Unveiling the Impact of Social and Environmental Determinants of Health on Lung Function Decline in Cystic Fibrosis through Data Integration using the US Registry

International Society for Clinical Biostatistics

Eleni-Rosalina Andrinopoulou, Emrah Gecili, Rhonda Szczesniak

July, 2024

www.erandrinopoulou.com e.andrinopoulou@erasmusmc.nl @ERandrinopoulou

Introduction: Motivation

A lot of information is available

➔ Electronic medical records

Introduction: Motivation

A lot of information is available

➔ Electronic medical records

Cystic Fibrosis

➔ genetic disorder affecting the lungs, pancreas, and other organs

➔ $>$ 75 percent of people with CF are diagnosed by age 2

US Cystic Fibrosis Registry

- ◇ $>$23,000 patients
- ◇ $>$1,400,000 observations (on average $>$10 years of follow-up)

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

www.erandrinopoulou.com e.andrinopoulou@erasmusmc.nl @ERandrinopoulou

Introduction: Motivation

A lot of information is available

➔ Electronic medical records

Different types of information

- ➔ Baseline characteristics: Sex, F508del, SESlow, Enzymes
- ➔ Biomarkers: FEV_1 % pred
- ➔ Nutritional status: BMI percentile
- ➔ Social and environmental determinants: Deprivation index

Deprivation index

➔ Socioeconomic variables from the American Community Survey (ACS): capture "community deprivation"

- ◇ Principal components analysis of six different 2015 ACS measures
- ◇ "Deprivation Index": the first component explains over 60% of the total variance
- ◇ Rescaling and normalizing forces the index to range from 0 to 1, with a higher index being more deprived

Cole Brokamp, Andrew F. Beck, Neera K. Goyal, Patrick Ryan, James M. Greenberg, Eric S. Hall. Material Community Deprivation and Hospital Utilization During the First Year of Life: An Urban Population-Based Cohort Study. Annals of Epidemiology. 30. 37-43. 2019

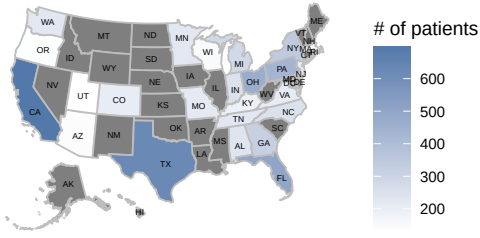
https://geomarker.io/dep_index/

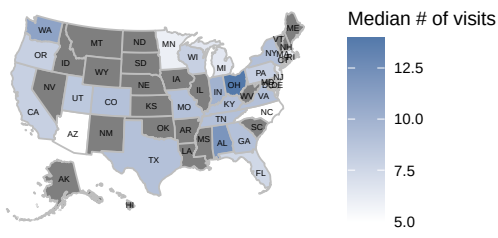
Introduction: Research question

Can we integrate registry data with social and environmental determinants of health to improve the accuracy of disease progression prognostication?

Introduction: Descriptive statistics

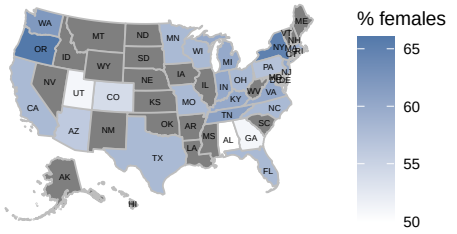
Baseline information

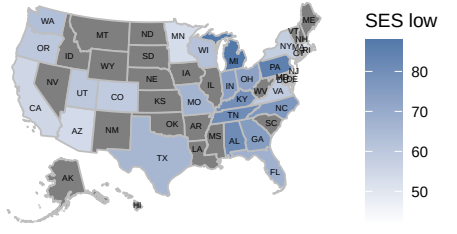




Introduction: Descriptive statistics

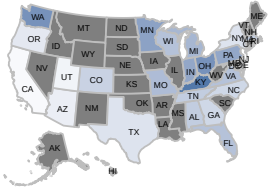
Baseline information



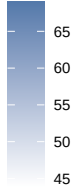


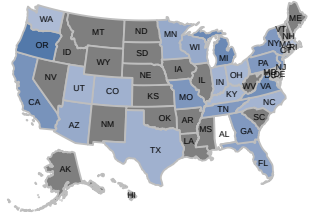
Introduction: Descriptive statistics

Baseline information

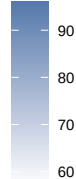


F508del Homozygous



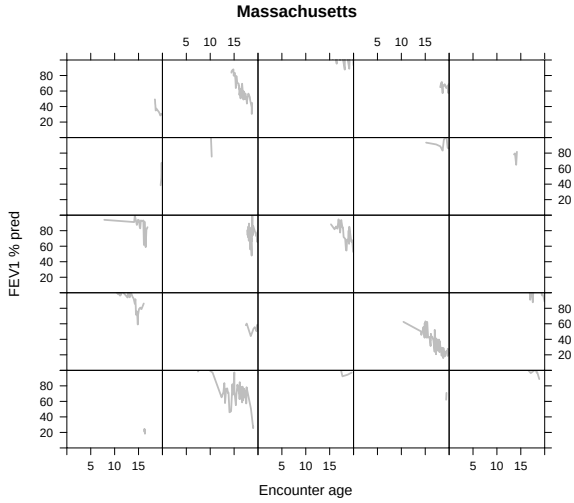


Enzymes Yes



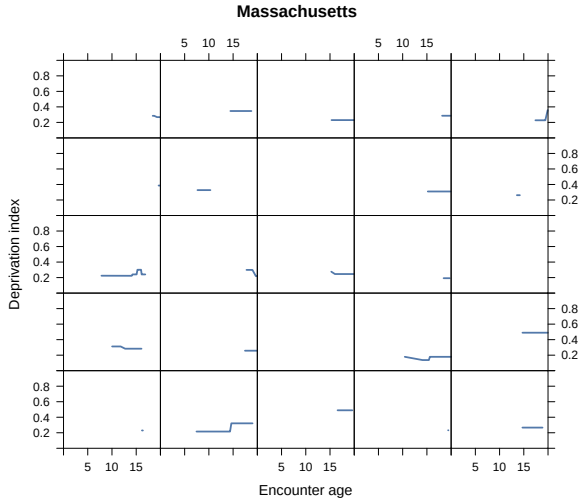
Introduction: Descriptive statistics

Lung function



Introduction: Descriptive statistics

Social and environmental determinants of health



Methods: Multivariate Mixed Models

$$\texttt{DepIndex:} y_{2i}(t) = x_{2i}^{\top}(t)\beta_1 + z_{2i}(t)^{\top} b_{2i} + \epsilon_{2i}(t)$$
$$\texttt{FEV1\%pred:} y_{1i}(t) = x_{1i}^{\top}(t)\beta_1 + z_{1i}(t)^{\top} b_{1i} + \epsilon_{1i}(t)$$

where

- $b_i^{\top} = (b_{1i}^{\top}, b_{2i}^{\top}) \sim N(0, D)$

Cannot directly measure the strength of the association and lack clinical relevance

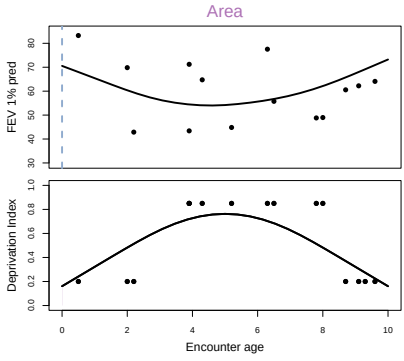
Methods: Multivariate Mixed Models

$$\texttt{DepIndex:} y_{2i}(t) = m_{2i}(t) + \epsilon_{2i}(t) = x_{2i}^{\top}(t)\beta_1 + z_{2i}(t)^{\top} b_{2i} + \epsilon_{2i}(t)$$

$$\texttt{FEV1\%pred:} y_{1i}(t) = x_{1i}^{\top}(t)\beta_1 + z_{1i}(t)^{\top} b_{1i} + \boldsymbol{\alpha}_{S2} \int_0^t m_{2i}(s)ds + \epsilon_{1i}(t)$$

where

- $b_i^{\top} = (b_{1i}^{\top}, b_{2i}^{\top}) \sim N(0, D)$



R package: https://github.com/ERandrinopoulou/multiLME

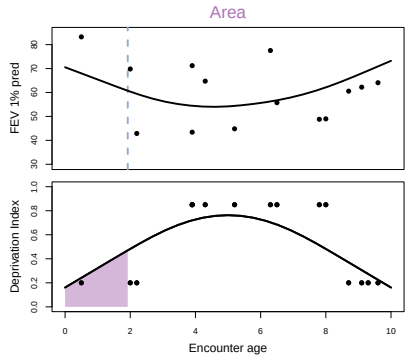
Methods: Multivariate Mixed Models

$$\texttt{DepIndex:} y_{2i}(t) = m_{2i}(t) + \epsilon_{2i}(t) = x_{2i}^{\top}(t)\beta_1 + z_{2i}(t)^{\top} b_{2i} + \epsilon_{2i}(t)$$

$$\texttt{FEV1\%pred:} y_{1i}(t) = x_{1i}^{\top}(t)\beta_1 + z_{1i}(t)^{\top} b_{1i} + \boldsymbol{\alpha}_{S2} \int_{t-d}^{t} m_{2i}(s)ds + \epsilon_{1i}(t)$$

where

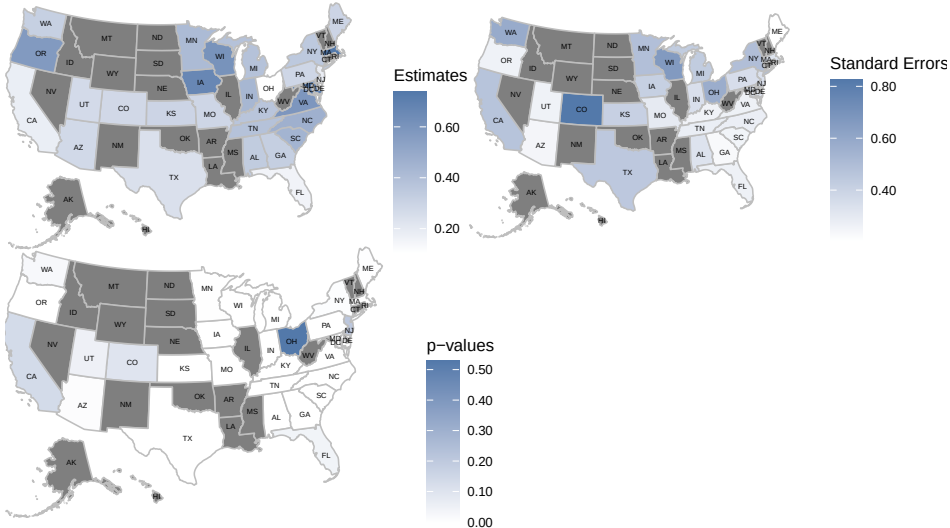
- $b_i^{\top} = (b_{1i}^{\top}, b_{2i}^{\top}) \sim N(0, D)$



R package: https://github.com/ERandrinopoulou/multiLME

Results: Multivariate Mixed Models

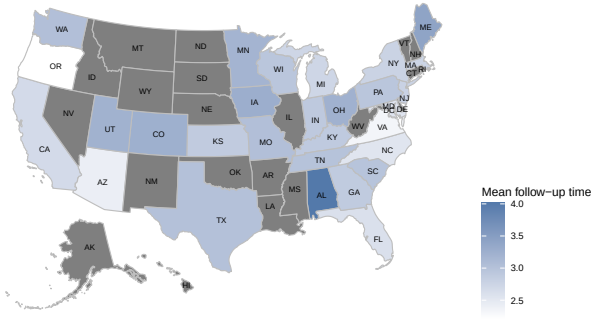
Estimate: for 0.1 unit increase in the area under the curve of the deprivation index



www.erandrinopoulou.com e.andrinopoulou@erasmusmc.nl @ERandrinopoulou

Results: Multivariate Mixed Models

Descriptive statistics: Baseline information



Methods: Multivariate Mixed Models

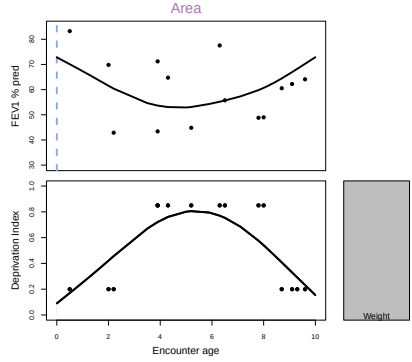
DepIndex: $y_{2i}(t) = m_{2i}(t) + \epsilon_{2i}(t) = x_{2i}^{\top}(t)\beta_1 + z_{2i}(t)^{\top} b_{2i} + \epsilon_{2i}(t)$

FEV1%pred: $y_{1i}(t) = x_{1i}^{\top}(t)\beta_1 + z_{1i}(t)^{\top} b_{1i} + \boldsymbol{\alpha}_{S2} \frac{1}{t} \int_0^t m_{2i}(s)ds + \epsilon_{1i}(t)$

Weight

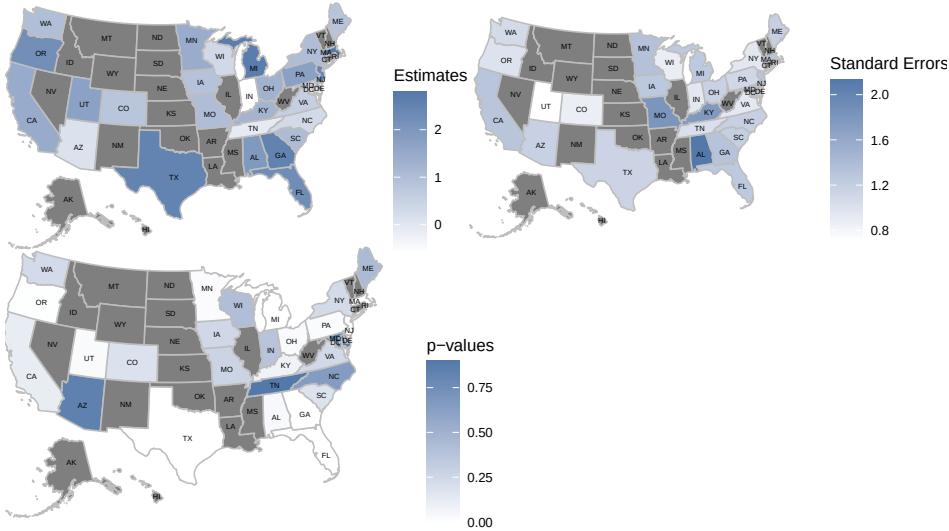
where

- $b_i^{\top} = (b_{1i}^{\top}, b_{2i}^{\top}) \sim N(0, D)$



Results: Multivariate Mixed Models

Estimate: for 0.1 unit increase in the normalized area under the curve of deprivation index



Results: Multivariate Mixed Models

➔ When the exposure time is short (e.g., 2 or 5 years), the association becomes weaker with some differences between the states.

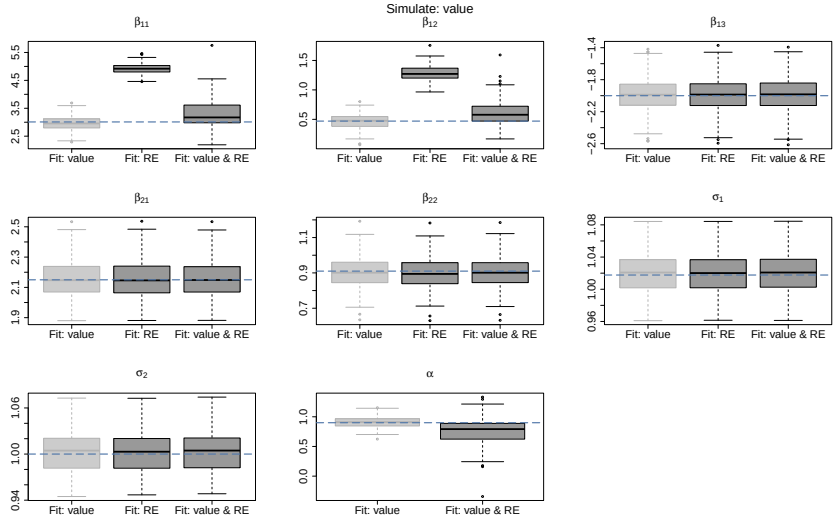
Sensitivity analysis:

➔ When 2 and 5 years are assumed, there is a stronger association when the diagnosis age is *below 18*. When 10 and 15 years are assumed, there is a stronger association when the diagnosis age is *above 12*.

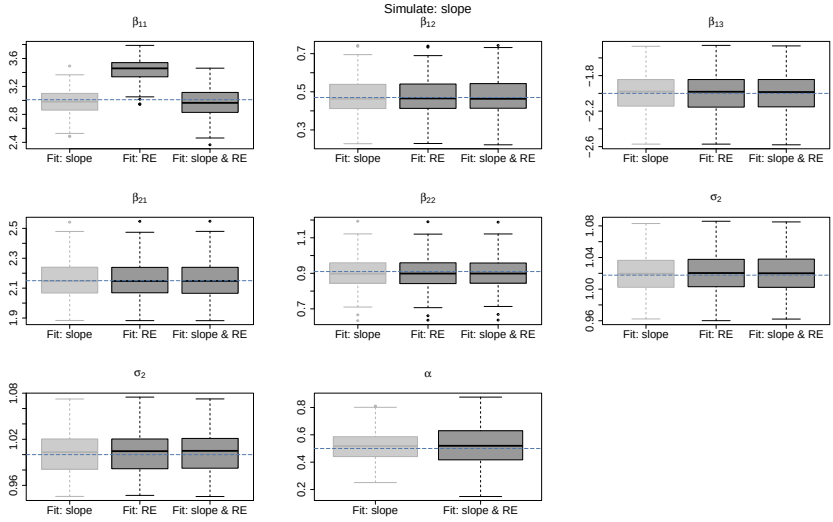
Simulation: Set up

- **Real-world data**: defining the connection between the different longitudinal outcomes is a difficult task
- Bias: when we ignore or over specify the relationship between different data sources.
  - **Simulate**: different forms of association (value/area/slope)
    - **Fit**: value/area/slope
    - RE
    - value/area/slope + RE

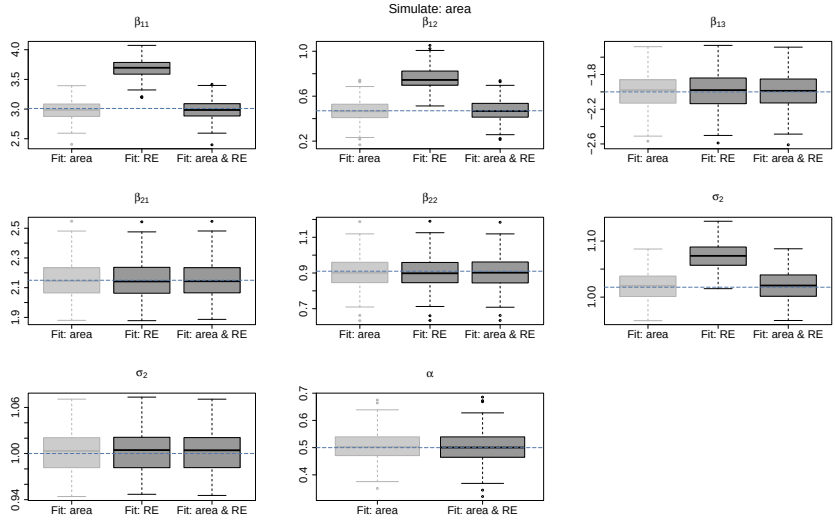
Simulation: Results



Simulation: Results



Simulation: Results



Conclusion

- ➔ A lot of data is available
- ➔ Better treatment and monitoring strategies if all information is used
- ➔ Challenge in combining different types of information
- ➔ Investigate other social and environmental determinants of health

www.erandrinopoulou.com e.andrinopoulou@erasmusmc.nl @ERandrinopoulou

Erasmus MC
University Medical Center Rotterdam

Thank you for your attention!

www.erandrinopoulou.com e.andrinopoulou@erasmusmc.nl @ERandrinopoulou