Scalable Joint Modeling of Longitudinal and Point Process Data: Disease Trajectory Prediction and Improving Management of Chronic Kidney Disease

Joe Futoma (jdf38@duke.edu)

UAI Bayesian Applications Workshop, 2016

June 29, 2016
Outline

❖ Motivation
❖ Working with EHR Data
❖ Proposed Joint Model
❖ Experiments & Results
❖ In Clinical Practice
Untreated diabetes & high blood pressure.

Normal kidney function, but with evidence of kidney damage.

No regular medical care.
Age 48
A few more ER trips
Kidney function falling
Age 49
Kidney function now 50%
Age 50
Gets a Duke Primary Care doc
Kidney function 30%
Age 51
CKD first noted on problem list
Three months later presents to ER with kidney failure symptoms and “crash starts” dialysis
Missed Opportunities:

To prevent or delay kidney failure

To prepare for kidney failure
42% starting dialysis have no prior nephrology care
<10% with moderate CKD
<50% with severe CKD
even aware of illness!
Chronic Kidney Disease (CKD)

Progressive loss of kidney function with significant morbidity, mortality, health system utilization, economic costs
Chronic Kidney Disease (CKD)

Heart disease

Progressive loss of kidney function with significant morbidity, mortality, health system utilization, economic costs
Progressive loss of kidney function with significant morbidity, mortality, health system utilization, economic costs
Outline

- Motivation
- Working with EHR Data
- Proposed Joint Model
- Experiments & Results
- In Clinical Practice
Real EHR data poses many challenges!
Data acquisition from DEDUCE

630,000 patients pulled, one encounter in previous year (winter 2015)

393,000 with at least 1 serum creatinine lab

115,000 with 10+ labs

30,000 patients age 65+ and meet criteria for moderate stage CKD
Quantifying CKD Progression

- Estimated glomerular filtration rate (eGFR) is an extremely noisy estimate of kidney function. Most common validated equation:

\[
eGFR = 141 \cdot \min(S_{cr}/\kappa, 1)^{\alpha} \cdot \max(S_{cr}/\kappa, 1)^{-1.209} \cdot 0.993^{\text{Age}} \cdot 1.018^{\mathbb{I}(\text{female})} \cdot 1.159^{\mathbb{I}(\text{black})}
\]

- \(S_{cr}\): Serum creatinine (mg/dL)
- \(\kappa\): 0.7 (female), 0.9 (male)
- \(\alpha\): -0.329 (female), -0.411 (male)
Estimated glomerular filtration rate (eGFR) is an extremely noisy

Renal Function Trajectory Is More Important than Chronic Kidney Disease Stage for Managing Patients with Chronic Kidney Disease

Steven J. Rosansky

WJB Dorn Veteran’s Hospital, Columbia, S.C., USA
Outline

- Motivation
- Working with EHR Data
- Proposed Joint Model
- Experiments & Results
- In Clinical Practice
Proposed Joint Model

- **Goal:** jointly model risks of future loss of kidney function, cardiac events
  - Heart attacks (AMI), Stroke (CVA)
- **Hierarchical latent variable model:** captures dependencies between disease trajectory and event risk
  - Submodels for longitudinal, event data with shared latent variables
- $\tilde{y}_i$: eGFRs at times $\tilde{t}_i$; $\tilde{u}_i$: event times (may be none); $x_i$ covariates
- **Conditional independence in joint likelihood:**
  \[
p(y_i, u_i | z_i, b_i, f_i, v_i; x_i) = p(y_i | z_i, b_i, f_i; x_i)p(u_i | z_i, b_i, f_i, v_i; x_i)
  \]
Longitudinal Submodel
Longitudinal Submodel

- Longitudinal values conditionally independent: \( p(y_i | z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij} | z_i, b_i, f_i) \)
Longitudinal Submodel

- Longitudinal values conditionally independent: $p(y_i | z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij} | z_i, b_i, f_i)$
- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean
Longitudinal Submodel

- Longitudinal values conditionally independent: 
  \[ p(y_i | z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij} | z_i, b_i, f_i) \]

- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

\[
\begin{align*}
y_i(t) &= m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \sim iid \; N(0, \sigma^2_{\epsilon}) \\
m_i(t) &= \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{zi} + \Phi_l(t)^\top b_i + f_i(t).
\end{align*}
\]
Longitudinal Submodel

- Longitudinal values conditionally independent: \( p(\vec{y}_i|z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij}|z_i, b_i, f_i) \)

- Values normally distributed, with mean a sum of 4 terms

  - Can also view as GP with highly structured mean

\[
\begin{align*}
    y_i(t) &= m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2_{\epsilon}) \\
m_i(t) &= \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{zi} + \Phi_l(t)^\top b_i + f_i(t).
\end{align*}
\]
Longitudinal Submodel

- Longitudinal values conditionally independent: 
  \[ p(y_i|z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij}|z_i, b_i, f_i) \]

- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

  \[
  y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2_{\epsilon})
  \]

  \[
  m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{zi} + \Phi_l(t)^\top b_i + f_i(t).
  \]
Longitudinal Submodel

- Longitudinal values conditionally independent: 
  \[ p(y_i | z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij} | z_i, b_i, f_i) \]

- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

\[
y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \sim iid N(0, \sigma^2_\epsilon) \\
m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{zi} + \Phi_l(t)^\top b_i + f_i(t).
\]

Population component (fixed intercept and slope)
Longitudinal Submodel

- Longitudinal values conditionally independent: 
  \[ p(y_i^T | z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij} | z_i, b_i, f_i) \]

- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

\[
y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2)\\
m_i(t) = \Phi_p(t)^T \Lambda x_{ip} + \Phi_z(t)^T \beta_{zi} + \Phi_l(t)^T b_i + f_i(t).
\]

Population component (fixed intercept and slope)

- Basis expansion, in practice [1, t]
- Coefficient matrix
- Baseline covariates
Longitudinal Submodel

- Longitudinal values conditionally independent: 
  \[ p(y_i|z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij}|z_i, b_i, f_i) \]
- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

\[
y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2_{\epsilon})
\]
\[
m_i(t) = \Phi_p(t)\Lambda x_{ip} + \Phi_z(t)\beta_{zi} + \Phi_l(t)b_i + f_i(t).
\]
Longitudinal Submodel

- Longitudinal values conditionally independent: 
  \[ p(y_i | z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij} | z_i, b_i, f_i) \]

- Values normally distributed, with mean a sum of 4 terms
  
  - Can also view as GP with highly structured mean

\[
y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2_{\epsilon})
\]

\[
m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{zi} + \Phi_l(t)^\top b_i + f_i(t).
\]

Subpopulation component (unique disease trajectory with splines)
Longitudinal Submodel

- Longitudinal values conditionally independent: 
  \[ p(y_i | z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij} | z_i, b_i, f_i) \]

- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

\[
y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2_{\epsilon})
\]

\[
m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t).
\]

Subpopulation component (unique disease trajectory with splines)
Longitudinal Submodel

- Longitudinal values conditionally independent: \( p(y_i | z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij} | z_i, b_i, f_i) \)
- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

\[
y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \sim iid \ N(0, \sigma^2_\epsilon)
\]
\[
m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t).
\]
Longitudinal Submodel

- Longitudinal values conditionally independent: \( p(y_i|z_i,b_i,f_i) = \prod_{j=1}^{N_i} p(y_{ij}|z_i,b_i,f_i) \)
- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

\[
y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2_{\epsilon})
\]
\[
m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{zi} + \Phi_l(t)^\top b_i + f_i(t).
\]

Individual component (random intercept and slope)
Longitudinal Submodel

- Longitudinal values conditionally independent: 
  \[ p(y_i | z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij} | z_i, b_i, f_i) \]

- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

\[ y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2_{\epsilon}) \]

\[ m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t). \]

Individual component (random intercept and slope)

- Basis expansion, in practice \([1,t]\)
- Random effect, \(b_i \sim N(0, \Sigma_b)\)
Longitudinal Submodel

- Longitudinal values conditionally independent: \( p(y_i | z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij} | z_i, b_i, f_i) \)

- Values normally distributed, with mean a sum of 4 terms
  
  - Can also view as GP with highly structured mean

\[ y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2_\epsilon) \]

\[ m_i(t) = \Phi_p(t) \Lambda x_{ip} + \Phi_z(t) \beta_{zi} + \Phi_l(t) b_i + f_i(t). \]
Longitudinal Submodel

- Longitudinal values conditionally independent: 
  \[ p(y_i|z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij}|z_i, b_i, f_i) \]

- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

\[
y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2_{\epsilon})
\]
\[
m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{zi} + \Phi_l(t)^\top b_i + f_i(t).
\]

Structured noise process (GP noise, transient trends)
Longitudinal Submodel

- Longitudinal values conditionally independent: \( p(y_i | z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij} | z_i, b_i, f_i) \)

- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

\[
y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \sim iid N(0, \sigma^2_{\epsilon})
\]
\[
m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{zi} + \Phi_l(t)^\top b_i + f_i(t).
\]

Structured noise process (GP noise, transient trends)

\[
K_{OU}(t_1, t_2) = \sigma_f^2 \exp\left\{-\frac{|t_1 - t_2|}{l}\right\}
\]
Longitudinal Submodel

- Longitudinal values conditionally independent: $p(y_i|z_i, b_i, f_i) = \prod_{j=1}^{N_i} p(y_{ij}|z_i, b_i, f_i)$

- Values normally distributed, with mean a sum of 4 terms
  - Can also view as GP with highly structured mean

\[
y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2_\epsilon)
\]
\[
m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{zi} + \Phi_l(t)^\top b_i + f_i(t).
\]
Point Process Submodel
Point Process Submodel

- Poisson Process model, conditional likelihood on \([T_i^-, T_i^+]\), events at \(\{u_{ik}\}_{k=1}^{K_i}\):
Poisson Process model, conditional likelihood on $[T_i^-, T_i^+]$, events at $\{u_{ik}\}_{k=1}^{K_i}$:

$$p(\tilde{u}_i | z_i, b_i, f_i, v_i) = \prod_{k=1}^{K_i} r_i(u_{ik}) \exp\{- \int_{T_i^-}^{T_i^+} r_i(t) dt\}$$
Point Process Submodel

- Poisson Process model, conditional likelihood on \([T_i^-, T_i^+]\), events at \(\{u_{ik}\}_{k=1}^{K_i}\):
  \[
p(u_i|z_i, b_i, f_i, v_i) = \prod_{k=1}^{K_i} r_i(u_{ik}) \exp\left\{-\int_{T_i^-}^{T_i^+} r_i(t) dt\right\}
\]

- Rate function: hazard function from Cox proportional hazards model
  - Common choice in survival analysis
Point Process Submodel

- Poisson Process model, conditional likelihood on $[T_i^-, T_i^+]$, events at $\{u_{ik}\}_{k=1}^{K_i}$:

$$p(u_i|z_i, b_i, f_i, v_i) = \prod_{k=1}^{K_i} r_i(u_{ik}) \exp\left\{- \int_{T_i^-}^{T_i^+} r_i(t) dt\right\}$$

- Rate function: hazard function from Cox proportional hazards model
  - Common choice in survival analysis

$$r_i(t) = r_0(t) \exp\{\gamma^\top x_i + \alpha m_i(t) + \delta m'_i(t) + v_i\}$$
Point Process Submodel

- Poisson Process model, conditional likelihood on \([T_i^-, T_i^+]\), events at \(\{u_{ik}\}_{k=1}^{K_i}\):

\[
p(u_i | z_i, b_i, f_i, v_i) = \prod_{k=1}^{K_i} r_i(u_{ik}) \exp\left\{ - \int_{T_i^-}^{T_i^+} r_i(t) dt \right\}
\]

- Rate function: hazard function from Cox proportional hazards model
  
  - Common choice in survival analysis

\[
r_i(t) = r_0(t) \exp\{\gamma^\top x_{ir} + \alpha m_i(t) + \delta m_i'(t) + v_i\}
\]

piecewise constant baseline rate
Point Process Submodel

- Poisson Process model, conditional likelihood on $[T_i^-, T_i^+]$, events at $\{u_{ik}\}_{k=1}^{K_i}$:
  \[
p(u_i | z_i, b_i, f_i, v_i) = \prod_{k=1}^{K_i} r_i(u_{ik}) \exp\left\{- \int_{T_i^-}^{T_i^+} r_i(t) dt\right\}
\]
- Rate function: hazard function from Cox proportional hazards model
  - Common choice in survival analysis
  \[
r_i(t) = r_0(t) \exp\{\gamma^\top x_{ir} + \alpha m_i(t) + \delta m'_i(t) + v_i\}
\]

piecewise constant baseline rate
coefficient vector
Point Process Submodel

- Poisson Process model, conditional likelihood on \([T_i^-, T_i^+]\), events at \(\{u_{ik}\}_{k=1}^{K_i}\):
  \[
p(\bar{u}_i|z_i, b_i, f_i, v_i) = \prod_{k=1}^{K_i} r_i(u_{ik}) \exp\{-\int_{T_i^-}^{T_i^+} r_i(t) dt\}
\]
- Rate function: hazard function from Cox proportional hazards model
  - Common choice in survival analysis
    \[
    r_i(t) = r_0(t) \exp\{\gamma^\top x_{ir} + \alpha m_i(t) + \delta m_i'(t) + v_i\}
    \]

piecewise constant
baseline rate
coefficient vector
baseline covariates
Poisson Process model, conditional likelihood on \([T_i^-, T_i^+]\), events at \(\{u_{ik}\}_{k=1}^{K_i}\):

\[
p(u_i | z_i, b_i, f_i, v_i) = \prod_{k=1}^{K_i} r_i(u_{ik}) \exp\left\{- \int_{T_i^-}^{T_i^+} r_i(t) dt \right\}
\]

Rate function: hazard function from Cox proportional hazards model

- Common choice in survival analysis

\[
r_i(t) = r_0(t) \exp\{\gamma^\top x_{ir} + \alpha m_i(t) + \delta m_i'(t) + v_i\}
\]
Point Process Submodel

- Poisson Process model, conditional likelihood on \([T_i^-, T_i^+]\), events at \(\{u_{ik}\}_{k=1}^{K_i}\):
  \[
p(u_i|z_i, b_i, f_i, v_i) = \prod_{k=1}^{K_i} r_i(u_{ik}) \exp\left\{ - \int_{T_i^-}^{T_i^+} r_i(t) dt \right\}
\]

- Rate function: hazard function from Cox proportional hazards model
  \[
r_i(t) = r_0(t) \exp\{\gamma^\top x_{ir} + \alpha m_i(t) + \delta m_i'(t) + v_i\}
\]

- Common choice in survival analysis

  - Piecewise constant baseline rate
  - Coefficient vector
  - Baseline covariates
  - Association between event risk and expected mean/slope of eGFR
  - Random effect (frailty term): \(v_i \sim N(0, \sigma_v^2)\)
Inference

- Variational inference: find distribution \( q \) in approx. family close in KL to true posterior
  - Equivalently, maximize a lower bound on marginal likelihood:
    \[
p(y, u) \geq \mathcal{L}(q) \equiv E_q[\log p(y, u, z, b, f, v, \Theta) - \log q(z, b, f, v, \Theta)]
    \]
- Mean-field assumption, fully factorized variational family:
  \[
  q(z, b, f, v, \Theta) = q(\Theta) \prod_{i=1}^{N} q_i(z_i|\nu_{z_i})q_i(b_i|\mu_{b_i}, \Sigma_{b_i})q_i(v_i|\mu_{v_i}, \sigma_{v_i}^2)q_i(f_i)
  \]
  - Variational distributions have same form as prior (multinomial, MVN, N)
  - For \( f \), adapt ideas from sparse GPs, use observation times as pseudo-inputs [Lloyd et al, 2014]
- Goal: learn optimal var. params. \( \lambda_i = \{\nu_{z_i}, \mu_{b_i}, \Sigma_{b_i}, \mu_{v_i}, \sigma_{v_i}^2, \mu_f, \Sigma_f\} \), pt. est. \( \hat{\Theta} \)
- ELBO has closed form, exact gradients with automatic differentiation
- Stochastic optimization, **subsample** observations for noisy unbiased gradients wrt \( \Theta \)
Related Work

- Longitudinal model from [Schulam & Saria, 2015]

- Joint models in biostatistics: [Rizopoulos, 2012], [Proust-Lima et al., 2014] for introductions
  - Typically fit via EM for MLE or MCMC for Bayesian setting

- In medicine: cross-sectional, data from single time point
  - E.g. [Tangri et al. 2011]; no dynamic predictions, limits clinical utility
Outline

❖ Motivation
❖ Working with EHR Data
❖ Proposed Joint Model
❖ Experiments & Results
❖ In Clinical Practice
Dataset

- 23,450 patients with moderate stage CKD and 10+ eGFR readings
  - CKD definition: 2 eGFR readings < 60mL/min, separated by 90+ days
- Preprocessing: mean in monthly bins
  - eGFR only valid estimate of kidney function at steady state
  - 22.9 readings on average (std. dev. 13.6, median 19.0)
  - Alignment: set $t=0$ to be first eGFR reading < 60mL/min
- Adverse events: AMI, CVA identified using ICD9 codes. Max 1 event / month
  - 13.4% had 1+ AMI code (mean w/ 1+: 4.1, std dev: 7.1, median: 2.0)
  - 17.4% had 1+ AMI code (mean w/ 1+: 6.4, std dev: 13.3, median: 3.0)
- Baseline covariates: baseline age, race, gender; hypertension, diabetes
Experimental Setup

- Use first 60% of eGFR trajectory/events to predict last 40%
- Evaluation metrics:
  - MSE and MAE for held-out eGFR values
  - AUROC, AUPR for predicting any event in \([T,T+c]\) as binary classification
- Longitudinal baseline: [Schulam & Saria, 2015]
- Point process baselines:
  - Cox regression; rate: \( r_i(t) = r_0(t) \exp\{\gamma^\top x_{ir}\} \)
  - Time-varying Cox using observed eGFR; rate: \( r_i(t) = r_0(t) \exp\{\gamma^\top x_{ir} + \alpha y_i(t)\} \)
## Quantitative Results

### Longitudinal Submodels

<table>
<thead>
<tr>
<th>Submodel</th>
<th>MSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Schulam &amp; Saria, 2015]</td>
<td>155.80</td>
<td>9.27</td>
</tr>
<tr>
<td>Joint Model (CVA)</td>
<td>147.31</td>
<td>9.01</td>
</tr>
<tr>
<td>Joint Model (AMI)</td>
<td>152.78</td>
<td>9.15</td>
</tr>
</tbody>
</table>

### CVA: AUROCs

<table>
<thead>
<tr>
<th></th>
<th>1 yr.</th>
<th>2 yr.</th>
<th>3 yr.</th>
<th>4 yr.</th>
<th>5 yr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint Model</td>
<td>0.786</td>
<td>0.746</td>
<td>0.727</td>
<td>0.742</td>
<td>0.740</td>
</tr>
<tr>
<td>Cox</td>
<td>0.574</td>
<td>0.597</td>
<td>0.602</td>
<td>0.606</td>
<td>0.587</td>
</tr>
<tr>
<td>Time-varying Cox</td>
<td>0.576</td>
<td>0.557</td>
<td>0.563</td>
<td>0.593</td>
<td>0.566</td>
</tr>
</tbody>
</table>

### AMI: AUROCs

<table>
<thead>
<tr>
<th></th>
<th>1 yr.</th>
<th>2 yr.</th>
<th>3 yr.</th>
<th>4 yr.</th>
<th>5 yr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint Model</td>
<td>0.755</td>
<td>0.704</td>
<td>0.737</td>
<td>0.654</td>
<td>0.663</td>
</tr>
<tr>
<td>Cox</td>
<td>0.704</td>
<td>0.676</td>
<td>0.617</td>
<td>0.599</td>
<td>0.640</td>
</tr>
<tr>
<td>Time-varying Cox</td>
<td>0.640</td>
<td>0.652</td>
<td>0.647</td>
<td>0.663</td>
<td>0.655</td>
</tr>
</tbody>
</table>

### CVA: AUPRs

<table>
<thead>
<tr>
<th></th>
<th>1 yr.</th>
<th>2 yr.</th>
<th>3 yr.</th>
<th>4 yr.</th>
<th>5 yr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint Model</td>
<td>0.423</td>
<td>0.389</td>
<td>0.370</td>
<td>0.405</td>
<td>0.400</td>
</tr>
<tr>
<td>Cox</td>
<td>0.065</td>
<td>0.101</td>
<td>0.123</td>
<td>0.137</td>
<td>0.134</td>
</tr>
<tr>
<td>Time-varying Cox</td>
<td>0.062</td>
<td>0.086</td>
<td>0.114</td>
<td>0.157</td>
<td>0.130</td>
</tr>
</tbody>
</table>

### AMI: AUPRs

<table>
<thead>
<tr>
<th></th>
<th>1 yr.</th>
<th>2 yr.</th>
<th>3 yr.</th>
<th>4 yr.</th>
<th>5 yr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint Model</td>
<td>0.163</td>
<td>0.128</td>
<td>0.172</td>
<td>0.166</td>
<td>0.119</td>
</tr>
<tr>
<td>Cox</td>
<td>0.052</td>
<td>0.059</td>
<td>0.051</td>
<td>0.065</td>
<td>0.083</td>
</tr>
<tr>
<td>Time-varying Cox</td>
<td>0.048</td>
<td>0.057</td>
<td>0.067</td>
<td>0.088</td>
<td>0.103</td>
</tr>
</tbody>
</table>
Joint Model Results

Predicted eGFR Trajectory and Event Probabilities for a Held-out Patient
Outline

- Motivation
- Working with EHR Data
- Proposed Joint Model
- Experiments & Results
- In Clinical Practice
In clinical practice...

- Implementation underway: Duke Connected Care, Accountable Care Organization responsible for cost and quality of care of 45,000 Medicare patients
  - 12,000 patients with at least moderate CKD, 1,000 end-stage kidney disease
  - Cost of end-stage disease >$60k per year vs <$10k for average patient
- Goal: incorporate risk stratification and joint model predictions into dashboard application
  - Identify, better manage care of high-risk patients
Conclusion

❖ Novel joint model for longitudinal, point process data
  ❖ First stochastic variational inference algorithm for joint models
❖ **Actionable** implementation in use for CKD patients!

❖ Future work:
  ❖ Multivariate in both longitudinal variables and events
  ❖ More flexible models (e.g. beyond Cox assumption)
  ❖ More clinically actionable metrics to evaluate models
Thank you!
Acknowledgments

- Contact: jdf38@duke.edu

Joint work with:

- Mark Sendak, M.P.P./M.D. Candidate
- C. Blake Cameron, M.D.
- Katherine Heller, Ph.D.


More Inference Details

- Sparse GP / pseudo-inputs: full MVN variational distribution at eGFR times, true conditional $p$ for rest

**Data:** data $y, u$; hyperparameters.
**Result:** point estimate $\hat{\Theta}$, approximate posteriors $q_i$.

Initialize global parameters $\Theta$.

repeat

Randomly sample data for $S$ patients, $\{y_s, u_s\}_{s=1}^S$.

for $s = 1:S$ in parallel do

Optimize local variational parameters for $q_s$ via gradient ascent.

end

Compute the noisy gradient for $\Theta$.

Update $\Theta$ using AdaGrad.

until convergence of the ELBO;

**Algorithm 1:** Stochastic Variational Inference algorithm for our Joint Model.
More Inference Details

Sparse GP / pseudo-inputs: full MVN variational distribution at eGFR times, true conditional $p$ for rest

$$q_i(f_i(t_i^\rightarrow), f_i(w_i), f_i(t_i^{\text{grid}})) = p(f_i(w_i), f_i(t_i^{\text{grid}})|f_i(t_i))q(f_i(t_i)|\mu_{f_i}, \Sigma_{f_i})$$

Data: data $y, u$; hyperparameters.
Result: point estimate $\Theta$, approximate posteriors $q_i$. Initialize global parameters $\Theta$.
repeat
  Randomly sample data for $S$ patients, $\{y_s, u_s\}_{s=1}^S$.
  for $s = 1:S$ in parallel do
    Optimize local variational parameters for $q_s$ via
gradient ascent.
  end
  Compute the noisy gradient for $\Theta$.
  Update $\Theta$ using AdaGrad.
until convergence of the ELBO;
Algorithm 1: Stochastic Variational Inference algorithm for our Joint Model.