

Dynamic prediction modelling in hand disorders after stroke using a latent class multivariate mixed model

Eleni-Rosalina Andrinopoulou, Ruud Selles, Rinske Nijland,
Carel Meskers, Gerard Ridders, Gert Kwakkel, Dimitris
Rizopoulos

International Society for Clinical Biostatistics, 14-18 July, 2019

Clinical Application

Data set collected Amsterdam

→ Patients measured after a stroke

Outcome of interest:

The Action Research Arm Test (ARAT) is a measure used by physical therapists and other health care professionals to assess upper extremity performance

Clinical Application: Data Details

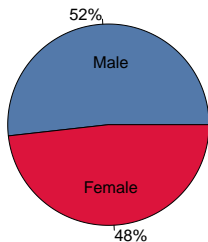
Number of patients:

450

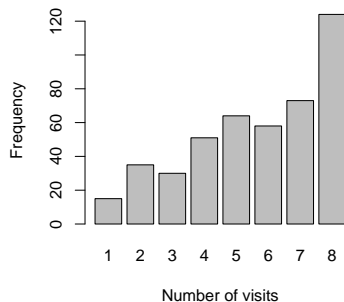
Mean age at stroke:

65

Gender:



Follow-up visits:



Clinical Application: Data Details (cont'd)

Clinical Application: Data Details (cont'd)

Clinical Application: Research Question

Guide clinical decision making → use **complete** biomarker information.

Can we utilize all available longitudinal measurements to predict the future ARAT measurements?

GemsTracker

Statistical Analysis

Special feature should be taken into account in longitudinal data

- Correlation between measurements obtained from the same patients
- Biological variation of the outcome
- Unbalanced datasets

Mixed-effects models

Let y_i represent the repeated measurements of an outcome for the i -th patient, $i = 1, \dots, n$

$$y_i(t) = x_i^\top(t)\beta + z_i^\top(t)b_i + \epsilon_i(t),$$

$$b_i \sim N(0, D),$$

$$\epsilon_i(t) \sim N(0, \sigma_i^2),$$

where

- ◇ $x_i^\top(t)\beta$ denotes the fixed part
- ◇ $z_i^\top(t)b_i$ denotes the random part

Sub-populations (1)

Time-dependent covariates (2)

Statistical Analysis: Sub-populations

Challenge (1)

Challenge (1)

Latent class models

$$y_i(t|c_i = \textcolor{red}{g}) = x_i^\top(t)\beta_{\textcolor{red}{g}} + z_i^\top(t)b_{i\textcolor{red}{g}} + \epsilon_i(t),$$

$$b_{i\textcolor{red}{g}} \sim N(0, D_{\textcolor{red}{g}}),$$

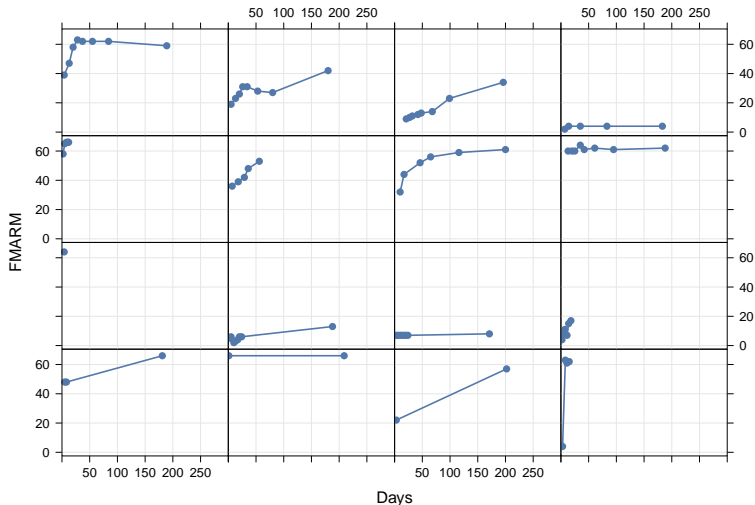
$$\epsilon_i(t) \sim N(0, \sigma_i^2),$$

$$Pr(c_i = \textcolor{red}{g}) \sim Dirichlet(A_c),$$

where

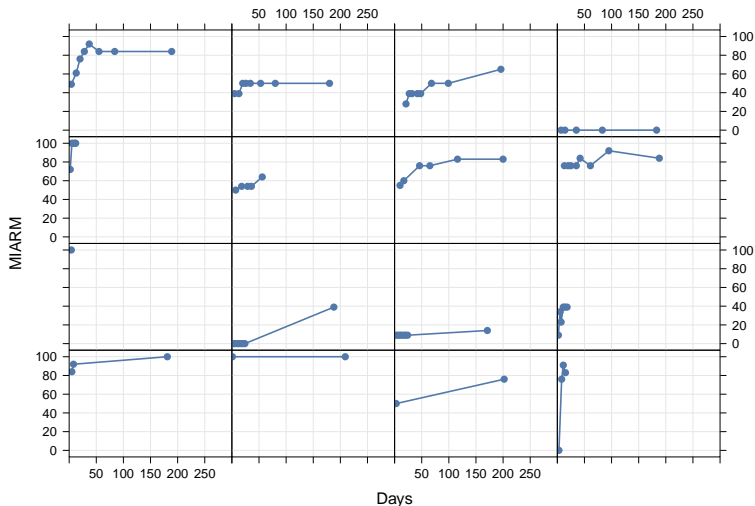
- ◇ $x_i^\top(t)\beta$ denotes the fixed part
- ◇ $z_i^\top(t)b_i$ denotes the random part
- ◇ $\textcolor{red}{g}$ indicates the class

Statistical Analysis: Time-dependent Challenge (2)



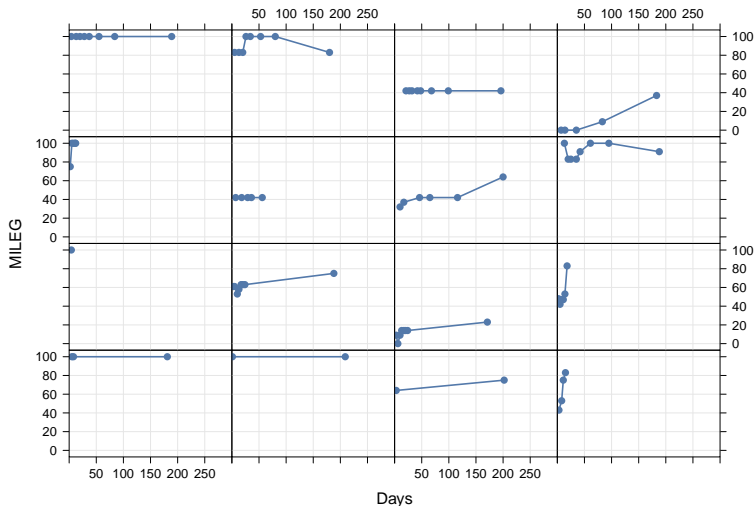
Statistical Analysis: Time-dependent (cont'd)

Challenge (2)



Statistical Analysis: Time-dependent (cont'd)

Challenge (2)



Statistical Analysis: Time-dependent (cont'd)

Challenge (2)

Multivariate model (k longitudinal outcomes)

$$h_k[E\{y_{ki}(t) \mid c_i = g\} \mid b_{kig}] = x_{ki}^\top(t)\beta_{kg} + z_{ki}^\top(t)b_{kig},$$

$$b_{ig} = (b_{i1g}^\top, \dots, b_{iKg}^\top) \sim N(0, D_g),$$

- ◇ $x_{ki}^\top(t)\beta_{kg}$ denotes the fixed part
- ◇ $z_{ki}^\top(t)b_{kig}$ denotes the random part
- ◇ $h_k(\cdot)$ denotes the link function and g indicates the class

Fixed Effects

Nonlinear time in days (with 3 knots)

Shoulder abduction

Finger extension

Recombinant tissue plasminogen activator (medication)

Neglect (lack of awareness of the recovering side)

Random Effects

Nonlinear time in days (with 3 knots)

2 classes

Statistical Analysis: Model Specification - MIARM, MILEG, FMARM

Fixed Effects

Nonlinear time in days (with 3 knots)

Random Effects

Nonlinear time in days (with 3 knots)

2 classes

Statistical Analysis: Results

Check the fitting of the model

Prediction

Prediction: ARAT data set

Using the proposed multivariate latent class mixed model

Prediction: Results

Assess the performance of the proposed model → **Important**

- ◇ Univariate mixed model (1 class)
- ◇ Multivariate mixed model (2 classes)

Prediction: Performance (cont'd)

Assess the performance of the proposed model:

→ Different methods and metrics exist (e.g. Mean absolute error)

→ Proper scoring rules

- ◇ Compare the predictive **distribution** of the outcome with the observed value

Logarithmic scoring rule

$$LR = \log[f_{y_{pred}}(y_{obs})],$$

where $f_{y_{pred}}$ is the predictive density at the observed values

→ Proper scoring rules

- ◇ Compare the predictive **distribution** of the outcome with the observed value

Continuous ranked probability score

$$CRPS = \int [P_{y_{pred}}(x) - P_{y_{obs}}(x)]^2 dx,$$

where $P_{y_{pred}}$ and $P_{y_{obs}}$ are the cumulative distribution function of the prediction and the observation respectively

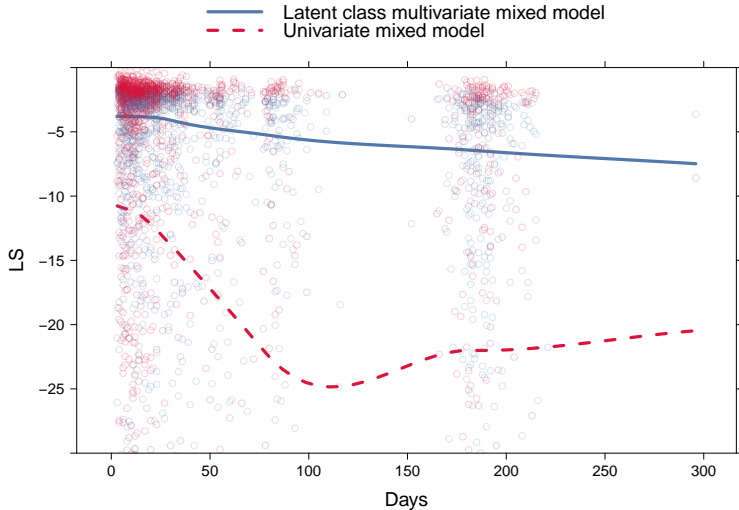
→ Cross-validation

- ◇ we split the data into 10 parts
- ◇ use 9 for fitting and 1 for predicting

predicting data: use 1 observation to predict the rest

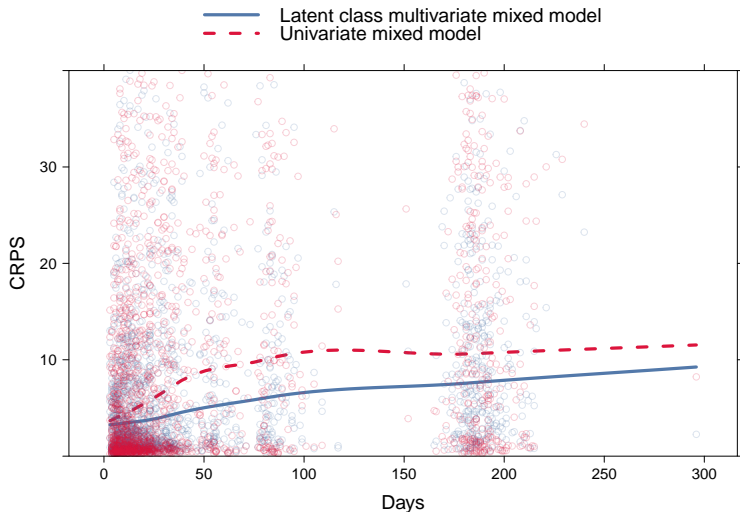
Prediction: Performance (cont'd)

Logarithmic scoring rule



Prediction: Performance (cont'd)

Continuous ranked probability score



Latent class multivariate mixed model

Future work

- ◇ More classes
- ◇ Extra outcomes
- ◇ Proper scoring rules

Thank you for your attention!