Predicting Disease Progression with a Model for Multivariate Longitudinal Clinical Data

Joseph Futoma   JDF38@DUKE.EDU
Dept. of Statistical Science
Duke University
Durham, NC 27708

Mark Sendak   MARK.SENDAK@DM.DUKE.EDU
Institute for Health Innovation
School of Medicine
Duke University
Durham, NC 27708

C. Blake Cameron   BLAKE.CAMERON@DM.DUKE.EDU
Division of Nephrology
Duke University
Durham, NC 27708

Katherine Heller   KHELLER@STAT.DUKE.EDU
Dept. of Statistical Science
Duke University
Durham, NC 27708

Abstract

Accurate prediction of the future trajectory of a disease is an important challenge in personalized medicine and population health management. However, many complex chronic diseases exhibit large degrees of heterogeneity, and furthermore there is not always a single readily available biomarker to quantify disease severity. Even when such a clinical variable exists, there are often additional related biomarkers that may help improve prediction of future disease state. To this end, we propose a novel probabilistic generative model for multivariate longitudinal data that captures dependencies between multivariate trajectories of clinical variables. We use a Gaussian process based regression model for each individual trajectory, and build off ideas from latent class models to induce dependence between their mean functions. We develop a scalable variational inference algorithm that we use to fit our model to a large dataset of longitudinal electronic patient health records. Our model's dynamic predictions have significantly lower errors compared to a recent state of the art method for modeling disease trajectories, and they are being incorporated into a population health rounding tool to be used by clinicians at our local accountable care organization.

1. Introduction

The ability of healthcare organizations to utilize predictions about individuals' future health to make appropriate interventions is becomingly increasingly important with the adoption of accountable care and alternative payment models (CMS (2016)). In particular, managing patients with complex, chronic diseases such as cardiovascular disease, diabetes, and chronic

kidney disease is especially difficult, as selection of optimal therapy may require integration of multiple conditions and risk factors that are considered in isolation under current approaches to care. Further, patients with multiple chronic conditions are among both the most expensive and highest utilizers of healthcare services (Johnson and et al. (2015)).

However, predicting future disease trajectory is an extremely challenging problem. One difficulty is the many underlying sources of variability that can drive the different potential manifestations of the disease. For instance, the underlying biological mechanisms of the disease can give rise to latent disease subtypes, or groups of individuals with shared characteristics (Saria and Goldenberg (2015)). For most complicated diseases, there are no clear definitions of subtypes, so this must be inferred from the data. In addition, there are individual-specific sources of variability that may not be directly observed, such as behavioral and genetic factors, environmental conditions, or temporary infections. Another challenge is the fact that observations are irregularly sampled, asynchronous, and episodic, precluding the use of many time series methods developed for data regularly sampled at discrete time intervals. The large degree of missing data, especially when modeling multivariate longitudinal data, also presents complications. The task is made even more difficult when the primary data source is the electronic health record (EHR) rather than a curated registry, as even selecting a relevant cohort of patients to model can take extensive review by a clinical expert.

In our work, we collaborated with Duke Connected Care, an accountable care organization affiliated with the Duke University Health System. Accountable care organizations are healthcare institutions that aim to link provider reimbursement with quality metrics and reductions in total cost of healthcare services to a specific population of patients. These organizations need personalized prediction tools capable of flagging specific patients in their populations at the highest risk of having poor outcomes (Parikh et al. (2016)). Data to develop such tools are already largely being collected automatically with the widespread adoption of EHRs. However, in order to be clinically relevant the prediction methods will need to be flexible enough to accommodate the inherent limitations of operational EHR data (Hersh and et al (2013)), dynamically update predictions as more information is collected, and scale seamlessly to the massive size of modern health records.

We focused our efforts on developing models for chronic kidney disease (CKD), which is characterized by a slow and typically symptomless loss of kidney function over time that results in complications that can cause poor health, premature death, increased health service utilization, and excess economic costs. A person's kidney function can be approximated by their estimated glomerular filtration rate (eGFR), which is calculated using a common clinical laboratory test (serum creatinine or cystatin C) and demographic information (age, sex and race) (Levey and et al. (2009)). Impairment of a person's eGRF is used to define and stage CKD. In addition, a number of other routinely measured laboratory values may be used to help detect abnormal or declining kidney function. Despite the fact that CKD can be easily identified using simple eGFR-based laboratory criteria, it is often not recognized (Szczech et al. (2014), Tuot et al. (2011)), and providing optimal care for CKD patients is notoriously difficult. Even among patients with recognized CKD, prediction of progression to kidney failure (requiring dialysis or kidney transplantation), or other complications such as early death from heart attack or stroke, is a formidable task (Mendelssohn et al. (2011)). Additionally, providers often fail to prescribe appropriate preventive treatment to slow disease progression or address complications (Smart et al. (2014)).
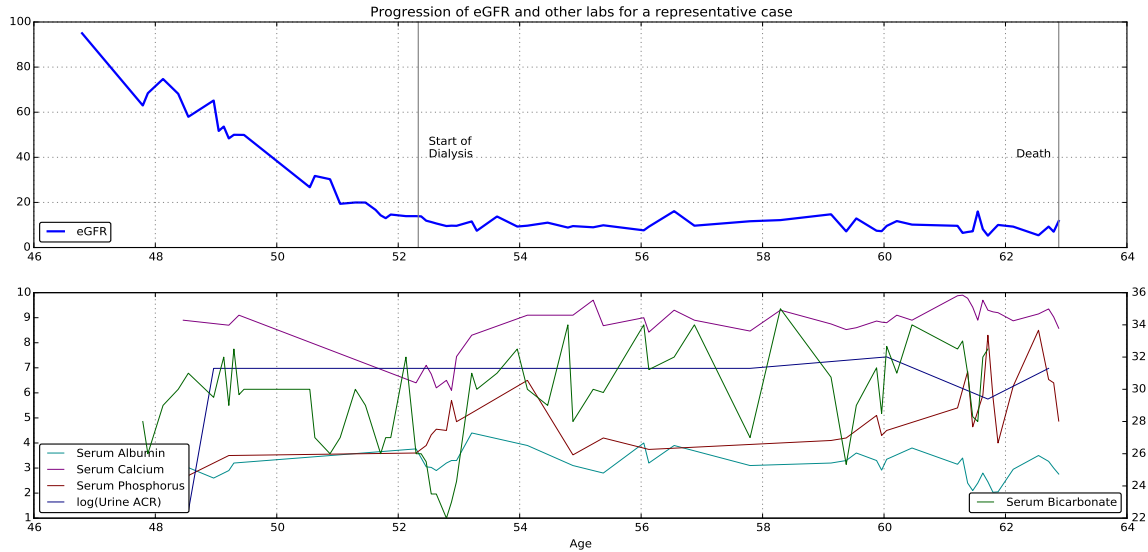
Figure 1: Clinical course of a patient who experienced a rapid progression of CKD. Top plot shows estimated glomerular filtration rate (eGFR), an estimate of kidney function (60-100 is normal, <60 indicates clinically significant kidney disease). The bottom plot shows the trajectory of five other clinical labs relevant to kidney disease.

These traits make CKD an ideal condition to develop disease progression models that can be used in high-impact care management programs. The difficulties with CKD care are best explained with a representative clinical case, illustrated in Figure 1. A 47 year-old man makes first contact with the health system with normal kidney function, although he possesses several risk factors for developing CKD. Over the next few years, his kidney function rapidly deteriorates (normal annual rate of kidney function loss at his age is only about 1-2 mL/min). However, his kidney disease goes unnoticed due to other more pronounced medical conditions, and he does not receive any treatment to slow progression to total kidney failure. At age 52 he is finally referred to a kidney specialist, but now it is too late to make advanced preparations for kidney failure, such as pre-emptive kidney transplantation or at-home dialysis, and kidney failure is inevitable. Within a few months of this first specialist appointment, he requires hospitalization for emergency dialysis initiation, a traumatic procedure that also makes him among the most expensive type of patient to treat (Johnson and et al. (2015)). While on dialysis over the next decade he suffers several cardiovascular complications before ultimately dying at age 63. A care management program utilizing individualized predictions from a disease progression model, like the one developed in this work, might have been able to act upon the multiple missed opportunities from this patient story.

Our aim in this work is to develop flexible, broadly applicable statistical models for multivariate longitudinal data. Ideally such a method would not only accurately model related time-varying clinical variables, but also leverage information from them to improve prediction of the disease trajectory of interest. For instance, in our particular application to CKD, while eGFR is the primary biomarker of interest, prediction of other labs can also

be clinically useful. To accomplish this, we utilize a hierarchical latent variable model that captures dependencies between multivariate longitudinal trajectories. Our model uses a Gaussian process (GP) with a highly structured mean function to model each longitudinal variable for each individual. The mean functions for the GPs are then made dependent through shared latent variables. Using our model, we study a large cohort of patients with CKD from the Duke University Health System EHR and make predictions about the future trajectory of their disease severity, as measured by eGFR, along with five other commonly recorded laboratory values that are known to be affected by CKD. The stochastic variational inference algorithm that we develop to fit the model scales well to the large dataset, and our model's predictions have significantly lower error than a state-of-the-art univariate method for predicting disease trajectories.

2. Related Work

Within the medical literature, the vast majority of predictive models are cross-sectional and only consider features at or up until the current time to predict outcomes at a fixed point in the future. Typically, these models only attempt to explain variability in the outcome of interest by conditioning on baseline covariates. This precludes the ability to generate dynamic individualized predictions, making them difficult to use for medical decision making in practice. For instance, Tangri and et al. (2011) is a commonly referenced Cox proportional hazards model for predicting time to kidney failure.

Within the statistics and machine learning communities, Markov models of many varieties are frequently used to generate dynamic predictions, e.g. autoregressive models, HMMs, and state space models (Murphy (2012)). However, these methods are generally only applicable in settings with discrete, regularly-spaced observation times, and in most applications the data consists of a single set of multivariate time series (e.g. financial returns), not a large collection of sparsely sampled multivariate time series (in our setting, disease or lab trajectories). GPs have been commonly been used in settings with continuous time observations, see e.g. (Shi et al. (2005)) for a thorough overview. Since they are prior distributions over functions, they are a natural modeling choice for disease trajectories, however, accurate forecasts for GPs require careful specification of the mean functions (Roberts and et al. (2013)).

There are several related works that tackle the problem of dynamic predictions for medical applications. (Rizopoulos (2011)) construct joint models with a focus on updating dynamic predictions about time to death as more values of a longitudinal biomarker are observed. They also account for individual heterogeneity using random effects. Related work on joint modeling (Proust-Lima and et al. (2014)) uses a mixture model to address heterogeneity. Although a few works from this joint modeling literature address multivariate data, typically the only dependency among variables is simplistic, coming from a shared vector of random effects. Most similar to our work is (Schulam and Saria (2015)), who present a model for a univariate marker of disease trajectory using a GP with a highly structured mean function.

There has been much recent interest in machine learning in modeling EHRs and other types of healthcare data. For instance, (Lian and et al. (2015)) use hierarchical point processes to predict hospital admissions, and (Ranganath and et al. (2015)) develop a dynamic factor model to learn relationships between diseases and predict future diagnosis codes. Some related papers use multitask GPs to model multivariate longitudinal clinical data from the

Intensive Care Unit (ICU) (Durichen and et al. (2015), Ghassemi and et al. (2015)). However, these works use independently trained models for each patient and do not hierarchically share any information across patients. This worked well in their examples, since in the ICU there is a relatively large number of observations per subject, but would not work as well with our much sparser EHR data from chronic disease patients. Closer to our work in the application is (Perotte and et al. (2015)), whose main goal is prediction of time of progression from CKD stage 3 to stage 4. Although they use a standard Kalman filter to model multivariate laboratory data, this is not the main aim of their work and it will also not be as flexible as our GP-based models. Finally, Futoma et al. (2016) present a method for jointly modeling CKD disease trajectory with adverse events. Though their application and data is similar to this work, they were primarily concerned with improving prediction of cardiac events using information from eGFR trajectories, while our focus is instead on leveraging additional longitudinal clinical variables to improve prediction of eGFR trajectories.

3. Proposed Multivariate Disease Trajectory Model

Our proposed hierarchical latent variable model jointly models each patient's multivariate longitudinal data by using a GP for each individual variable, with shared latent variables inducing dependence between the mean functions. Note that our model reduces to the method presented in (Schulam and Saria (2015)) in the univariate setting.

Let $\vec{y_i}(t) = (y_{i1}(t), \ldots, y_{iP}(t))^\top \in \mathbb{R}^P$ denote the $P-$dimensional trajectory of measurements for individual i, and let $\vec{y_{ip}} = \{y_{ip}(t_{ipj})\}_{j=1}^{n_{ip}}$ be the n_{ip} observations for variable p at times t_{ipj} for this individual. Let c_i, $\vec{z_i}$, $\vec{b_i}$, and $\vec{f_i}$ be latent variables specific to individual i, to be defined subsequently. We will assume for each person that the longitudinal variables are conditionally independent given these latent variables: $p(\vec{y_i}|c_i, \vec{z_i}, \vec{b_i}, \vec{f_i}) = \prod_{p=1}^P p(\vec{y_{ip}}|c_i, \vec{z_i}, \vec{b_i}, \vec{f_i})$.

For each longitudinal variable, we propose the following generative model:

$$y_{ip}(t) \sim N(\mu_{ip}(t), \sigma_p^2) \tag{1}$$

$$\mu_{ip}(t) \sim \mathcal{GP}(\Phi_{pop}^{(p)}(t)^\top \Lambda^{(p)} x_i + \Phi_z^{(p)}(t)^\top \beta_{z_{ip}}^{(p)} + \Phi_l^{(p)}(t)^\top b_{ip}, K_p) \tag{2}$$

$$K_p(t, t') = a_p^2 \exp\{-l_p^{-1}|t - t'|\} \tag{3}$$

$$z_{ip}|c_i \sim \text{Multinomial}(\Psi_{c_i}^{(p)}) \tag{4}$$

with priors across all P variables:

$$c_i \sim \text{Multinomial}(\pi_i), \ \pi_{ig} = \frac{e^{w_g^\top x_i}}{\sum_{g'=1}^G e^{w_{g'}^\top x_i}} \tag{5}$$

$$\vec{b_i} = (b_{i1}, \ldots, b_{iP})^\top \sim N(0, \Sigma_b). \tag{6}$$

The first term in the mean function (2) admits population-level fixed effects for each lab using observed baseline covariates $x_i \in \mathbb{R}^q$, e.g. gender, race, age. $\Lambda^{(p)} \in \mathbb{R}^{d_p \times q}$ is a coefficient matrix with $\Phi_{pop}^{(p)}(t) \in \mathbb{R}^{d_p}$ a basis expansion of time; in practice we use a linear expansion of time with $d_p = 2$ so this term allows for a fixed intercept and slope.

The second term in (2) is a subpopulation component, where it is assumed individual i's trajectory for variable p belongs to a latent subpopulation, denoted $z_{ip} \in \{1, \ldots, G_p\}$. Each

subpopulation is associated with a unique trajectory; in particular, $\Phi_z^{(p)}(t) \in \mathbb{R}^{d_z}$ is a fixed B-spline basis expansion of time (for simplicity, assumed to be the same for each variable: degree two, with the same eight interior knots evenly spaced in time) with $\beta_g^{(p)} \in \mathbb{R}^{d_z}$ the coefficient vector for subpopulation g and variable p.

The prior for each z_{ip} in (4) depends on the individual's "global" cluster $c_i \in \{1, \ldots, G\}$. Drawing on ideas from latent structure analysis for multivariate categorical data (Lazarsfeld and Henry (1968)), the c_i induce dependence among the P trajectory-specific clusters z_{ip}, as they have conditionally independent multinomial priors. Each of the G columns $\Psi_g^{(p)}$ in the matrix $\Psi^{(p)} \in \mathbb{R}^{G_p \times G}$ defines a distribution over the G_p values that z_{ip} can take. The c_i then has a multinomial logistic regression prior in (5) that depends on the baseline covariates x_i, where $\{w_g\}_{g=1}^G$ are regression coefficients with $w_1 \equiv 0$ for identifiability.

The third term in (3) is a random effects component, allowing for individual-specific long-term deviations in trajectory that are learned dynamically as more data becomes available. In practice $\Phi_l(t) \in \mathbb{R}^{d_l}$ is a linear expansion of time with $d_l = 2$, so that $b_{ip} \in \mathbb{R}^{d_l}$ is a random slope and intercept vector for patient i. The overall vector $\vec{b_i}$ has a multivariate normal prior distribution in (6), making the random effects dependent across labs.

Finally, $K_p(t, t') = a_p^2 \exp\{-l_p^{-1}|t - t'|\}$ is the Ornstein-Uhlenbeck (OU) covariance function for the GP each variable, with parameters a_p, l_p. This kernel is well-suited for modeling transient deviations from the mean function, as it is mean-reverting and has no long-range dependence between deviations. We can rewrite (2) as $\mu_{ip}(t) = \Lambda^{(p)} x_i + \Phi_z(t)^\top \beta_{z_{ip}} + \Phi_l(t)^\top b_{ip} + f_{ip}(t)$, where $f_{ip}(t) \sim \mathcal{GP}(0, K_p)$, so that we can explicitly represent the short-term deviations f_{ip} from the GP in order to learn them during inference.

4. Inference

The computational problem associated with fitting our model is estimation of the posterior distribution of latent variables and model parameters given the observed data. As is common with complex probabilistic generative models, exact computation of the posterior is intractable and requires approximation to compute. To this end, we develop a mean field variational inference [Jordan et al., 1999] algorithm to find an approximation to the posterior.

Variational inference converts the task of posterior inference into an optimization problem of finding a distribution q in some approximating family of distributions that is close in KL divergence to the true posterior. The problem is equivalent to maximizing what is known as the evidence lower bound (ELBO) [Bishop, 2006]:

$$\mathcal{L}(q) = E_q[\log p(y, z, b, f, c, \Theta) - \log q(z, b, f, c, \Theta)], \tag{7}$$

which forms a lower bound on the marginal likelihood $p(y)$ of our model.

VARIATIONAL APPROXIMATION

The model parameters to be learned are $\Theta = \left\{\{\Lambda^{(p)}, \beta^{(p)}, \Psi^{(p)}, a_p, l_p, \sigma_p^2\}_{p=1}^P, W, \Sigma_b\right\}$, and the local latent variables specific to each person are their global cluster assignment c_i, subpopulation assignments z_{ip}, random effects $\vec{b_i}$ and structured noise functions f_{ip}. The

joint distribution for our model can be expressed as:

$$p(y, z, b, f, c, \Theta) = p(\Theta) \prod_{i=1}^{N} p(b_i|\Theta) p(c_i|\Theta) \prod_{p=1}^{P} p(\vec{y_{ip}}|z_{ip}, b_{ip}, f_{ip}, \Theta) p(z_{ip}|c_i, \Theta) p(f_{ip}|\Theta) \quad (8)$$

We make the mean field assumption for the variational distribution, which assumes that in the approximate posterior q, all the latent variables are independent. This implies that $q(z, b, f, v, \Theta) = q(\Theta) \prod_{i=1}^{N} q_i(c_i, z_i, b_i, f_i)$, where:

$$q_i(c_i, z_i, b_i, f_i) = q_i(c_i|\nu_{c_i}) q_i(b_i|\mu_{b_i}, \Sigma_{b_i}) \prod_{p=1}^{P} q_i(z_{ip}|\nu_{z_{ip}}) q_i(f_{ip}|\mu_{f_{ip}}, \Sigma_{f_{ip}}) \quad (9)$$

The assumed variational distributions for each latent variable are in the same family as their prior distribution, i.e. c_i and z_{ip} are multinomials, and b_i is multivariate normal. For f_{ip} we use a multivariate normal evaluated at all times at which variable p is observed for patient i. To evaluate f_{ip} at additional times e.g. during model evaluation we use the conditional GP framework and treat the observed values as pseudo-inputs, from the sparse GP literature (Titsias (2009)). Finally, for all model parameters Θ we learn a point estimate, so their variational distributions are delta functions. We impose vague normal priors on $\Lambda^{(p)}, \beta^{(p)}$, and W, a uniform prior on $\Psi^{(p)}$, and learn MLEs for other parameters. Thus, the goal of our variational algorithm is to learn optimal variational parameters $\lambda_i = \{\nu_{c_i}, \nu_{z_{ip}}, \mu_{b_i}, \Sigma_{b_i}, \mu_{f_{ip}}, \Sigma_{f_{ip}}\}$ for each individual i, as well as a point estimate $\hat{\Theta}$ for model parameters. In practice, we optimize Cholesky decompositions of covariance matrices.

SOLVING THE OPTIMIZATION PROBLEM

We use the automatic differentiation package *autograd* [1] in Python to compute exact gradients in order to optimize the lower bound, since the lower bound has an analytic closed-form expression. At each iteration of the algorithm, we optimize the local variational parameters using exact gradients. To optimize the global parameters, we turn to stochastic optimization, a commonly used tool in variational inference. Rather than using the entire large dataset to compute exact gradients of the ELBO with respect to Θ, we can compute an unbiased noisy gradient based on a sampled batch of observations (Hoffman et al. (2013)). To set the learning rate we use RMSProp, which adaptively allows for a different learning rate for each parameter (Tieleman and Hinton (2012)).

5. Empirical Study

In this section we describe our experimental setup and results on our dataset.

DATASET

Our dataset contains laboratory values from 44,519 patients with stage 3 CKD or higher extracted from the Duke University Health System EHR. IRB approval (#Pro00066690) was obtained for this work. We started with an initial cohort of roughly 600,000 patients

---

1. https://github.com/HIPS/autograd

that had at least one encounter in the health system in the year prior to Feb. 1, 2015. This includes inpatient, outpatient, and emergency department visits over a span of roughly 20 years. From this, we filtered to patients who had at least five recorded values for serum creatinine, the laboratory value we used to calculate eGFR. Finally, we selected patients that had Stage 3 CKD or higher, indicative of moderate to severe kidney damage and defined as two eGFR measurements less than 60 mL/min separated by at least 90 days.

We also choose to model five related lab values that have important clinical significance for CKD. The first, serum albumin, is an overall marker of health and nutrition. The second, serum bicarbonate, can indicate acid accumulation from inadequate acid elimination by the kidney. The third, serum calcium, can indicate improperly functioning kidneys if levels are too high. The fourth, serum phosphorus, can indicate phosphorus accumulation due to inadequate elimination by the kidney, and is associated with cardiovascular death and bone disorders. Finally, urine albumin to creatinine ratio (ACR) is a risk factor and cause of kidney failure (we use a log transform following common practice). While all 44,519 patients have at least five eGFR measurements, there are 884, 242, 78, 16159, and 4321 patients who have no recorded values for the other labs, and the median number of measurements for each lab (among patients with at least one) is 14, 8, 11, 12, 3, and 4, respectively.

As a final preprocessing step, since the recorded eGFR values are extremely noisy and eGFR is only a valid estimate of kidney function at steady state, we take the mean of eGFR readings in monthly time bins for each patient. We also do this to the other lab values, to reduce the overall noise in short time spans. Future work will more explicitly model periods of rapid fluctuation and high variance as they may be related to long term risk. In order to align the patients on a common time axis, for each patient we fix $t = 0$ to be their first recorded eGFR reading below 60 mL/min. The baseline covariates used for x_i were baseline age, race and gender, and indicator variables for hypertension and diabetes, as well as an overall intercept.

EVALUATION

After learning a point estimate for the global model parameters during training, they are held fixed. Then, an approximate posterior over the local latent variables is learned for each patient in the held-out test set. Predictions about future lab values are made by drawing samples from the approximate posterior predictive. We compare our method with the method of (Schulam and Saria (2015)), trained independently to each of the 6 labs. For each test patient, we learn their parameters three times, using data up until times $t = 1$, $t = 2$, and $t = 4$, each time recording the mean absolute error of predictions for each lab in future time windows. These values are then averaged over all patients in the test set, and finally averaged over 10 cross validation folds to produce Table 1. We use the 10 fold cross validation with one-sided, paired t-tests to test for significant improvements in performance.

RESULTS

Table 1 displays the results of both methods. We set $G = G_p = 10$, but results were not sensitive to this choice. Our proposed method generally outperforms (Schulam and Saria (2015)), especially when less data is used to make predictions (e.g. $t = 1$). This makes sense, as our method is able to learn correlations between variables to improve predictions.

Table 1: Mean Absolute Errors across all labs from 10 fold cross validation. Bold indicates p-value from one-sided, paired t-test comparing methods was $< .05$. *,**,*** indicate $p < .01, < .001, < .0001$, respectively.

Predictions with data up to...		t = 1				t = 2			t = 4	
Lab	Model	(1, 2]	(2, 4]	(4, 8]	(8, 19]	(2, 4]	(4, 8]	(8, 19]	(4, 8]	(8, 19]
eGFR	Schulam	8.84	10.36	12.04	13.78	8.82	11.10	13.17	9.33	12.15
	Proposed	**8.76****	**10.18****	**11.79****	13.68	**8.67****	**10.99***	13.13	9.34	12.18
Serum Alb.	Schulam	0.59	0.79	1.06	1.49	0.60	0.87	1.27	0.62	0.96
	Proposed	**0.32****	**0.37****	**0.43****	**0.56****	**0.33****	**0.42****	**0.56****	**0.38****	**0.52****
Serum Bicarb.	Schulam	1.91	2.02	2.13	2.27	1.89	2.04	2.20	1.90	**2.12**
	Proposed	1.85	1.98	2.09	2.29	1.88	2.06	2.29	1.94	2.26
Serum Calc.	Schulam	0.71	1.02	1.56	2.78	0.72	1.21	2.28	0.81	1.55
	Proposed	**0.36****	**0.41****	**0.49****	**0.61****	**0.38****	**0.47****	**0.61****	**0.42 ****	**0.57****
Serum Phos.	Schulam	1.02	1.28	1.44	1.40	1.11	1.33	1.29	1.12	1.15
	Proposed	**0.66****	**0.81***	**1.09**	1.62	**0.72****	**1.01**	1.46	**0.84***	1.27
Urine ACR	Schulam	1.15	1.30	1.45	1.62	1.14	1.32	1.52	1.14	1.37
	Proposed	**0.84****	**0.96****	**1.11***	**1.40**	**0.89****	**1.09***	1.40	**0.99***	1.31

Further, in cases where the methods had to predict future labs for a patient that did not yet have any observed values for that lab, (Schulam and Saria (2015)) could only predict using baseline covariates, while our multivariate approach could also leverage information from the other related labs. Current clinical practice for managing care of CKD patients uses clinical judgment alone to predict future disease status. Incorporation of a model such as ours to predict eGFR trajectory and other labs would provide a useful tool for providers to assess the risk of future decline in kidney function.

6. Discussion

In this paper, we proposed a flexible model for multivariate longitudinal data, and applied it to disease trajectory modeling in CKD. Our model yields good performance on the task of predicting future kidney function and related lab values. Although our work is a promising early work for developing machine learning models from EHR data and applying them to real clinical tasks, many inherent limitations to working with operational EHR data (Hersh and et al (2013)) must be overcome in order for the methods to be widely used in practice.

There are many interesting avenues for future work. We plan to consider models for other diseases, including diabetes and cardiovascular disease, which are frequently comorbid with CKD. Incorporation of a larger number of longitudinal variables will require care in order to ensure tractability, especially if we choose to model frequently recorded vitals. Another interesting direction to consider is joint models for adverse event data, such as in (Futoma et al. (2016)). Jointly modeling multiple event processes (e.g. emergency department visits, heart attacks, strokes) will allow us to learn correlations between different types of events. Given that much of the data recorded in the EHR is in the form of administrative billing codes, future work should incorporate these into the models as well, perhaps in an unsupervised fashion. Finally, models incorporating additional outcomes such as medical costs, hospitalizations, and patient quality of life are of significant practical interest.

The screenshot in Figure 2 shows a prototype of the rounding tool being developed with Duke Connected Care, where our predictions will help flag high risk patients for chart review during rounding sessions. Deployment of a flexible, scalable predictive tool such as ours offers ACOs the opportunity to transform the way care is provided to large patient populations.
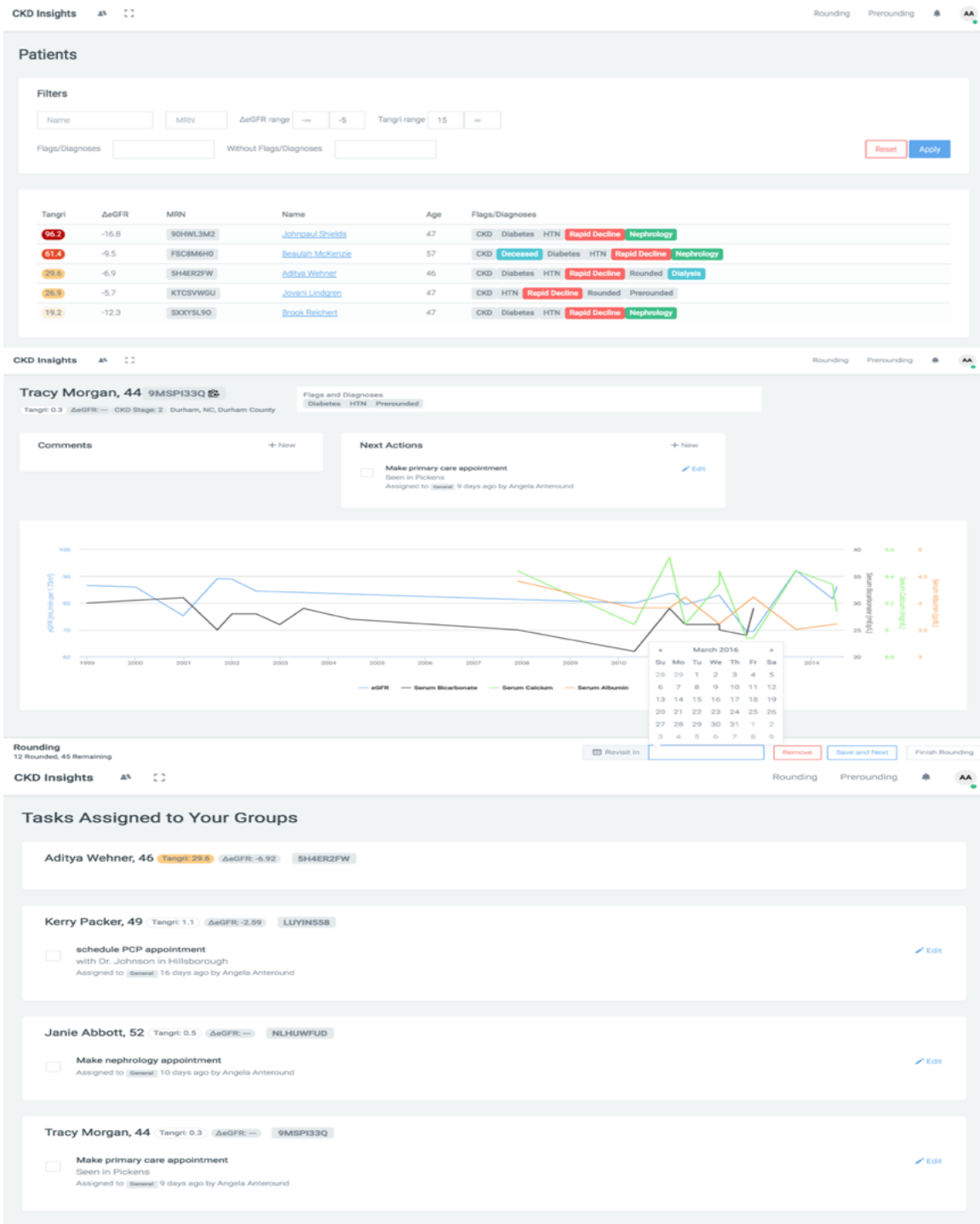
Figure 2: Snapshots from our CKD rounding application (with synthetic data). The top panel shows a pre-rounding table of patients to be rounded on, along with risk scores and appropriate flags. The middle panel displays patient data and other relevant information so that this patient's care can be efficiently managed and an appropriate intervention made, if applicable. The bottom panel shows a list of tasks assigned to each group present at rounds.

Acknowledgments

References

CMS. Quality payment program: Delivery system reform, medicare payment reform, and macr, 2016. URL `https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-MIPS-and-APMs.html`.

R. Durichen and M. Pimentel et al. Multitask gaussian processes for multivariate physiological time-series analysis. In *IEEE Transactions on Biomedical Engineering*, pages 314–322, 2015.

J. Futoma, M. Sendak, C. Cameron, and K. Heller. Scalable joint modeling of longitudinal and point process data for disease trajectory prediction and improving management of chronic kidney disease. In *UAI, to appear*, 2016.

M. Ghassemi and M. Pimentel et al. A multivariate timeseries modeling approach to severity of illness assessment and forecasting in icu with sparse, heterogeneous clinical data. In *AAAI*, 2015.

W. Hersh and M. Weiner et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. In *Medical Care*, 2013.

M. Hoffman, D. Blei, C. Wang, and J. Paisley. Stochastic variational inference. In *JMLR*, pages 1303–1347, 2013.

T. Johnson and D.Rinehart et al. For many patients who use large amounts of healthcare services, the need is intense yet temporary. In *Health Affairs*, 2015.

P. Lazarsfeld and N. Henry. Latent structure analysis. In *Houghton Mifflin*, 1968.

A. Levey and L. Stevens et al. A new equation to estimate glomerular filtration rate. In *Annals of Internal Medicine*, pages 604–612, 2009.

W. Lian and R. Henao et al. A multitask point process predictive model. In *ICML*, 2015.

D. Mendelssohn, B. Curtis, and K. Yeates et al. Suboptimal initiation of dialysis with and without early referral to a nephrologist. In *Nephrology Dialysis Transplantation*, 2011.

K. Murphy. Machine learning: A probabilistic perspective. In *MIT Press*, 2012.

R. Parikh, M. Kakad, and D. Bates. Integrating predictive analytics into high-value care: The dawn of precision delivery. In *JAMA*, pages 651–652, 2016.

A. Perotte and R. Ranganath et al. Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis. In *JAMIA*, 2015.

C. Proust-Lima and et al. Joint latent class models for longitudinal and time-to-event data: A review. In *Statistical Methods in Medical Research*, 2014.

R. Ranganath and A. Perotte et al. The survival filter: Joint survival analysis with a latent time series. In *UAI*, 2015.

D. Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. In *Biometrics*, 2011.

S. Roberts and M. Osborne et al. Gaussian processes for time series modeling. In *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2013.

S. Saria and A. Goldenberg. Subtyping: What is it and its role in precision medicine. In *IEEE Intelligent Systems*, 2015.

P. Schulam and S. Saria. A framework for individualizing predictions of disease trajectories by exploiting multi-resolution structure. In *NIPS*, 2015.

J. Shi, R. Murray-Smith, and D. Titterington. Hierarchical gaussian process mixtures for regression. In *Statistics and Computing*, 2005.

N. Smart, G. Dieberg, M. Ladhanni, and T. Titus. Early referral to specialist nephrology services for preventing the progression to end-stage kidney disease. In *Cochrane Database of Systematic Reviews*, 2014.

L. Szczech, R. Stewart, and H. Su et al. Primary care detection of chronic kidney disease in adults with type-2 diabetes: The add-ckd study. In *PLoS ONe*, 2014.

N. Tangri and L. Stevens et al. A predictive model for progression of chronic kidney disease to kidney failure. In *JAMA*, pages 1553–1559, 2011.

T. Tieleman and G. Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude, 2012.

M. Titsias. Variational learning of inducing variables in sparse gaussian processes. In *AISTATS*, 2009.

D. Tuot, L. Plantinga, and C. Hsu et al. Chronic kidney disease awareness among individuals with clinical markers of kidney dysfunction. In *Clinical Journal of the American Society of Nephrology*, 2011.