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

@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl

Clinical Application

Clinical Application: Motivation

Data set collected in Amsterdam

➔ Patients followed after 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

@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl

Clinical Application: Data Details

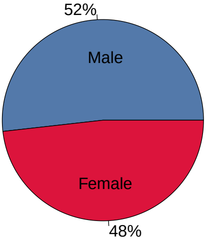
Number of patients:

450

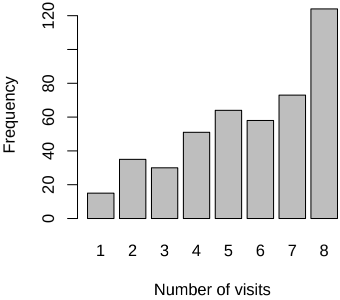
Mean age at stroke:

65

Gender:

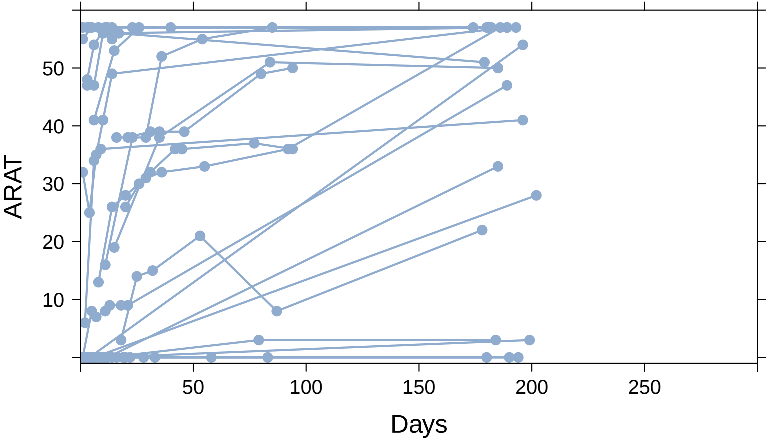


Follow-up visits:

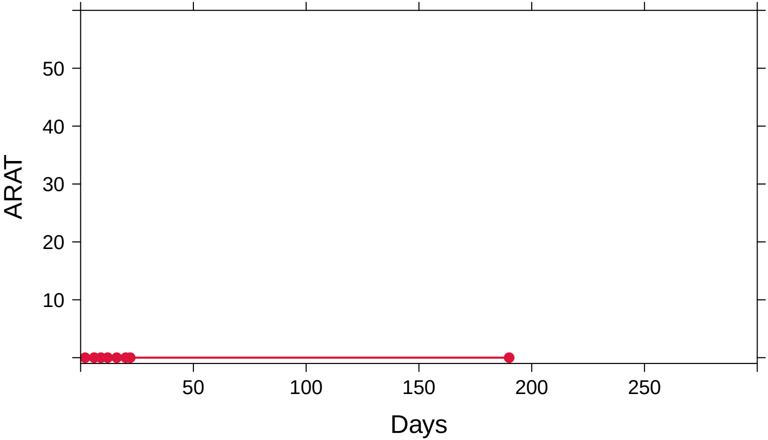


@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl

Clinical Application: Data Details (cont'd)



Clinical Application: Data Details (cont'd)

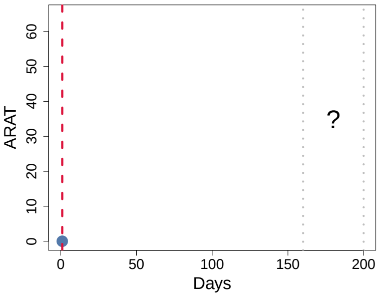


Clinical Application: Research Question

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

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

GemsTracker



Statistical Analysis

Statistical Analysis: Data Characteristics

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

@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl

Statistical Analysis: Mixed-effects models

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

$$y_i(t) = x_i^\top(t)\beta + z_i^\top(t)b_i + \epsilon_i(t),$$
$$b_i \sim N(0, D),$$
$$\epsilon_i(t) \sim N(0, \sigma_i^2),$$

where

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

Statistical Analysis: Challenges

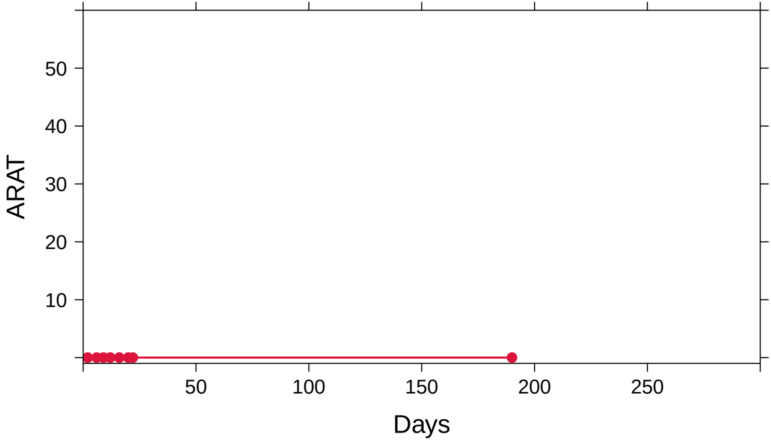
(1) **Sub-populations**

(2) **Time-dependent covariates**

@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl

Statistical Analysis: Sub-populations

Challenge *(1)*



Statistical Analysis: Sub-populations

Challenge *(1)*

Latent class models

$$y_i(t|c_i = g) = x_i^\top(t)\beta_g + z_i^\top(t)b_{ig} + \epsilon_i(t),$$

$$b_{ig} \sim N(0, D_g),$$

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

$$Pr(c_i = g) \sim Dirichlet(A_c),$$

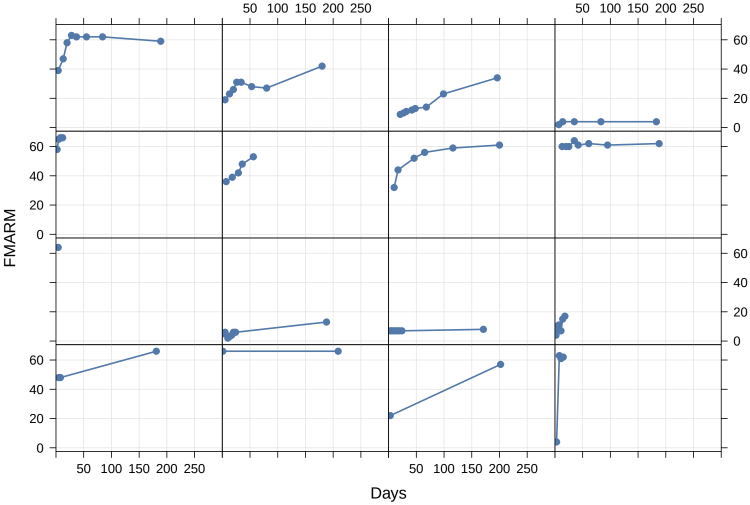
where

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

@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl

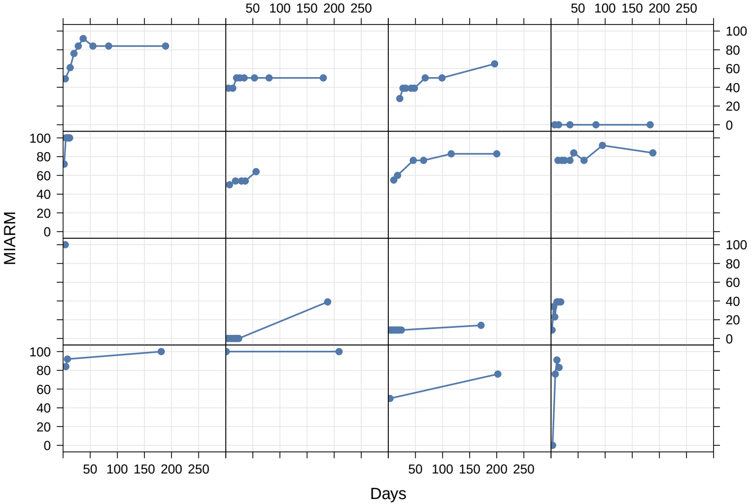
Statistical Analysis: Time-dependent

Challenge *(2)*



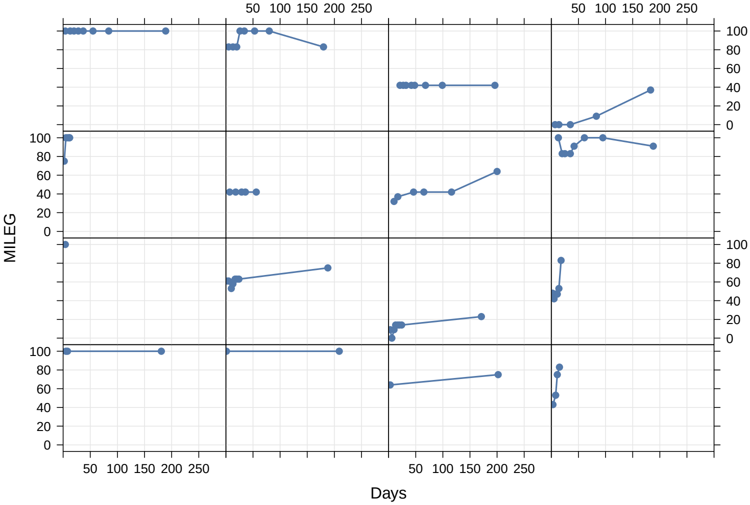
Statistical Analysis: Time-dependent (cont'd)

Challenge *(2)*

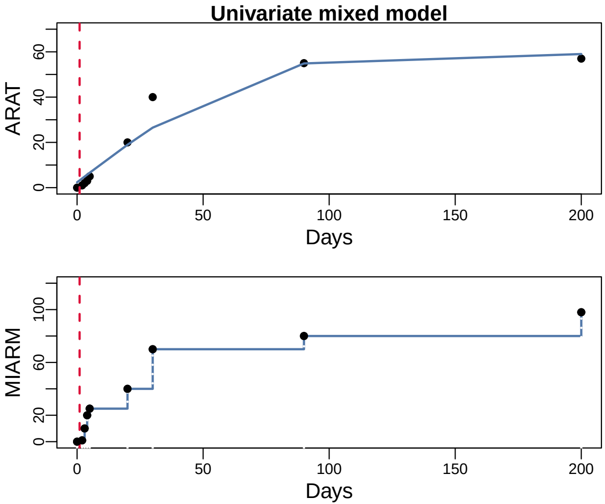


Statistical Analysis: Time-dependent (cont'd)

Challenge *(2)*



Statistical Analysis: Time-dependent (cont'd)



Statistical Analysis: Time-dependent (cont'd)

Challenge *(2)*

Multivariate model (k longitudinal outcomes)

$$h_k[E\{y_{ki}(t \mid c_i = g)|b_{kig}\}] = x_{ki}^\top(t)\beta_{kg} + z_{ki}^\top(t)b_{kig},$$
$$b_{ig} = (b_{i1g}^\top, \ldots, b_{iKg}^\top) \sim N(0, D_g),$$

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

Statistical Analysis: Model Specification - ARAT

Bayesian framework

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)

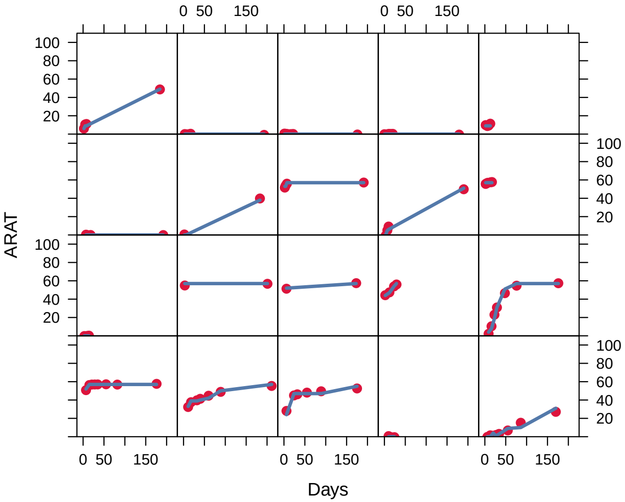
Random Effects

Nonlinear time in days (with 3 knots)

2 classes

Statistical Analysis: Results

Check the fitting of the model



Prediction

@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl

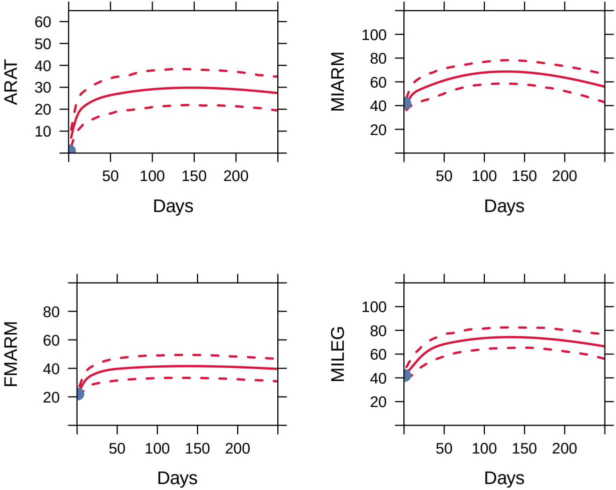
Prediction: ARAT data set

Predictions using the proposed latent class multivariate mixed model

Monte Carlo simulation scheme

- Draw parameters from the MCMC
- Draw b_{ig} from the posterior
- Calculate predictions

Prediction: Results



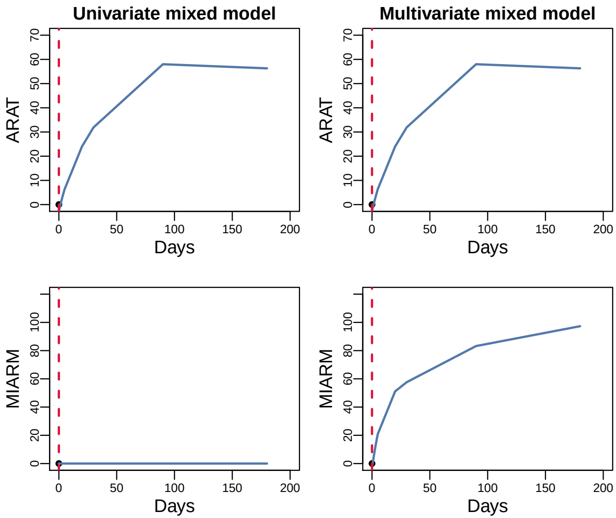
Prediction: Performance

Assess the performance of the proposed model $\rightarrow$ **Important**

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

@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl

Prediction: Performance (cont'd)



@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl

Prediction: Performance (cont'd)

Assess the performance of the proposed model:

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

Prediction: Performance (cont'd)

➔ **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

@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl

Prediction: Performance (cont'd)

➔ **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 disctribution function of the prediction and the observation respectively

Prediction: Performance (cont'd)

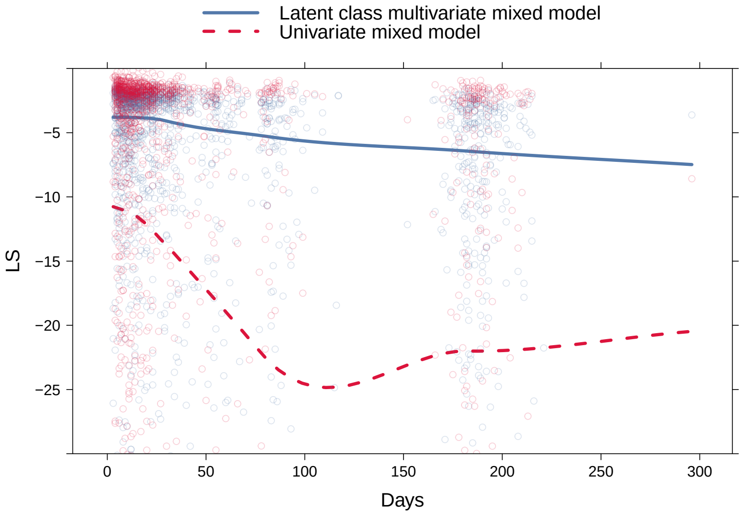
➔ **Cross-validation**

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

predicting data: use 1 observation to predict the rest

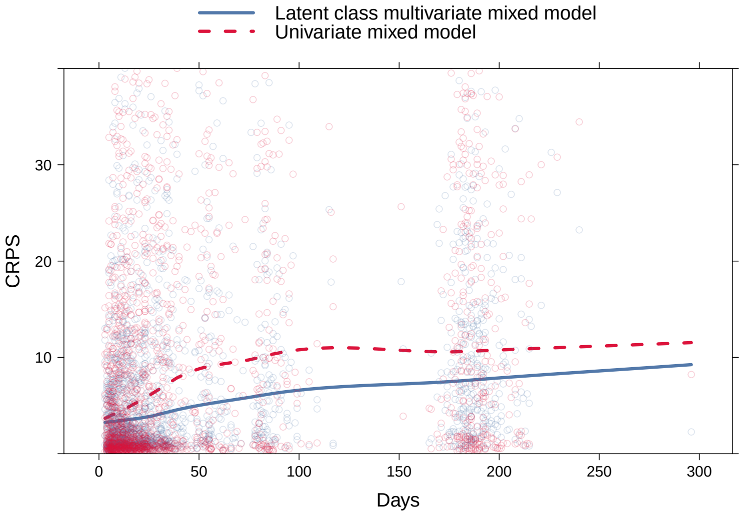
Prediction: Performance (cont'd)

Logarithmic scoring rule



Prediction: Performance (cont'd)

Continuous ranked probability score



Conclusion

Latent class multivariate mixed model

Future work

- More classes
- Extra outcomes
- Proper scoring rules

@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl

Thank you for your attention!

The slides are available at: https://www.erandrinopoulou.com

@ERandrinopoulou https://github.com/erandrinopoulou eandrinopoulou@erasmusmc.nl