Estimating Cystic Fibrosis Lung Function Decline: An Empirical Study

Eleni-Rosalina Andrinopoulou, on behalf of the Research Methods of Calculating Lung Function Decline Workgroup committee: Cystic Fibrosis Foundation, Cincinnati Children's Hospital Medical Center and Erasmus MC

North American Cystic Fibrosis Conference, October, 2020

www.erandrinopoulou.com eandrinopoulou@erasmusmc.nl @ERandrinopoulou

Disclosure

I have no actual or potential conflict of interest in relation to this presentation.

Introduction

Introduction: Data set

US Cystic Fibrosis Foundation Patient Registry (CFFPR) 2003-2016

➔ Included: 35,252 patients

Introduction: Data set

US Cystic Fibrosis Foundation Patient Registry (CFFPR) 2003-2016

➔ Included: 35,252 patients

➔ Excluded: patients with missing pulmonary function (808 patients)

➔ Excluded: aged $<$ 6 years (137 patients)

➔ Excluded: lung transplant prior to 2003 (580 patients)

Total of 33,727 patients with 1,276,456

Introduction: Goal

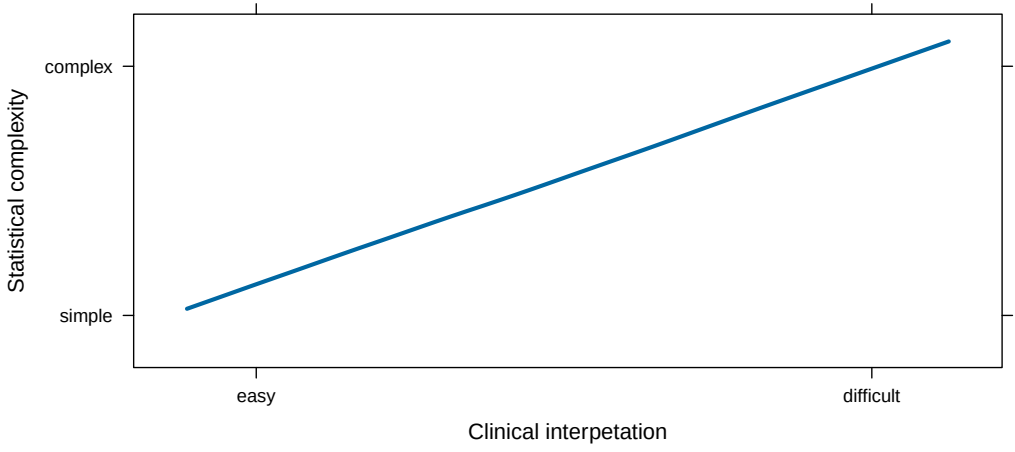
Variability in analytic approaches to model FEV_1 decline

Introduction: Goal

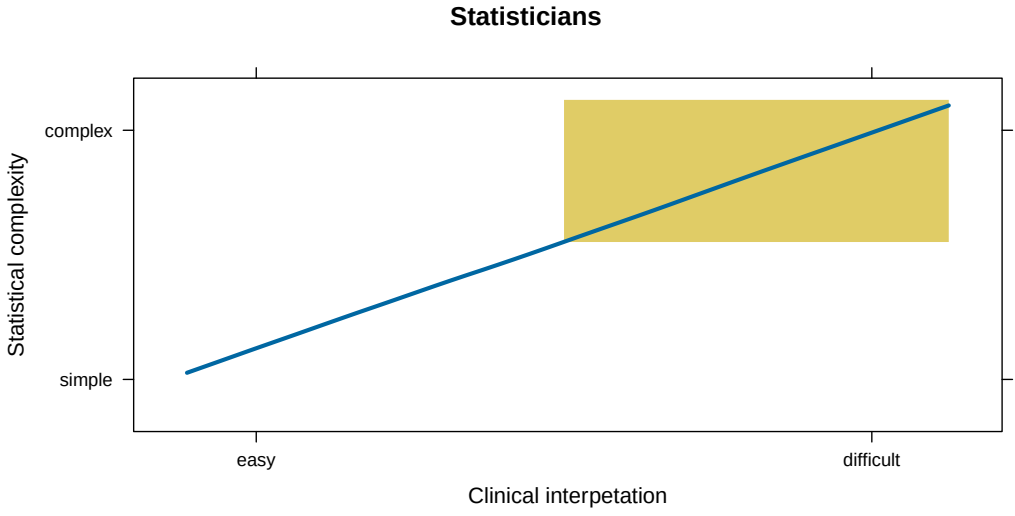
Variability in analytic approaches to model FEV_1 decline

- ➔ Rationale for selected statistical methods
- ➔ Differences between models
- ➔ Differences between scenarios
  - ◇ small sample size
  - ◇ smaller follow-up period

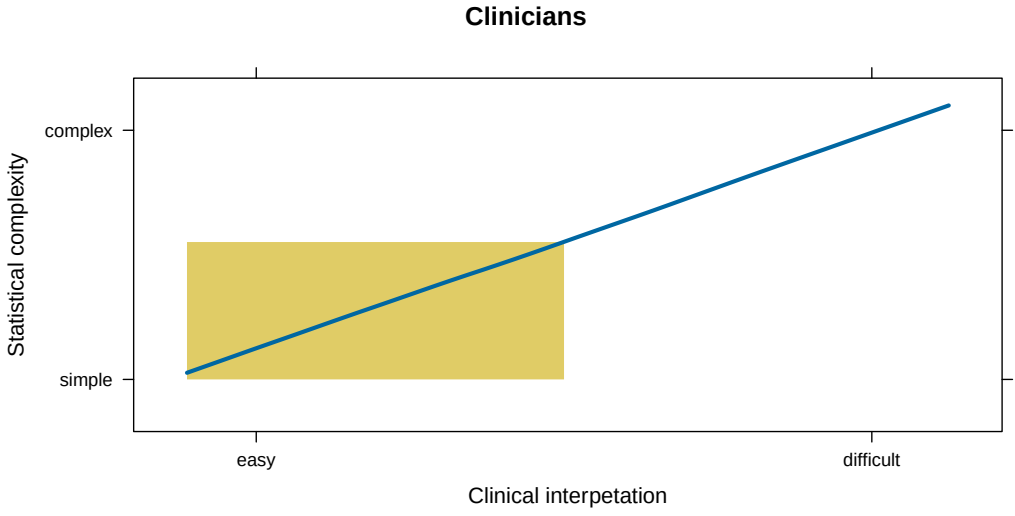
Introduction: Challenge



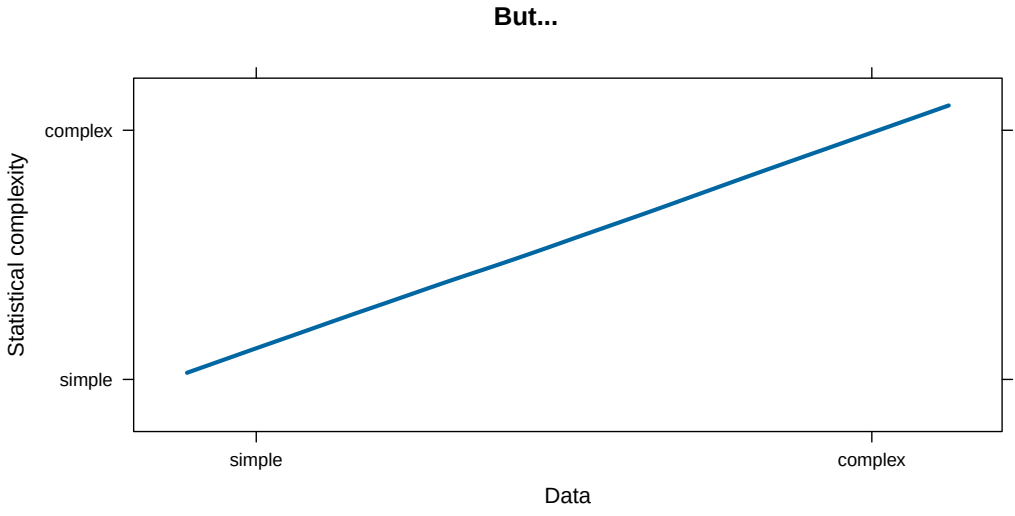
Introduction: Challenge



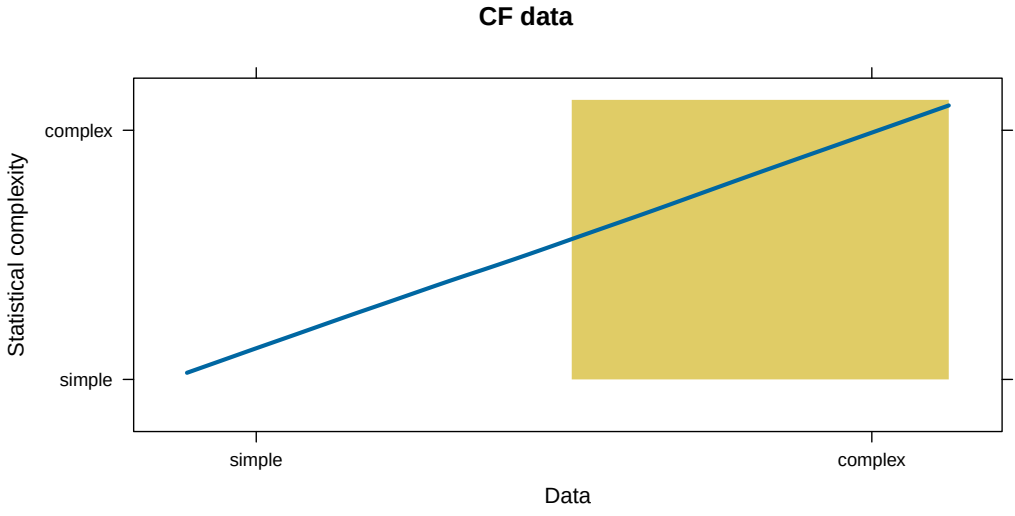
Introduction: Challenge



Introduction: Challenge

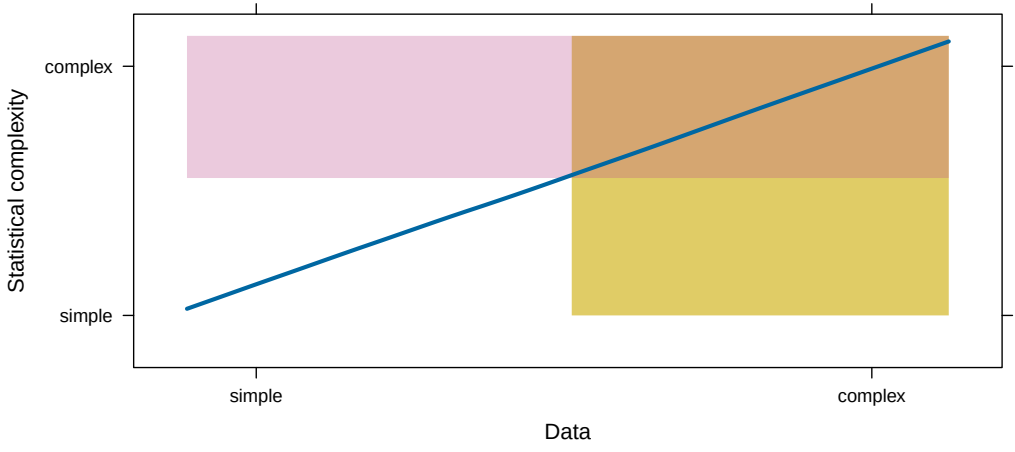


Introduction: Challenge



Introduction: Challenge

Therefore...



www.erandrinopoulou.com eandrinopoulou@erasmusmc.nl @ERandrinopoulou

Methods

Methods: Features

Characteristics of FEV_1

- ➔ Multiple measurements within the same patient
- ➔ Unbalanced design
- ➔ Measurement error

Methods: Complex statistical models

- ➔ Mixed effects models (e.g. random effects structure)
- ➔ Generalized estimating equation (e.g. correlation structure)
- ➔ Joint models of longitudinal and survival data

Methods: Keep in mind!

Each data set is unique

➔ There is no One Model Fits All Needs solution

Methods: Mixed models

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

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

where
$\epsilon_i(t) \sim N(0, \sigma^2)$
$b_i \sim N(0, D)$

Methods: Mixed models

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

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

where
$\epsilon_i(t) \sim N(0, \sigma^2)$
$b_i \sim N(0, D)$

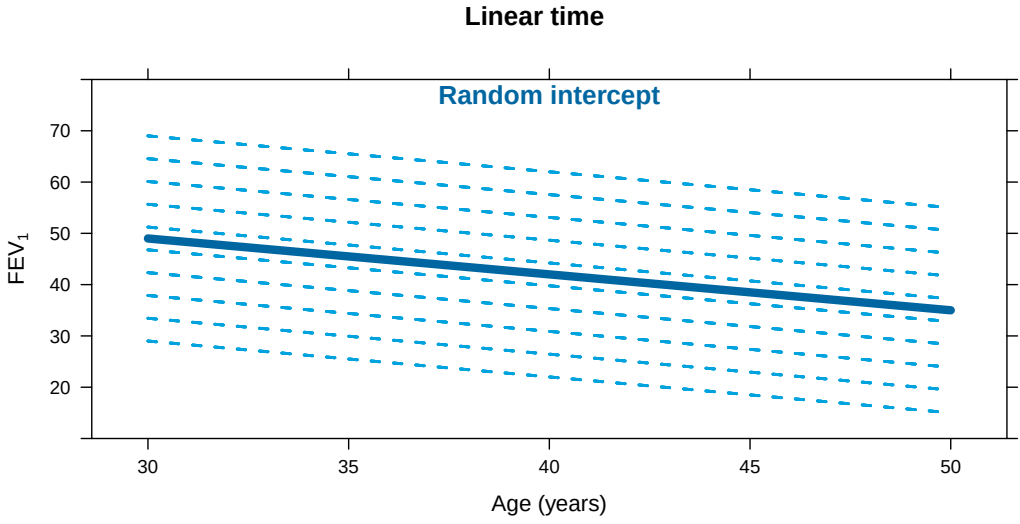
Methods: Mixed models

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

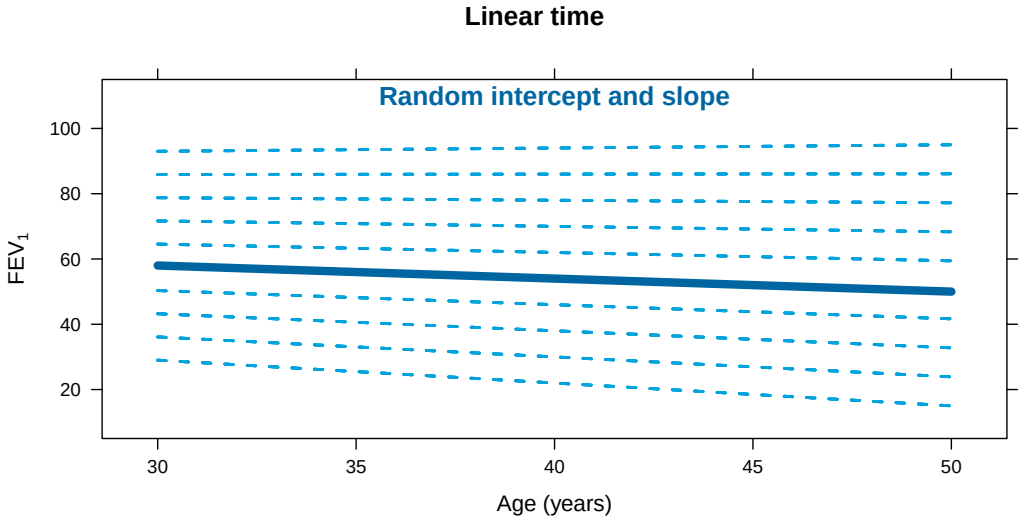
$$y_i(t) = x_i^\top(t)\beta + z_i^\top(t)b_i + \epsilon_i(t),$$

where
$\epsilon_i(t) \sim N(0, \sigma^2)$
$b_i \sim N(0, D)$

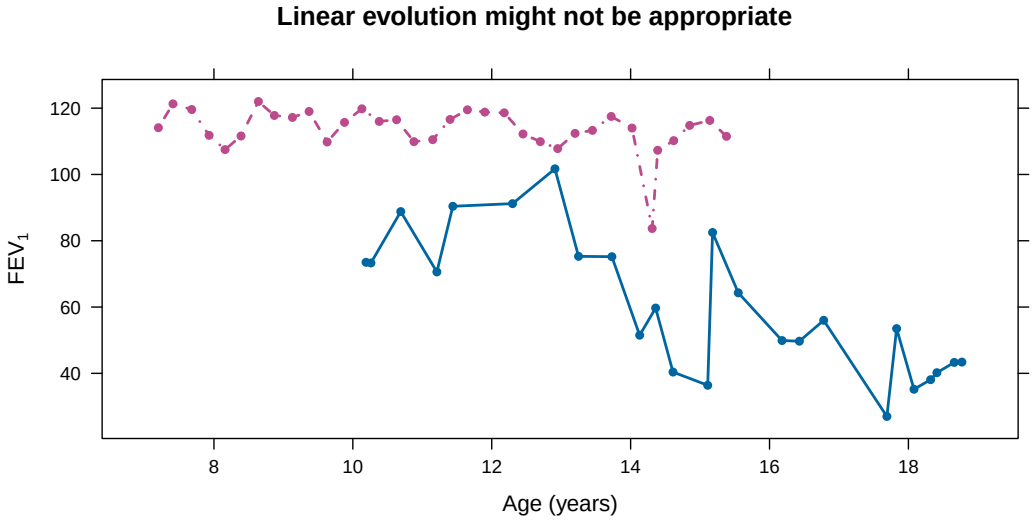
Methods: Mixed models



Methods: Mixed models



Methods: Nonlinear effects



Methods: Nonlinear effects

Assume polynomials

Methods: Nonlinear effects

Assume polynomials

Even better, assume splines!

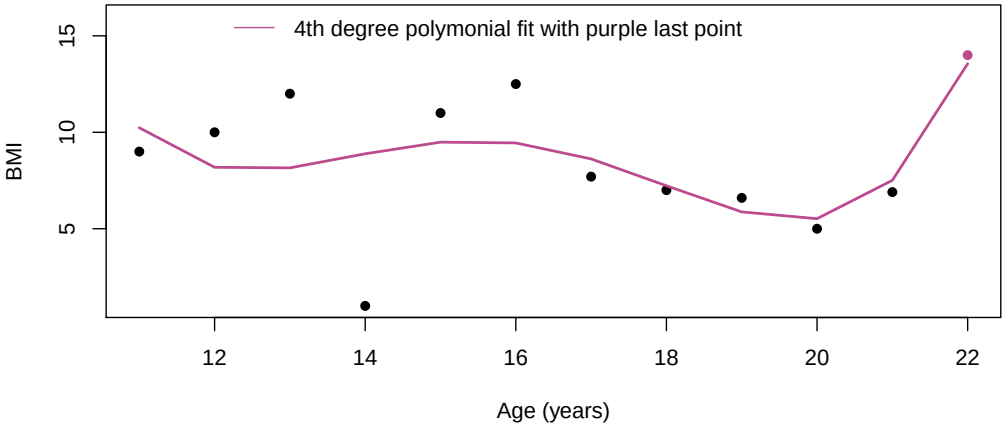
Methods: Nonlinear effects

Assume polynomials

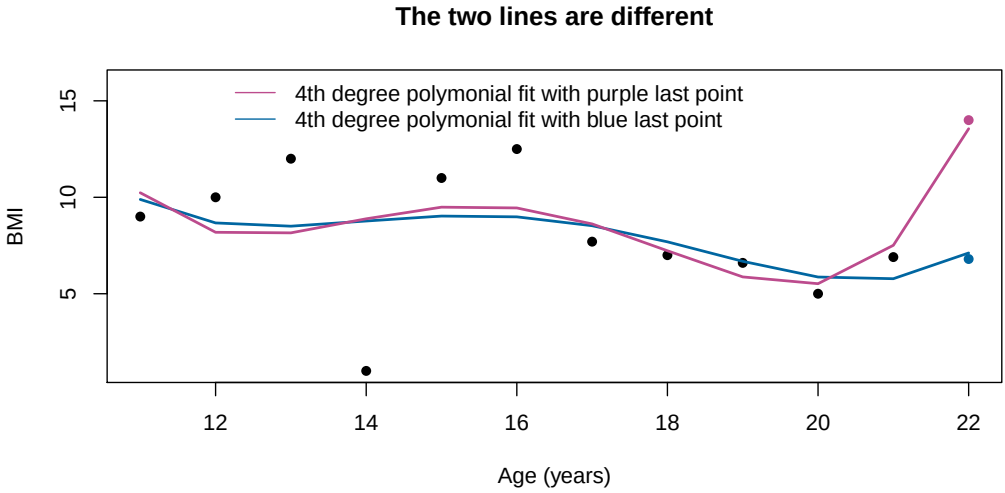
Even better, assume splines!

☹ **But... Multiple not interpretable coefficients**

Methods: Nonlinear effects

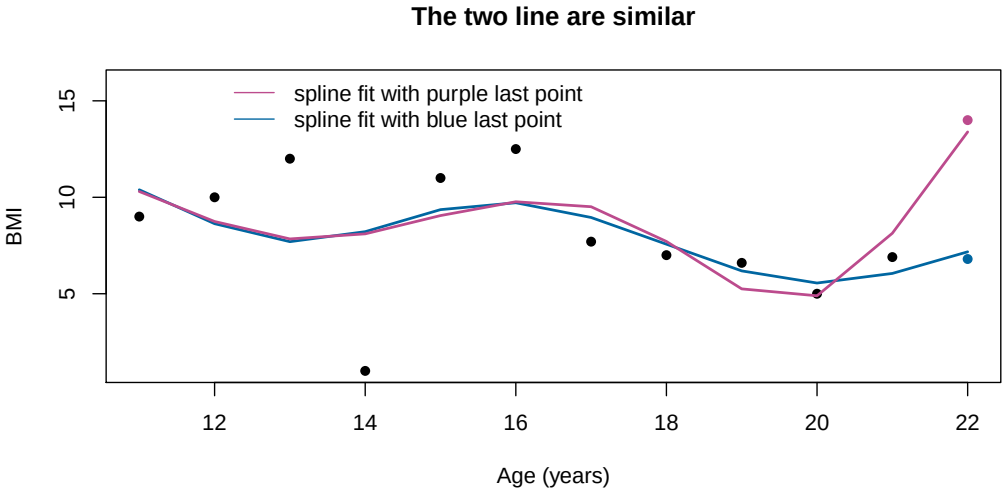


Methods: Nonlinear effects



Methods: Nonlinear effects

Methods: Nonlinear effects



Methods: Generalized estimating equation

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

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

where
$\epsilon_i(t) \sim N(0, V_i)$

➔ exponential covariance pattern for V_i

Methods: Mixed models with corr structure

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

$$y_i(t) = x_i(t)\beta + b_{0i} + \epsilon_i(t),$$

where
$\epsilon_i(t) \sim N(0, V_i)$
$b_i \sim N(0, D)$

➔ exponential covariance pattern for V_i

Taylor-Robinson, D., Whitehead, M., Diderichsen, F., Olesen, H. V., Pressler, T., Smyth, R. L., & Diggle, P. (2012). Understanding the natural progression in FEV1 decline in patients with cystic fibrosis: a longitudinal study. Thorax, 67(10), 860-866.

Methods: Joint Models

Other factors might influence the FEV_1 evolution

- ➔ Death
- ➔ Transplantation
- ➔ PEx

Model jointly!

Methods: Joint Models

Joint Models for Longitudinal and Time-to-Event Data

➔ Step 1: Fit a mixed-effects model

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

where

- $m_i(t)$: underlying value of longitudinal outcome

Methods: Joint Models

Joint Models for Longitudinal and Time-to-Event Data

➔ Step 2: Fit a survival model

$$h_i(t) = h_0(t) \exp[w_i\gamma + m_i(t)\ \alpha],$$

where

- $m_i(t)$ underlying value of longitudinal outcome
- α quantifies the strength of the association between the marker and the risk of an event
- w_i baseline covariates

Methods: Joint Models

Focus:

- ➔ On the longitudinal outcome
- ➔ On the survival outcome

Methods: Scenarios

Data conditions:

- ➔ Overall CFFPR data
- ➔ Varying the number of patients, e.g. small: center-based 150, medium: national registry, 3000 large: US registry, 30,000).
- ➔ Impact of follow-up on estimating FEV_1 decline, e.g. $<$2 years, 2-5 years, $>$5 years
- ➔ FEV_1 collection frequency, e.g. annual maximum, quarterly mean, quarterly maximum, encounter-level
- ➔ Impact of PEx in estimating FEV_1 decline, e.g. include or exclude FEV1 measurements taken during a PEx; include or exclude PEx as a rolling covariate

Methods: Model selection

Within each statistical approach

- ➔ Mixed models $\rightarrow$ AIC
- ➔ Joint models $\rightarrow$ AIC
- ➔ Generalized estimating equations $\rightarrow$ QIC

Methods: Model selection

Within each statistical approach

- ➔ Mixed models → AIC
- ➔ Joint models → AIC
- ➔ Generalized estimating equations → QIC

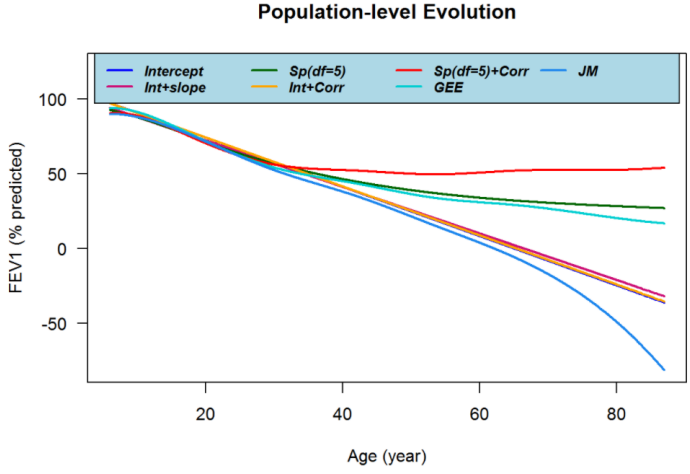
Between methods

- ➔ Mean Absolute Deviation
- ➔ Root Mean Square Error
- ➔ Correlation

Results

Results: Evolution over time

Using all data



Results: Within models

Using all data

Linear VS nonlinear:

- ➔ All models indicated an approximately linear rate of decline until age 30
- ➔ Nonlinear models fit better than linear models

Results: Within models

Using all data

Linear VS nonlinear:

- ➔ All models indicated an approximately linear rate of decline until age 30
- ➔ Nonlinear models fit better than linear models

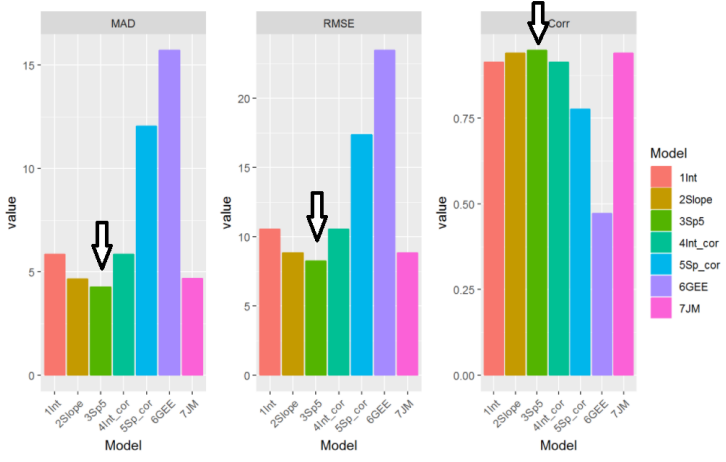
Mixed models - lowest AIC:

- ➔ Random intercept
- ➔ Correlation structure

Results: Between models

Using all data

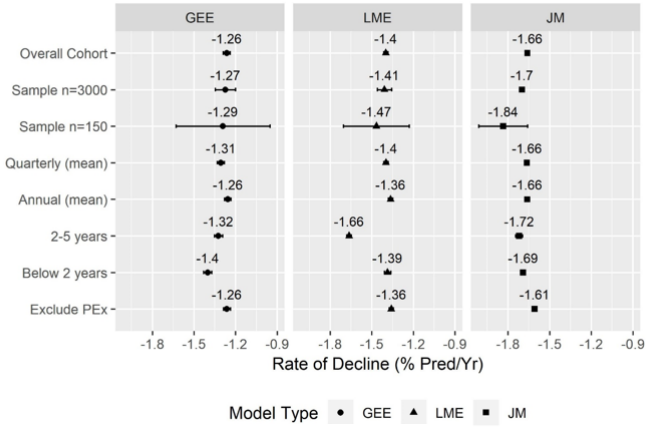
Not corrected for overoptimism



Results: Between models

Scenarios

Estimated rate of FEV1 decline assuming linear progression



www.erandrinopoulou.com eandrinopoulou@erasmusmc.nl @ERandrinopoulou

Discussion

www.erandrinopoulou.com eandrinopoulou@erasmusmc.nl @ERandrinopoulou

Discussion

Methods

- ➔ Overview of statistical analysis
- ➔ Different scenarios

Discussion

Methods

- ➔ Overview of statistical analysis
- ➔ Different scenarios

Result highlights

- ➔ Non linear evolution (declines more rapidly at earlier ages)
- ➔ Impact were similar across scenarios expect for lenght of f.u. and sample size

Thank you for your attention!

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

The presenter is supported by the National Institutes of Health / National Heart, Lung and Blood Institute (grant R01 HL141286)

www.erandrinopoulou.com eandrinopoulou@erasmusmc.nl @ERandrinopoulou