki}; $k = 1, 2$
- $T_{rr'.i}$ is a time from state r to state r' for individual i; $r, r' = 1, 2, 3$

1. Simulate separately N transition times $T_{12.i}$ from $Exp(h_{12})$ and $T_{13.i}$ from $Exp(h_{13})$ distributions
   (a) t_{1i} =min$(T_{12.i}, T_{13.i})$
      i. If $t_{1i} = T_{12.i}$: $s_i(t_{1i}) = 2$
      ii. If $t_{1i} = T_{13.i}$: $s_i(t_{1i}) = 3$
      iii. If $t_{1i} > C$: $t_{1i} = C$ and $s_i(t_{1i}) = 99(censored)$
2. For children with $t_{1i} = T_{12.i}$, simulate transition time $T_{23.i}$ from $Exp(h_{23})$
   (a) $t_{2i} = t_{1i} + T_{23.i}$ and $s_i(t_{2i}) = 3$
      i. If $t_{2i} > C$: $t_{2i} = C$ and $s_i(t_{2i}) = 2$

The 3-month follow-up was added to the data because autoantibodies (state 2) were measured at particular time points. Now, if child's transition age to the state 2 is for example 2.4 years (27 months), he/she is observed at 2.3 years in state 1 and at 2.6 years in state 2. Adding of the follow-up to the data causes also some not allowed $3->2$ transitions. In these cases, transition to state 3 has observed before the follow-up visit for autoantibodies and thus transition to state 2 was eliminated from the data.

2.2 With covariates

Let $\beta_{rr'}$ be covariate effect on transition rate from state r to state r', $r, r' = 1, 2, 3$, and $x_i(t)$ the covariate value for individual i at time t. Transition intensities become as:

$$h_{rr'}(t) = \begin{cases} h_{rr'};\text{ if } x_i(t) = 0 \\ \beta_{rr'} h_{rr'};\text{ if } x_i(t) = 1 \end{cases}$$

3 Model equations

Here, a model for the hazards is not specified yet. Given time interval $(t_{j-1}, t_j]$, $j = 1, ..., J$ is the number of the observation, the cumulative hazard function for leaving state 1 is

$$H_1\left(t_{j-1}, t_j\right) = \int_{t_{j-1}}^{t_j} h_{12}(u) + h_{13}\left(u\right) du \tag{1}$$

and for leaving state 2 is

$$H_2\left(t_{j-1}, t_j\right) = \int_{t_{j-1}}^{t_j} h_{23}(u) du. \tag{2}$$

$t_0 = 0$ for every child. If there is covariate, transition intensities $h_{rr'}(t)$ consist of baseline intensity multiplied by the covariate effect:

$$h_{rr'}(t) = h_{rr'.0}(t) \exp(\beta_{rr'} z_i(t)). \tag{3}$$

Let s_j be a state of the measurement j. Transition probabilities from state r to state r' are of the form $p_{rr'}\left(t_{j-1}, t_j\right) = P\left(s_{j-1} = r' \mid s_j = r\right)$. If we assume three-state progressive model (back-transitions are not allowed), transition probabilities for panel-observed data can be given by

1. Leaving from state 1 (1st row of the transition probability matrix)
   (a) $p_{11}\left(t_{j-1}, t_j\right) = \exp\left(-H_1\left(t_{j-1}, t_j\right)\right) = \exp\left(-\int_{t_{j-1}}^{t_j} h_{12}\left(u\right) + h_{13}\left(u\right) du\right)$
   (b) $p_{12}(t_{j-1}, t_j) = \int_{t_{j-1}}^{t_j} p_{11}\left(t_{j-1}, u\right) h_{12}(u) p_{22}\left(u, t_j\right) du$
   (c) $p_{13}\left(t_{j-1}, t_j\right) = 1 - p_{11}\left(t_{j-1}, t_j\right) - p_{12}\left(t_{j-1}, t_j\right)$ OR

(d) $p_{13} = p_{13}^{1}(t_{j-1}, t_j) + p_{13}^{2}(t_{j-1}, t_j)$, where

i. $p_{13}^{1}(t_{j-1}, t_j)$ stands for the straight transition from state 1 to 3

- $p_{13}^{1}(t_{j-1}, t_j) = \int_{t_{j-1}}^{t_j} p_{11}(t_{j-1}, v)\, h_{13}(v) p_{33}(v, t_j)\, dv$
  $= \int_{t_{j-1}}^{t_j} \exp\left(-\int_{t_{j-1}}^{v} h_{12}(u) + h_{13}(u)\, du\right) h_{13}(v)\, dv$

ii. $p_{13}^{2}(t_{j-1}, t_j)$ transition from state 1 to 3 via state 2

- $p_{13}^{2}(t_{j-1}, t_j) = \int_{t_{j-1}}^{t_j} \exp\left(-\int_{t_{j-1}}^{v} h_{12}(u) + h_{13}(u)\, du\right) h_{12}(v)\, p_{23}(v, t_j)\, dv$
  $= \int_{t_{j-1}}^{t_j} \left(\exp\left(-\int_{t_{j-1}}^{v} h_{12}(u) + h_{13}(u)\, du\right) h_{12}(v) \int_{v}^{t_j} \exp\left(-\int_{v}^{w} h_{23}(u)\, du\right) h_{23}(w)\, dw\right) dv$

2. Leaving from state 2 (2nd row of the transition probability matrix)

(a) $p_{21}(t_{j-1}, t_j) = 0$

(b) $p_{22}(t_{j-1}, t_j) = \exp(-H_2(t_{j-1}, t_j))$

(c) $p_{23}(t_{j-1}, t_j) = 1 - p_{22}(t_{j-1}, t_j)$ OR

(d) $p_{23} = \int_{t_{j-1}}^{t_j} p_{22}(t_{j-1}, v)\, h_{23}(v) p_{33}(v, t_j)\, dv$
$= \int_{t_{j-1}}^{t_j} \exp\left(-\int_{t_{j-1}}^{v} h_{23}(u)\, du\right) h_{23}(v)\, dv * 1$

3. Leaving from state 3 (3rd row of the transition intensity matrix)

(a) $p_{31}(t_{j-1}, t_j) = 0$

(b) $p_{32}(t_{j-1}, t_j) = 0$

(c) $p_{33}(t_{j-1}, t_j) = 1$

3.1 Contributions to the likelihood

Individual contribution to the likelihood can be calculated by multiplying over all of the appropriate transition probabilities

$$L_i = \left[\prod_{j=1}^{J-1} p_{rr'}(t_{i,j-1}, t_{i,j})\right] C(t_{i,J-1}, t_{i,J})$$

where $C(t_{i,J-1}, t_{i,J})$ is needed for the possible different definitions of the last state:

- If state at $t_{i,J}$ is right censored
  - $C(t_{i,J-1}, t_{i,J}) = \sum_{s=0}^{1} p_{rr'}(t_{i,J-1}, t_{i,J})$
    * Individual has not reached state 3 but state at the study end (at $t_{i,J}$) is uncertain:
      · If $r = 1$ (previous observed state (at $t_{i,J-1}$) is 1) state at $t_{i,J}$ can be 1 or 2: $p_{11}(t_{i,J-1}, t_{i,J}) + p_{12}(t_{i,J-1}, t_{i,J})$

- If $r = 2$ (previous observed state (at $t_{i,J-1}$) is 2) we know that state at $t_{i,J}$ can only be 2: $p_{22}\left(t_{i,J-1}, t_{i,J}\right)$

- If state at $t_{i,J}$ is 3
  - $C\left(t_{i,J-1}, t_{i,J}\right) = \sum_{r=0}^{1} p_{rr'}\left(t_{i,J-1}, t_{i,J}\right) q_{r3}\left(t_J\right)$
    - Individual has reached state 3 but state before the diagnosis (at $t_{i,J-1}$) is uncertain:
      - If $r = 1$ (previous observed state (at $t_{i,J-1}$) is 1) state at $t_{i,J-1}$ can be 1 or 2: $p_{11}\left(t_{i,J-1}, t_{i,J}\right) h_{13}\left(t_J\right) + p_{12}\left(t_{i,J-1}, t_{i,J}\right) h_{23}\left(t_J\right)$
      - If $r = 2$ (previous observed state (at $t_{i,J-1}$) is 2) state at $t_{i,J-1}$ can only be 2: $p_{22}\left(t_{i,J-1}, t_{i,J}\right) h_{23}\left(t_J\right)$

- Otherwise
  - $C\left(t_{i,J-1}, t_{i,J}\right) = p_{rr'}\left(t_{i,J-1}, t_{i,J}\right)$

Final likelihood can be calculated by multiplying over all of the individual contributions (s: state, x: covariate):

$$\prod_{i}^{N} L_i(\boldsymbol{\theta}|\boldsymbol{s}, \boldsymbol{x})$$

3.2 Exponential model

3.2.1 Without covariates

If a model for the hazards is exponential, transition-specific hazards are specified by constants $h_{rr'}(t) = h_{rr'}$ for all states r, r' at any time t. Cumulative hazard function for leaving state 1 becomes as

- $\exp\left(-H_1\left(t_{j-1}, t_j\right)\right) = \exp\left(-\int_{t_{j-1}}^{t_j} h_{12}(u) + h_{13}\left(u\right) du\right)$
  $= \exp\left(-\left(h_{12} + h_{13}\right)\left(t_j - t_{j-1}\right)\right)$

  and for leaving state 2

- $\exp\left(-H_2\left(t_{j-1}, t_j\right)\right) = \exp\left(-\int_{t_{j-1}}^{t_j} h_{23}(u) du\right)$
  $= \exp\left(-h_{23}\left(t_j - t_{j-1}\right)\right)$.

Transition probabilities become as

- $p_{11}(t_{j-1}, t_j) = \exp\left(-\left(h_{12} + h_{13}\right)\left(t_j - t_{j-1}\right)\right)$
- $p_{12}(t_{j-1}, t_j) = \int_{t_{j-1}}^{t_j} p_{11}\left(t_{j-1}, u\right) h_{12}(u) p_{22}\left(u, t_j\right) du$
  $= \int_{t_{j-1}}^{t_j} \exp\left(-H_1\left(t_{j-1}, u\right)\right) h_{12}(u) \exp\left(-H_2\left(u, t_j\right)\right) du$
  $= \frac{h_{12}}{h_{12}+h_{13}-h_{23}}\left(\exp\left(-h_{23}\left(t_j - t_{j-1}\right)\right) - \exp\left(-\left(h_{12} + h_{13}\right)\left(t_j - t_{j-1}\right)\right)\right)$

- $p_{13}(t_{j-1}, t_j) = \exp\left(-\left(h_{12} + h_{13}\right)\left(t_j - t_{j-1}\right)\right) h_{13}(t_j) +$
  $\frac{h_{12}}{h_{12}+h_{13}-h_{23}} \left(\exp\left(-h_{23}\left(t_j - t_{j-1}\right)\right) - \exp\left(-\left(h_{12} + h_{13}\right)\left(t_j - t_{j-1}\right)\right)\right) h_{23}(t_j)$
- $p_{22}(t_{j-1}, t_j) = \exp\left(-h_{23}\left(t_j - t_{j-1}\right)\right)$
- $p_{23}(t_{j-1}, t_j) = \exp\left(-h_{23}\left(t_j - t_{j-1}\right)\right) h_{23}(t_j)$
- $p_C(t_{j-1}, t_j) = \exp\left(-\left(h_{12} + h_{13}\right)\left(t_j - t_{j-1}\right)\right) +$
  $\frac{h_{12}}{h_{12}+h_{13}-h_{23}} \left(\exp\left(-h_{23}\left(t_j - t_{j-1}\right)\right) - \exp\left(-\left(h_{12} + h_{13}\right)\left(t_j - t_{j-1}\right)\right)\right)$

3.2.2 With covariates

If a model for the hazards is exponential and we have time-dependent covariate, transition-specific hazards are specified by $h_{rr'}(t) = h_{rr'.0} \exp(\beta_{rr'} z_i(t))$ for all states r, r' at time t.

Model 1 If b_i is a pubertal onset age for child i:

$$h_{rr'i}(t) = \begin{cases} h_{rr'.0}, & t < b_i - 1 \\ h_{rr'.0} \exp(\beta_{rr'}), & t \geq b_i - 1 \end{cases}$$

Model 2 If b_i is a pubertal onset age for child i:

$$h_{rr'i}(t) = \begin{cases} h_{rr'.0}, & t < b_i - 1 \\ h_{rr'.0} \exp(\beta_{rr'}), & b_i - 1 \leq t < b_i + 1 \\ h_{rr'.0}, & t \geq b_i + 1 \end{cases}$$

Cumulative hazard function for leaving state 1 becomes as

- $\exp\left(-H_1\left(t_{j-1}, t_j\right)\right)$
  $= \exp\left(-\int_{t_{j-1}}^{t_j} \left(h_{12}(u) + h_{13}(u)\right) du\right)$

$$= \begin{cases} \exp\left(-\int_{t_{j-1}}^{t_j} \left(h_{12.0} + h_{13.0}\right) du\right), & t_{j-1} < b_i - 1 \\ \exp\left(-\int_{t_{j-1}}^{t_j} \left(h_{12.0} \exp(\beta_{12}) + h_{13.0} \exp(\beta_{13})\right) du\right), & b_i - 1 \leq t_{j-1} < b_i + 1 \\ \exp\left(-\int_{t_{j-1}}^{t_j} \left(h_{12.0} + h_{13.0}\right) du\right), & t_{j-1} \geq b_i + 1 \end{cases}$$

$$= \begin{cases} \exp\left(-\left(h_{12.0} + h_{13.0}\right)\left(t_j - t_{j-1}\right)\right), & t_{j-1} < b_i - 1 \\ \exp\left(-\left(h_{12.0} \exp(\beta_{12}) + h_{13.0} \exp(\beta_{13})\right)\left(t_j - t_{j-1}\right)\right), & b_i - 1 \leq t_{j-1} < b_i + 1 \\ \exp\left(-\left(h_{12.0} + h_{13.0}\right)\left(t_j - t_{j-1}\right)\right), & t_{j-1} \geq b_i + 1 \end{cases}$$

and for leaving state 2

- $\exp\left(-H_2\left(t_{j-1}, t_j\right)\right)$

  $= \exp\left(-\int_{t_{j-1}}^{t_j} h_{23}(u)du\right)$

$$= \begin{cases} \exp\left(-\int_{t_{j-1}}^{t_j} h_{23}du\right), & t_{j-1} < b_i - 1 \\ \exp\left(-\int_{t_{j-1}}^{t_j} h_{23}\exp(\beta_{23})du\right), & b_i - 1 \leq t_{j-1} < b_i + 1 \\ \exp\left(-\int_{t_{j-1}}^{t_j} h_{23}du\right), & t_{j-1} \geq b_i + 1 \end{cases}$$

$$= \begin{cases} \exp\left(-h_{23.0}\left(t_j - t_{j-1}\right)\right), & t_{j-1} < b_i - 1 \\ \exp\left(-h_{23.0}\exp(\beta_{23})\left(t_j - t_{j-1}\right)\right), & b_i - 1 \leq t_{j-1} < b_i + 1 \\ \exp\left(-h_{23.0}\left(t_j - t_{j-1}\right)\right), & t_{j-1} \geq b_i + 1 \end{cases}$$

Transition probabilities become as

$$p_{11}(t_{j-1}, t_j) = \begin{cases} \exp\left(-\left(h_{12.0} + h_{13.0}\right)\left(t_j - t_{j-1}\right)\right), & t_{j-1} < b_i - 1 \\ \exp\left(-\left(h_{12.0}\exp(\beta_{12}) + h_{13.0}\exp(\beta_{13})\right)\left(t_j - t_{j-1}\right)\right), & b_i - 1 \leq t_{j-1} < b_i + 1 \\ \exp\left(-\left(h_{12.0} + h_{13.0}\right)\left(t_j - t_{j-1}\right)\right), & t_{j-1} \geq b_i + 1 \end{cases}$$

$$p_{12}(t_{j-1}, t_j) = \begin{cases} \displaystyle\int_{t_{j-1}}^{t_j} \exp\left(-\int_{t_{j-1}}^{v} (h_{12.0} + h_{13.0})du\right) & \\ \displaystyle h_{12.0}\exp\left(-\int_{v}^{t_j} h_{23.0}du\right)dv, & t_{j-1} < b_i - 1 \\ \displaystyle\int_{t_{j-1}}^{t_j} \exp\left(-\int_{t_{j-1}}^{v} (h_{12}(u) + h_{13}(u))du\right) & \\ \displaystyle h_{12}(v)\exp\left(-\int_{v}^{t_j} h_{23}\left(u\right)du\right)dv, & b_i - 1 \leq t_{j-1} < b_i + 1 \\ \displaystyle\int_{t_{j-1}}^{t_j} \exp\left(-\int_{t_{j-1}}^{v} (h_{12.0} + h_{13.0})du\right) & \\ \displaystyle h_{12.0}\exp\left(-\int_{v}^{t_j} h_{23.0}du\right)dv, & t_{j-1} \geq b_i + 1 \end{cases}$$

$$p_{12}(t_{j-1},t_j)=\begin{cases}\int_{t_{j-1}}^{t_j}\exp\left(-\int_{t_{j-1}}^{v}(h_{12.0}+h_{13.0})du\right) & \\ h_{12.0}\exp\left(-\int_{v}^{t_j}h_{23.0}du\right)dv, & t_{j-1}<b_i-1\\ \int_{t_{j-1}}^{t_j}\exp\left(-\int_{t_{j-1}}^{v}(h_{12.0}\exp(\beta_{12})+h_{13.0}\exp(\beta_{13}))du\right) & \\ h_{12.0}\exp(\beta_{12})\exp\left(-\int_{v}^{t_j}h_{23.0}\exp(\beta_{23})du\right)dv, & b_i-1\le t_{j-1}<b_i+1\\ \int_{t_{j-1}}^{t_j}\exp\left(-\int_{t_{j-1}}^{v}(h_{12.0}+h_{13.0})du\right) & \\ h_{12.0}\exp\left(-\int_{v}^{t_j}h_{23.0}du\right)dv, & t_{j-1}\ge b_i+1\end{cases}$$

$$p_{12}(t_{j-1},t_j)=\begin{cases}\frac{h_{12.0}}{h_{12.0}+h_{13.0}-h_{23.0}}(\exp(-h_{23.0}\,(t_j-t_{j-1}))- & \\ \exp\left(-\left(h_{12.0}+h_{13.0}\right)(t_j-t_{j-1})\right)), & t_{j-1}<b_i-1\\ \frac{h_{12.0}\exp(\beta_{12})}{h_{12.0}\exp(\beta_{12})+h_{13.0}\exp(\beta_{13})-h_{23.0}\exp(\beta_{23})} & \\ (\exp(-h_{23.0}\exp(\beta_{23})\,(t_j-t_{j-1}))- & \\ \exp\left(-\left(h_{12.0}\exp(\beta_{12})+h_{13.0}\exp(\beta_{13})\right)(t_j-t_{j-1})\right)), & b_i-1\le t_{j-1}<b_i+1\\ \frac{h_{12.0}}{h_{12.0}+h_{13.0}-h_{23.0}}(\exp(-h_{23.0}\,(t_j-t_{j-1}))- & \\ \exp\left(-\left(h_{12.0}+h_{13.0}\right)(t_j-t_{j-1})\right)), & t_{j-1}\ge b_i+1\end{cases}$$

$$p_{22}(t_{j-1},t_j)=\begin{cases}\exp\left(-h_{23.0}\,(t_j-t_{j-1})\right), & t_{j-1}<b_i-1\\ \exp\left(-h_{23}\exp(\beta_{23})\,(t_j-t_{j-1})\right), & b_i-1\le t_{j-1}<b_i+1\\ \exp\left(-h_{23.0}\,(t_j-t_{j-1})\right), & t_{j-1}\ge b_i+1\end{cases}$$

4 Puberty

4.1 Different scenarios

Let $z_i\,(t)$ stands for the puberty status of the child i at time t, and pub_i for the pubertal onset timing of child i.

Let's set

$$h_{12}=0.05, h_{13}=0.01, h_{23}=0.5$$

$$\beta_{12} = \log(2), \beta_{13} = \log(3), \beta_{23} = \log(4),$$

where $h_{rr'}$s are baseline transition intensities and $\beta_{rr'}$s are the parameter estimates of puberty for different transitions.

1. Puberty affects permanently

   $$z_i\left(t\right) = \begin{cases} 0 & t < pub_i \\ 1 & t \geq pub_i \end{cases}$$

   Simulated using Exponential distribution with piecewise-constant rate based on pubertal onset ages:

   $$h_{12i}(t) = \begin{cases} 0.05, 0 \leq t < pub_i \\ 0.05 * 2, t \geq pub_i \end{cases}$$

   $$h_{13i}(t) = \begin{cases} 0.01, 0 \leq t < pub_i \\ 0.01 * 3, t \geq pub_i \end{cases}$$

   $$h_{23i}(t) = \begin{cases} 0.5, t < pub_i - t_{12i} \\ 0.5 * 4, t \geq pub_i - t_{12i} \end{cases}$$

2. Puberty affects temporarily for certain time a around pubertal onset

   $$z_i\left(t\right) = \begin{cases} 0 & t < pub_i - \frac{a}{2} \\ 1 & pub_i \leq t < pub_i + \frac{a}{2} \\ 0 & t \geq pub_i + a \end{cases}$$

   Simulated using Exponential distribution with piecewise-constant rate based on pubertal onset ages:

   $$h_{12i}(t) = \begin{cases} 0.05; 0 \leq t < pub_i \text{ or } t \geq pub_i + 2 \\ 0.05 * 2; pub_i \leq t < pub_i + 2 \end{cases}$$

   $$h_{13i}(t) = \begin{cases} 0.01; 0 \leq t < pub_i \text{ or } t \geq pub_i + 2 \\ 0.01 * 3; pub_i \leq t < pub_i + 2 \end{cases}$$

   $$h_{23i}(t) = \begin{cases} 0.5; t < pub_i - t_{12,i} \text{ or } t \geq pub_i - t_{12,i} + 2 \\ 0.5 * 4; pub_i - t_{12,i} \leq t < pub_i - t_{12,i} + 2 \end{cases}$$

3. Puberty affects temporarily and with evenly reducing effect (based on uniform distribution) for certain time a after pubertal onset

   $$z_i\left(t\right) = \begin{cases} 0 & t < pub_i \\ 1 - \frac{t - pub_i}{(pub_i + a) - pub_i} & pub_i \leq t < pub_i + a \\ 0 & t \geq pub_i + a \end{cases}$$

4. Puberty affects temporarily and with evenly increasing and reducing effect (based on uniform distribution) for certain time a around pubertal onset

$$z_i\left(t\right)=\begin{cases}0 & t<pub_i-\frac{a}{2}\\ \frac{t-pub_i}{\left(pub_i+\frac{a}{2}\right)-pub_i} & pub_i-\frac{a}{2}\leq t<pub_i\\ 1-\frac{t-pub_i}{\left(pub_i+\frac{a}{2}\right)-pub_i} & pub_i\leq t<pub_i+\frac{a}{2}\\ 0 & t\geq pub_i+\frac{a}{2}\end{cases}$$

5 Weibull distribution

5.1 Parametrization

If a continuous variable T has a Weibull distribution with scale parameter $\lambda > 0$ and shape parameter $k > 0$ we can denote $T \sim Weibull(\lambda, k)$. The probability density function is given by

$$f(t)=\begin{cases}\lambda k t^{k-1}\exp(-\lambda t^k) & t\geq 0\\ 0 & t<0,\end{cases} \tag{4}$$

where $S(t) = \exp(-\lambda t^k)$ is a survivor function and $h(t) = \lambda k t^{k-1}$ is a hazard function. Then, the cumulative distribution function $F(t) = 1 - \exp(-\lambda t^k)$.

5.2 Transition probabilities (likelihood contributions)

In Weibull model, transition-specific hazards are time-dependent:

$$h_{rr'}(t)=\lambda_{rr'}k_{rr'}t^{k_{rr'}-1}$$

The transition probabilities are :

$$\begin{aligned}p_{11}\left(t_{j-1},t_j\right)=\exp\left(-H_1(t_{j-1},t_j)\right)&=\exp\left(-\int_{t_{j-1}}^{t_j}\left(h_{12}(u)+h_{13}\left(u\right)\right)du\right)\\ &=\exp\left(-\int_{t_{j-1}}^{t_j}\left(\lambda_{12}k_{12}t^{k_{12}-1}+\lambda_{13}k_{13}t^{k_{13}-1})\right)du\right)\\ &=\exp\left(-\left(\frac{1}{k_{12}}\left(\lambda_{12}k_{12}t^{k_{12}}\right)\Big|_{t_{j-1}}^{t_j}+\frac{1}{k_{13}}\left(\lambda_{13}k_{13}t^{k_{13}}\right)\Big|_{t_{j-1}}^{t_j}\right)\right)\\ &=\exp\left(-\lambda_{12}(t_j^{k_{12}}-t_{j-1}^{k_{12}})-\lambda_{13}(t_j^{k_{13}}-t_{j-1}^{k_{13}})\right)\end{aligned} \tag{5}$$

$$p_{22}\left(t_{j-1},t_j\right)=\exp\left(-H_2(t_{j-1},t_j)\right)=\exp\left(-\int_{t_{j-1}}^{t_j}h_{23}(u)du\right)$$

$$= \exp\left(-\lambda_{23}(t_j^{k_{23}} - t_{j-1}^{k_{23}})\right) \tag{6}$$

$$p_{12}\left(t_{j-1}, t_j\right) = \int_{t_{j-1}}^{t_j} \exp\left(-H_1(t_{j-1}, u)\right) h_{12}(u) \exp\left(-H_2(u, t_j)\right) du \tag{7}$$

$$p_{23}\left(t_{j-1}, t_j\right) = p_{22}\left(t_{j-1}, t_j\right) h_{23}(t_j) \tag{8}$$

The integral for $p_{12}(t_{j-1}, t_j)$ does not have a closed-form solution. The integrand can be approximated using composite Simpson's rule. If the interval (t_{j-1}, t_j) is split up with into n sub-intervals, with n being an even number, and $h = \frac{t_j - t_{j-1}}{n}$ is the length of the intervals, the composite Simpson's rule is given by:

$$\int_{t_{j-1}}^{t_j} f(u)du = \frac{h}{3} \sum_{l=1}^{n/2} \left(f(u_{2l-2}) + 4f(u_{2l-1}) + f(u_{2l})\right)$$

$$= \frac{h}{3} \left(f(u_0) + 2 \sum_{l=1}^{n/2-1} f(u_{2l}) + 4 \sum_{l=1}^{n/2} f(u_{2l-1}) + f(u_n) \right),$$

where $u_l = t_{j-1} + lh$ for $l = 0, 1, ..., n-1, n$. In fact, $u_0 = t_{j-1}$ and $u_n = t_j$.

5.3 Simulating data

Given survival up to time $u > 0$, the conditional Weibull survivor function is $S(t|u) = \exp(-\lambda(t^k - u^k))$. The cumulative distribution function $F(t|u) = 1 - \exp(-\lambda(t^k - u^k))$ can be used to simulate conditional Weibull event times by the inversion method. Replace $F(t|u)$ with $U \sim U(0, 1)$, put T for t and solve for T:

$$U = 1 - \exp\left(-\lambda(T^k - u^k)\right)$$

$$1 - U = \exp\left(-\lambda(T^k - u^k)\right)$$

$$-\log\left(1 - U\right) = \lambda T^k - \lambda u^k$$

$$\lambda T^k = -\log\left(1 - U\right) + \lambda u^k$$

$$T^k = -\frac{1}{\lambda} \log\left(1 - U\right) + u^k$$

$$T = \left(-\frac{1}{\lambda} \log\left(1 - U\right) + u^k\right)^{\frac{1}{k}}$$

$$\log(T) = \log\left(-\frac{1}{\lambda}\log(1-U) + u^k\right)^{\frac{1}{k}}$$

$$\log(T) = \frac{1}{k}\log\left(-\frac{1}{\lambda}\log(1-U) + u^k\right).$$

Then

$$T|(T > u) = \exp\left(\frac{1}{k}\log\left(-\frac{1}{\lambda}\log(1-U) + u^k\right)\right). \tag{9}$$

Implementation can be checked by setting $u = 0$ and comparing sample mean and sample variance to the theretical ones:

$$E(T) = \lambda^{-\frac{1}{k}}\Gamma\left(1 + k^{-1}\right)$$

$$Var(T) = \lambda^{-\frac{2}{k}}\left(\Gamma\left(1 + \frac{2}{k}\right) - \Gamma\left(1 + \frac{1}{k}\right)^2\right),$$

where $\Gamma(n) = (n-1)!$ is the gamma function.

5.4 Model with puberty

When adding a covariate, transition-specific hazards became as:

$$h_{rr'}(t) = h_{rr'.0}(t)\exp(\beta_{rr'}z_i(t)) = \lambda_{rr'}k_{rr'}t^{k_{rr'}-1}\exp(\beta_{rr'}z_i(t)).$$

Then, $T \sim Weibull(\lambda\exp(\beta), k)$ when puberty is "on", and $T \sim Weibull(\lambda, k)$ when puberty is "off". Let b_i stand for the age at pubertal onset and let's assume that puberty affects to the transitions during two year period $(b_i - 1, b_i + 1)$ around the onset. The hazard function becomes as

$$h_{rr'}(t) = \begin{cases} \lambda_{rr'}k_{rr'}t^{k_{rr'}-1} & 0 < t \leq b_i - 1 \\ \lambda_{rr'}\exp(\beta_{rr'})k_{rr'}t^{k_{rr'}-1} & b_i - 1 \leq t < b_i + 1 \\ \lambda_{rr'}k_{rr'}t^{k_{rr'}-1} & t \geq b_i + 1, \end{cases} \tag{10}$$

where $\exp(\beta_{rr'})$ is the relative hazard during the pubertal period.

5.4.1 Simulating data

Given survival up to time $u > 0$, the conditional Weibull survivor function is $S(t|u) = \exp(-\lambda(t^k - u^k))$ which is the same as $p(u, t)$.

- For $1- > 2$ and $1- > 3$ transitions, separately:

1. Simulate event time t_i for individual i by using Equation 9 (u=0).
   (a) If $t_i > b_i - 1$, simulate $t_{ik}|(t_{ik} > b_i - 1)$ so that $\lambda = \lambda\exp(\beta)$ and $u = b_i - 1$ in Equation 9, and replace t_i by t_{ik}.

i. If $t_{ik} > b_i + 1$, simulate $t_{il}|(t_{il} > b_i + 1)$ so that $u = b_i + 1$ in Equation 9, and replace t_{ik} by t_{il}.

(b) Put final simulated age as t_i.

2. When implemented separately for both transitions, choose first transition appearing and put $t_i = t_{i1}$ (transition age to state 2) or $t_i = t_{i2}$ (transition age to state 3).

- For $2- > 3$ transition:

1. Pick children with $1- > 2$ transition.
2. Simulate event time t_{i3} for individual i:

   (a) If $t_{i1} < b_i - 1$, use Equation 9 by putting $u = t_{i1}$.

   i. If $t_{i3} > b_i - 1$, simulate $t_{i3.1}|(t_{i3.1} > b_i - 1)$ so that $\lambda = \lambda \exp(\beta)$ and $u = b_i - 1$ in Equation 9, and replace t_{i3} by $t_{i3.1}$.

   A. If $t_{i3.1} > b_i+1$, simulate $t_{i3.2}|(t_{i3.2} > b_i+1)$ so that $u = b_i+1$ in Equation 9, and replace $t_{i3.1}$ by $t_{i3.2}$.

   ii. Put final simulated age as t_{i3}.

   (b) If $b_i - 1 \leq t_{i1} < b_i + 1$, use Equation 9 by putting $\lambda = \lambda \exp(\beta)$ and $u = t_{i1}$.

   i. If $t_{i3} > b_i + 1$, simulate $t_{i3.1}|(t_{i3.1} > b_i + 1)$ so that $u = b_i + 1$ in Equation 9, and replace t_{i3} by $t_{i3.1}$.

   ii. Put final simulated age as t_{i3}.

   (c) If $t_{i1} \geq b_i + 1$, use Equation 9 by putting $u = t_{i1}$.

5.5 Proportional hazards vs. accelerated failure time

- The Weibull distribution has a proportional hazards property:

  In 2-group case, if hazard for individual at group 1 is $h_0(t) = \lambda\gamma t^{\gamma-1}$, then the hazard for individual in group 2 is $\psi h_0(t)$. The hazard function for individual i in group 2 is

$$h_i(t) = \psi\lambda\gamma t^{\gamma-1}$$

  which is the hazard function with scale parameter $\psi\lambda$ and shape parameter γ. Survival times in both groups have Weibull distribution with shape parameter γ and the hazard of death at time t for an individual in second group is proportional to that of individual in first group.

- The Weibull distribution has a accelerated failure time property:

  Survival times are assumed to have a Weibull distribution $W(\lambda, \gamma)$, λ is scale parameter and γ shape parameter, so that the baseline hazard function is $h_0(t) = \lambda\gamma t^{\gamma-1}$.

According to the general accelerated failure time model, the hazard function for ith individual is then given by

$$h_i(t) = e^{-\eta_i}\lambda\gamma(e^{-\eta_i}t)^{\gamma-1} = (e^{-\eta_i})^{\gamma}\lambda\gamma t^{\gamma-1}$$

so that the survival time of the individual has a $W(\lambda e^{-\gamma\eta_i}, \gamma)$ distribution. η_i stands for the linear component of the model.

If the baseline hazard function is the hazard function of a $W(\lambda, \gamma)$ distribution, the survival times under

- the general proportional hazards model have a $W(\lambda e^{\boldsymbol{\beta}'\boldsymbol{x}_i}, \gamma)$ distribution
- the accelerated failure time model have a $W(\lambda e^{-\gamma\boldsymbol{\alpha}'\boldsymbol{x}_i}, \gamma)$ distribution.

Then, it follows that the β-coefficients of the proportional hazards model can be produced from the accelerated failure time model by multiplying the α-coefficients of the accelerated failure time model by $-\gamma$.

6 Omia juttuja

- Sopivaa Weibull-jakaumaa voi tutkia log-cumulative hazard plot -kuvion avulla. Weibull-jakauman survivor-funktio on

$$S(t) = \exp(-\lambda t^{\gamma})$$

Ja log-cumulative hazard

$$S(t) = \log(-\log S(t)) = \log\lambda + \gamma\log t$$

Jos sijoitetaan $S(t)$:n paikalle Kaplan-Meier estimaatti $\hat{S}(t)$ ja tehdÃÃn log-cum hazard kuvio, sen pitÃisi antaa lÃhes suora viiva. TÃllÃ¶in viivan interceptin eksponentti on scale-parametrin ja slope shape-parametrin estimaatti. Jos slope (shape-parametri) on lÃhellÃ ykkÃ¶stÃ, survival timet saattavat noudattaa exponentiaalista jakaumaa.

-> Voisi valita simulointia varten sopivat jakaumat oikean aineiston perusteella

7 Comparison between continuous and panel-observed

Let's compare transition probabilities $1- > 2$ between continuous and panel-observed data when follow-up frequency draw near 0 years: $t_{j-1}- > t_j$. Transition time $t_{i,1}$ at continuously observed data and transition time t_j at panel-observed data denote the same age.

Transition probabilities are

1. $p_{12}(0, t_{i,1}) = p_{11}(0, t_{i,1})\, h_{12}(t_{i,1})$ (continuously observed)

2. $p_{11}(0, t_{j-1})p_{12}(t_{j-1}, t_j)$ (panel-observed),

   where $p_{12}(t_{j-1}, t_j) = \int_{t_{j-1}}^{t_j} p_{11}\,(t_{j-1}, u)\, h_{12}(u) p_{22}\,(u, t_j)\, du$

   $= \int_{t_{j-1}}^{t_j} \exp(-(H_{12}(u) - H_{12}(t_{j-1}) + H_{13}(u) - H_{13}(t_{j-1}))) h_{12}(u) \exp(-(H_{23}(t_j) - H_{23}(u)))du$

   $- > \int_{t_{j-1}}^{t_j} h_{12}(u)$, when $t_{j-1} - > t_j$

   $\int_{t_{j-1}}^{t_j} h_{12}(u) = H_{12}(t_j) - H_{12}(t_{j-1}) - > h_{12}(t_j)$

   because it's the difference between the cumulative hazard functions between two measured time points (difference in one unit) and hazard stays same until it changes at next measurement right after time t_j.