Multilevel joint modeling of longitudinal and binary outcomes for hierarchically structured data

Grace C. Zhou[1,2], Seongho Song[2], Rhonda D. Szczesniak[1,2]

[1]Cincinnati Children's Hospital Medical Center, Cincinnati, OH
[2]University of Cincinnati, Cincinnati, OH

Outline

1. **Introduction**
   – Literature reviews
   – Objectives
2. **Model framework**
   – The joint model
   – The flexible link function
   – Bayesian inferences
   – Model selection
3. **Simulation**
4. **Application**
   – Motivating Data
   – Results
5. **Discussion**
   – Conclusion
   – Limitation
   – Extension

Introduction

- Literature reviews
  - Henderson et al. (2000)
    - Joint modelling for longitudinal analysis
  - Horrocks and van Den Heuvel (2009)
    - Joint modelling for binary outcomes via a Bayesian perspective
  - Li et al. (2015)
    - A new class of flexible link functions
  - Su et al. (2020)
    - Flexible link functions in a joint hierarchical Gaussian process model

Introduction cont.

- Objectives
  - Address the need to account for center effect in a multicenter cohort analysis using Bayesian joint model (JM);
  - Develop a prognostic lung disease progression;
  - Compare two flexible link functions: splogit and spep.

Model framework

- The Joint Model

$$\begin{cases} Y_{ijk} = \boldsymbol{X}_{ij}^{T}(t_{ijk})\boldsymbol{\alpha} + b_i + U_{ij} + W_{ij}(t_{ijk}) + \epsilon_{ijk} \\ Pr(R_{ijk} = 1) = F_{sp}\{\boldsymbol{V}_{ij}^{T}(t_{ijk})\boldsymbol{\beta} + \rho_1 \times b_i + \rho_2 \times U_{ij}; r_i\} \end{cases}$$

Model framework cont.

- Symmetric Power Link Family

$$F_{sp}(x;r) = F_0^r\left(\frac{x}{r}\right)I_{(0,1]}(r) + \left[1 - F_0^{1/r}(-rx)\right]I_{(1,+\infty)}(r)$$

Model framework cont.

- Bayesian Inferences

$$
\begin{aligned}
\pi(\boldsymbol{\Psi}, \boldsymbol{b}, \boldsymbol{U}, \boldsymbol{W}) &\propto \pi(\boldsymbol{D}_{obs}, \boldsymbol{b}, \boldsymbol{U}, \boldsymbol{W} | \boldsymbol{\Psi}) \pi(\boldsymbol{\Psi}) \\
&\propto \pi(\boldsymbol{D}_{obs} | \boldsymbol{b}, \boldsymbol{U}, \boldsymbol{W}, \boldsymbol{\Psi}) \pi(\boldsymbol{b} | \boldsymbol{U}, \boldsymbol{W}, \boldsymbol{\Psi}) \pi(\boldsymbol{U} | \boldsymbol{W}, \boldsymbol{\Psi}) \pi(\boldsymbol{W} | \boldsymbol{\Psi}) \pi(\boldsymbol{\Psi}) \\
&\propto \prod_{i=1}^{n} \prod_{j=1}^{n_i} I\left(Y_{ij}, R_{ij} \middle| X_{ij}, V_{ij}, W_{ij}, \boldsymbol{\Psi}, b_i, U_{ij}\right) \pi(b_i | \sigma_1) \pi\left(U_{ij} \middle| \sigma_2\right) \pi\left(W_{ij} \middle| \tau, \rho\right) \pi(\boldsymbol{\Psi})
\end{aligned}
$$

- Model Selection
  - Watanabe-Akaike information criterion (WAIC)
  - Watanabe (2010), Gelman et al. (2014)

Simulation

- $\text{JM}_1 \begin{cases} Y_{ijk} = U_{ij} + \epsilon_{ijk} \\ Pr(R_{ijk} = 1) = F_{sp}\{\beta_0 + \beta_1 t_{ijk} + \rho_2 \times U_{ij}; r\} \end{cases}$

- $\text{JM}_2 \begin{cases} Y_{ijk} = b_i + U_{ij} + \epsilon_{ijk} \\ Pr(R_{ijk} = 1) = F_{sp}\{\beta_0 + \beta_1 t_{ijk} + \rho_1 \times b_i + \rho_2 \times U_{ij}; r\} \end{cases}$

- $\text{JM}_3 \begin{cases} Y_{ijk} = b_i + U_{ij} + \epsilon_{ijk} \\ Pr(R_{ijk} = 1) = F_{sp}\{\beta_0 + \beta_1 t_{ijk} + \rho_1 \times b_i + \rho_2 \times U_{ij}; r_i\} \end{cases}$

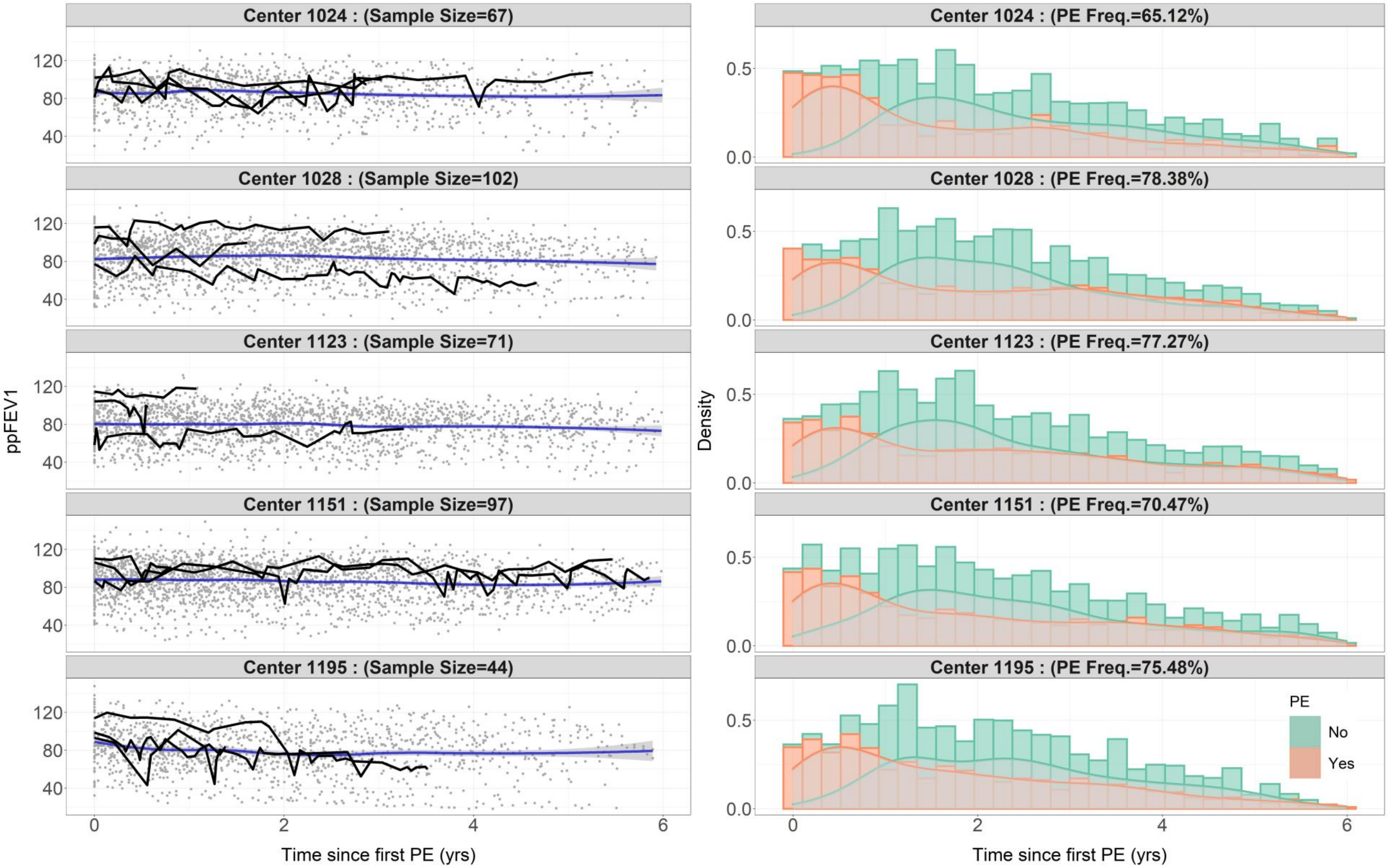
- $\text{JM}_4 \begin{cases} Y_{ijk} = b_i + U_{ij} + W_{ij}(t_{ijk}) + \epsilon_{ijk} \\ Pr(R_{ijk} = 1) = F_{sp}\{\beta_0 + \beta_1 t_{ijk} + \rho_1 \times b_i + \rho_2 \times U_{ij}; r_i\} \end{cases}$

Simulation Cont.

	splogit			spep		
	WAIC	WAIC1	WAIC2	WAIC	WAIC1	WAIC2
Misspecified(JM1)	9455.853	8369.268	1086.585	9256.988	8299.090	957.898
No center-index (JM2)	9362.319	8360.999	1001.320	9179.925	8292.143	887.782
No covariance (JM3)	9328.386	8360.075	968.311	9157.464	8291.870	865.594
Proposed (JM4)	8508.909	7574.098	934.811	**8352.834**	7541.872	810.962

Application

- Motivating Data



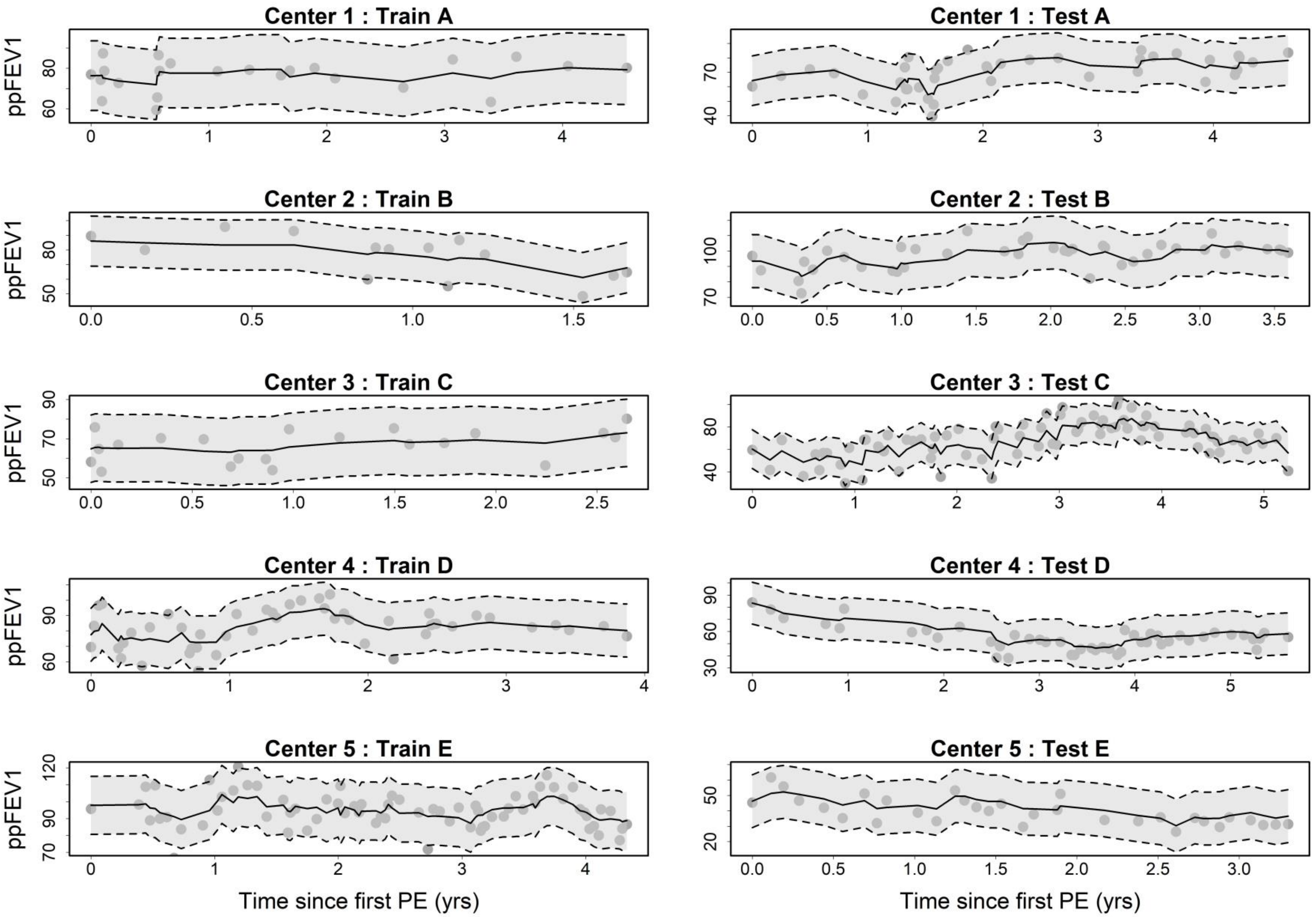
Application Cont.

	spep		
	WAIC	WAIC1	WAIC2
Misspecified(JM1)	49960.1	44302.0	5658.1
No center-index (JM2)	49867.6	44244.8	5622.8
No covariance (JM3)	49812.0	44228.0	5584.0
Proposed (JM4)	**47082.3**	42704.8	4377.5

Application Cont.

	Training				Testing			
	FEV1		PE		FEV1		PE	
	RMSE	SE	AUC	95% CI	RMSE	SE	AUC	95% CI
Misspecified(JM1)	10.755	0.142	0.748	(0.734, 0.763)	10.397	0.233	0.623	(0.594, 0.651)
No center-index (JM2)	10.686	0.141	0.748	(0.734, 0.763)	10.399	0.233	0.622	(0.593, 0.651)
No covariance (JM3)	10.671	0.141	0.755	(0.741, 0.77)	10.398	0.233	0.639	(0.610, 0.667)
Proposed (JM4)	7.768	0.102	0.882	(0.873, 0.892)	6.879	0.154	0.631	(0.604, 0.658)

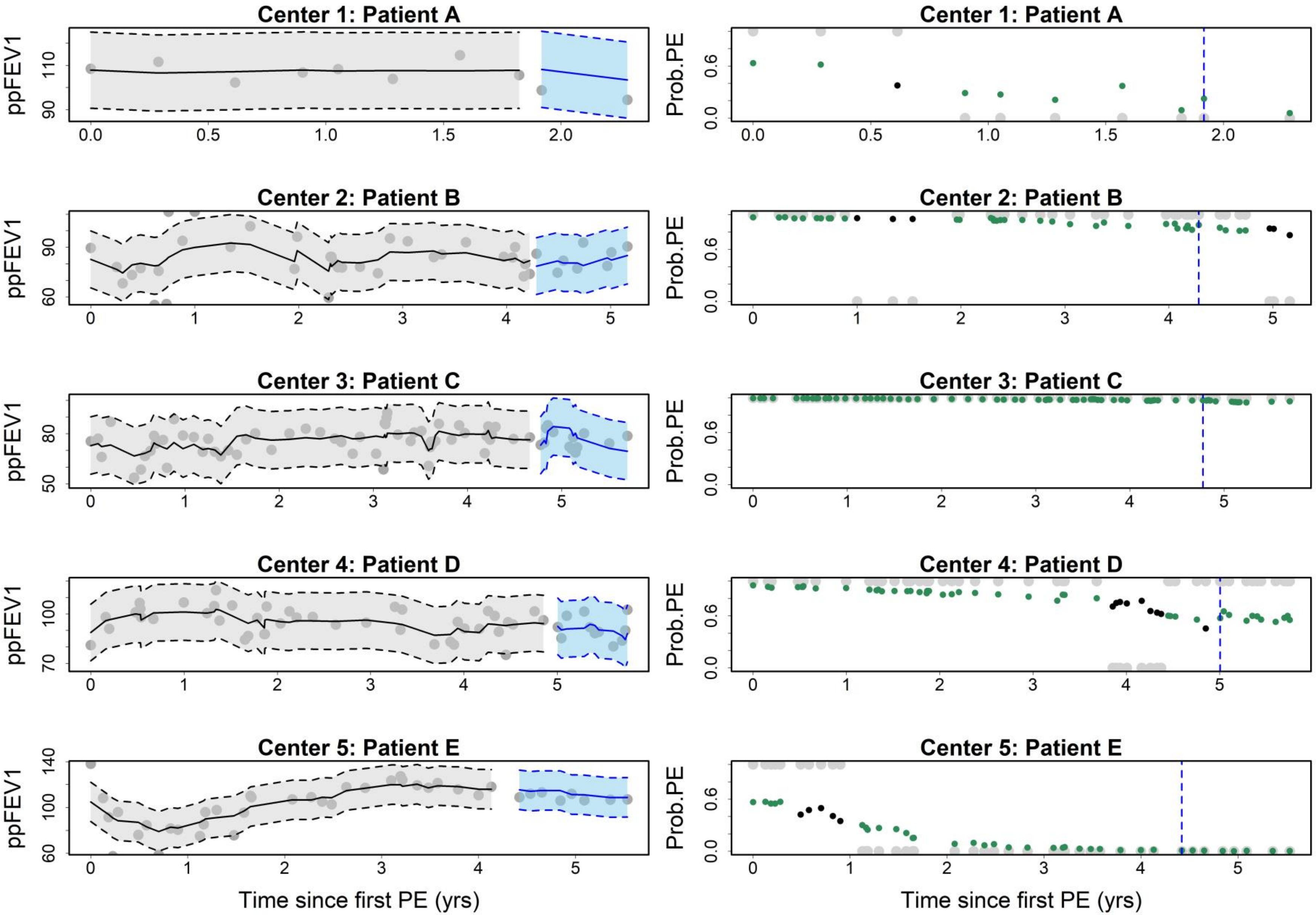
Application Cont.



Application Cont.

	Training				Masking			
	FEV1		PE		FEV1		PE	
	RMSE	SD	AUC	95% CI	RMSE	SD	AUC	95% CI
Misspecified(JM1)	10.755	0.142	0.748	(0.734, 0.763)	10.354	0.272	0.655	(0.626, 0.683)
No center-index (JM2)	10.686	0.141	0.748	(0.734, 0.763)	10.298	0.270	0.628	(0.598, 0.658)
No covariance (JM3)	10.671	0.141	0.755	(0.741, 0.77)	10.353	0.271	0.612	(0.582, 0.642)
Proposed (JM4)	7.768	0.102	0.882	(0.873, 0.892)	8.850	0.233	0.785	(0.760, 0.809)

Application Cont.



Cincinnati Children's
changing the outcome together
University of CINCINNATI

Discussion

- Conclusion
  - Ignoring center effect would induce bias
  - Our proposed model demonstrates capability to fit heterogeneous nature of CF data
  - Flexible link function spep facilitate more accuracy than splogit

Discussion

- Limitation
  - GLM submodel is not flexible to capture re-occurrence of PE events
  - Future work is needed to improve predictive performance

Discussion

- Extension
  - Survival submodel instead of GLM submodel
  - Joint latent class mixed model
  - Leave-one-out cross validation (LOO-CV)

Acknowledgement

- Grant R01 HL141286 from the National Institutes of Health

- Co-authors