

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

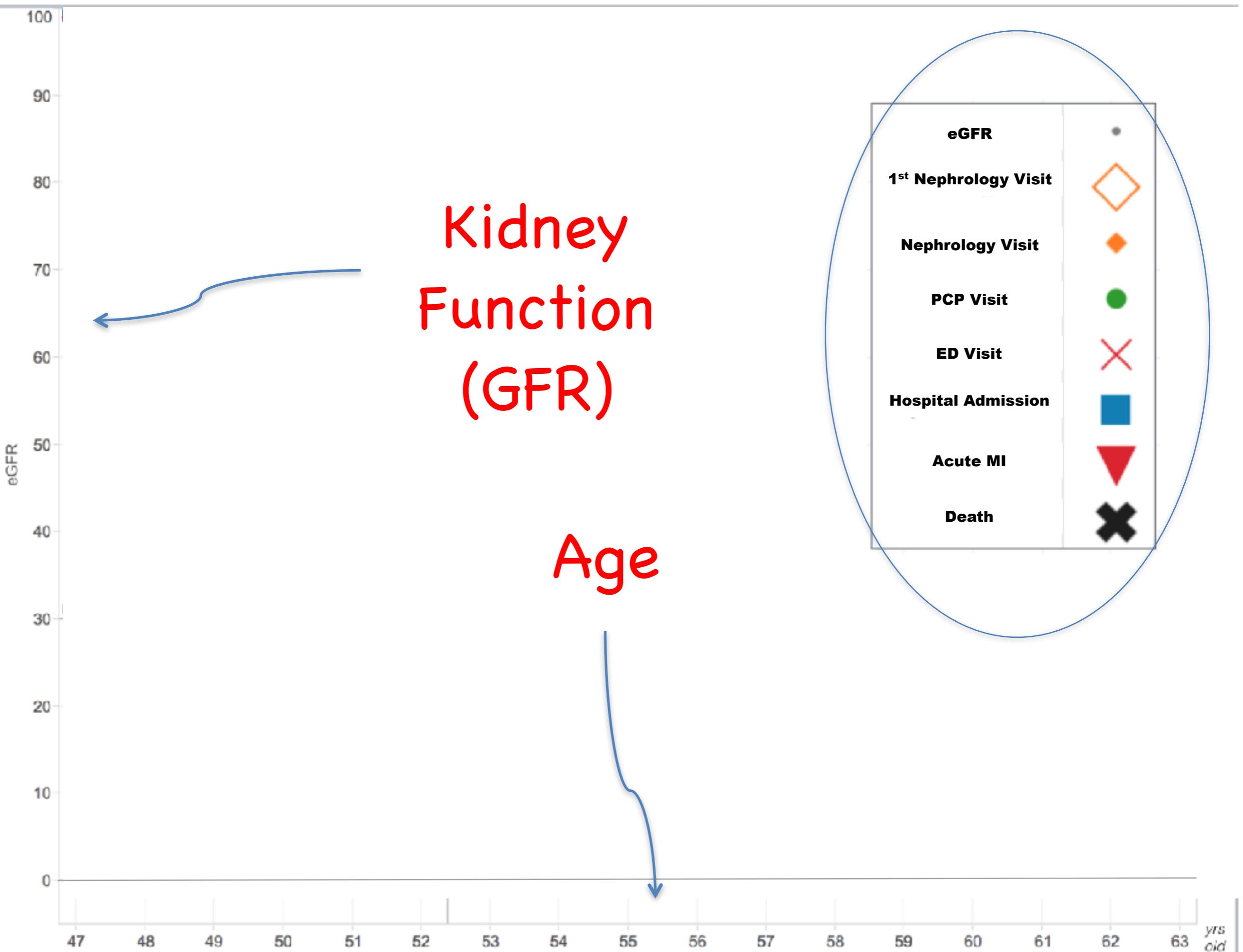


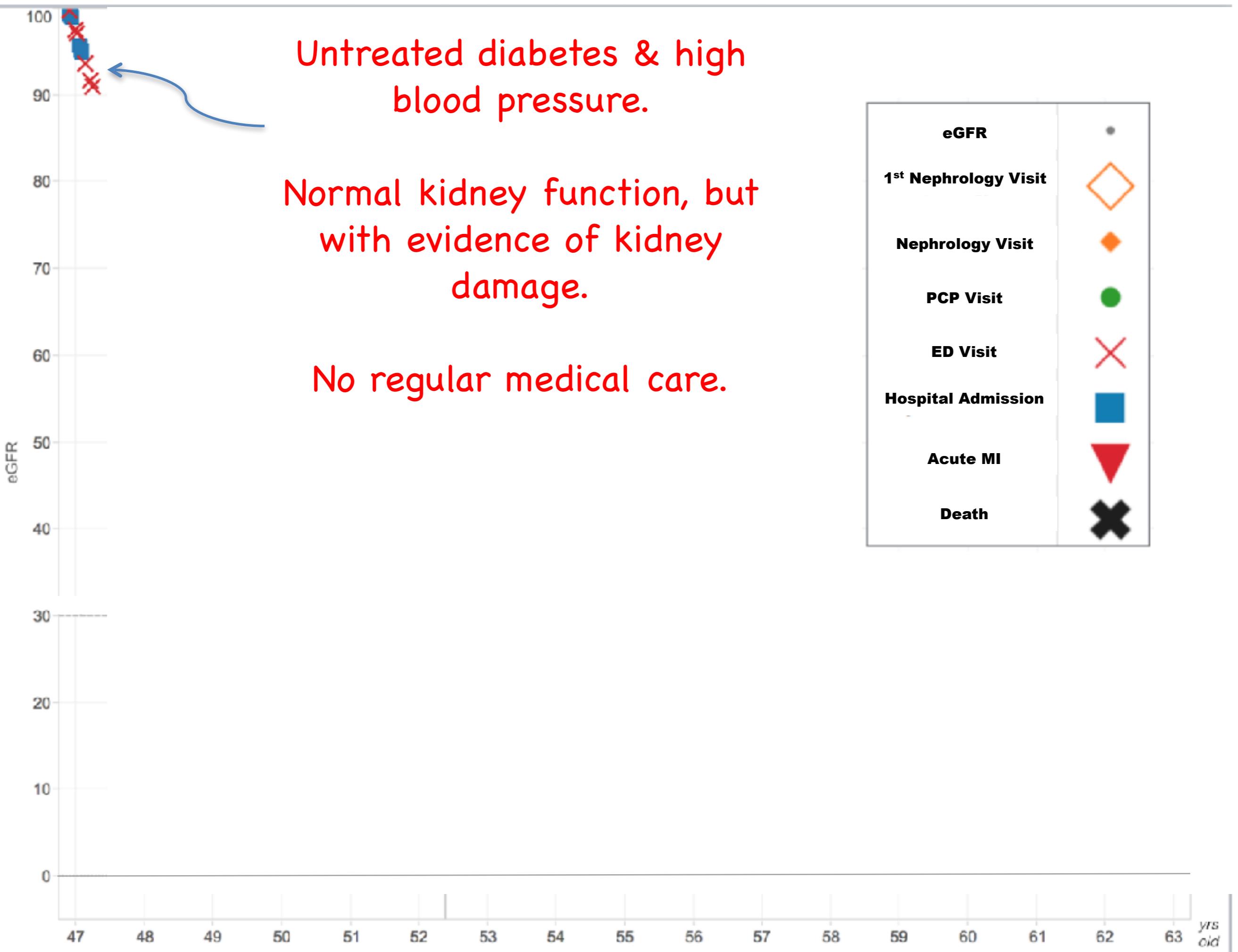
Outline

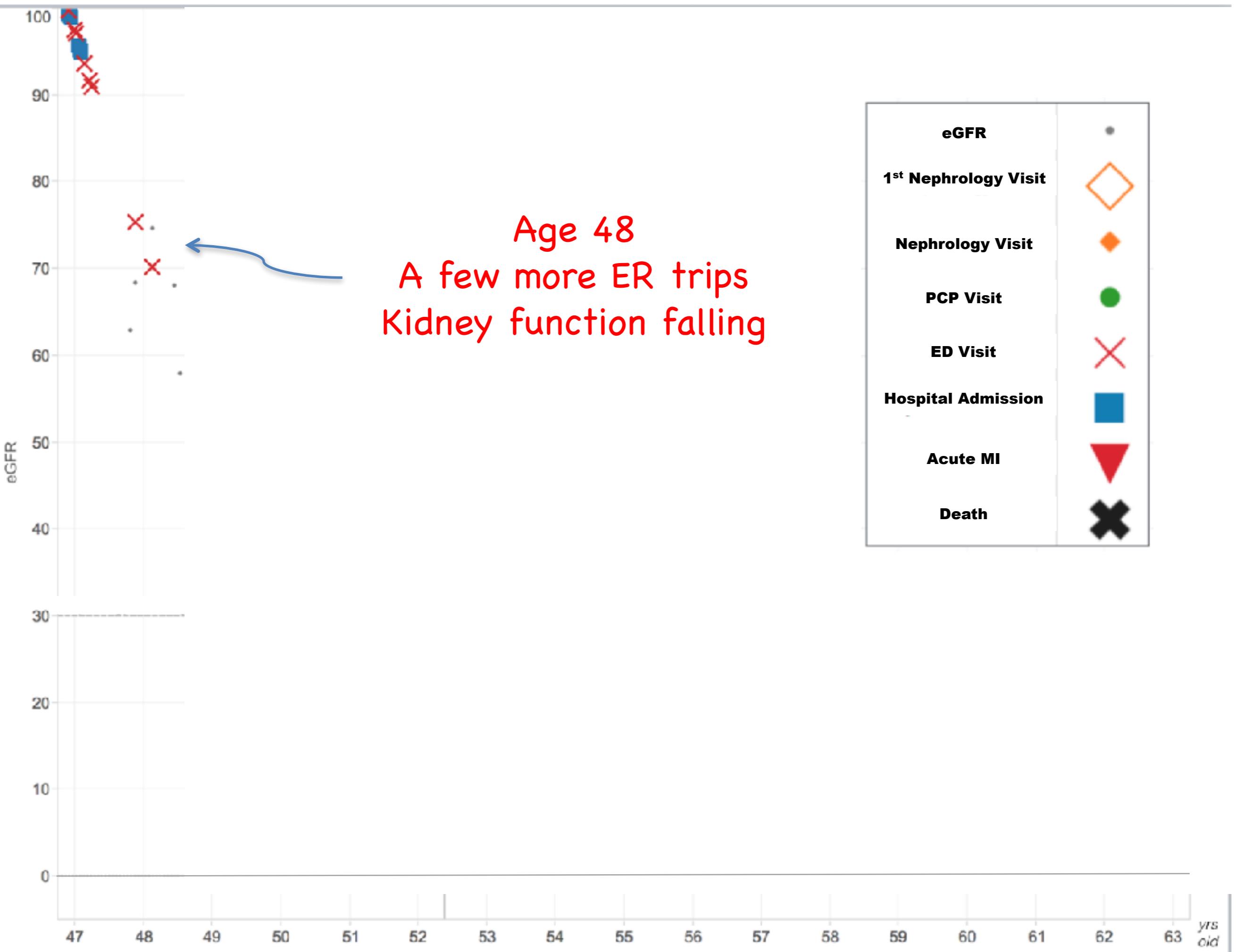
- ❖ Motivation
- ❖ Working with EHR Data
- ❖ Proposed Joint Model
- ❖ Experiments & Results
- ❖ In Clinical Practice

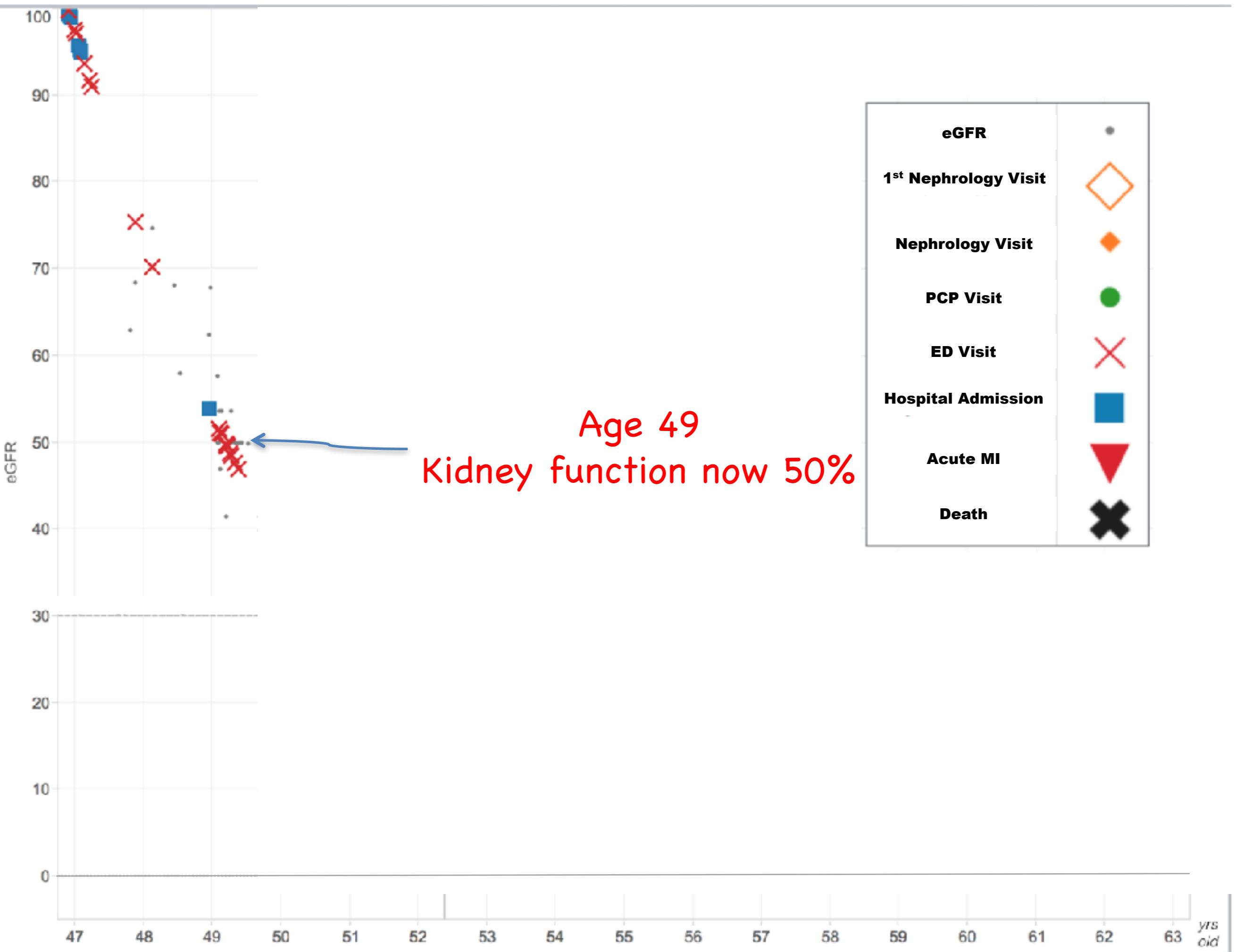
Kidney Function (GFR)

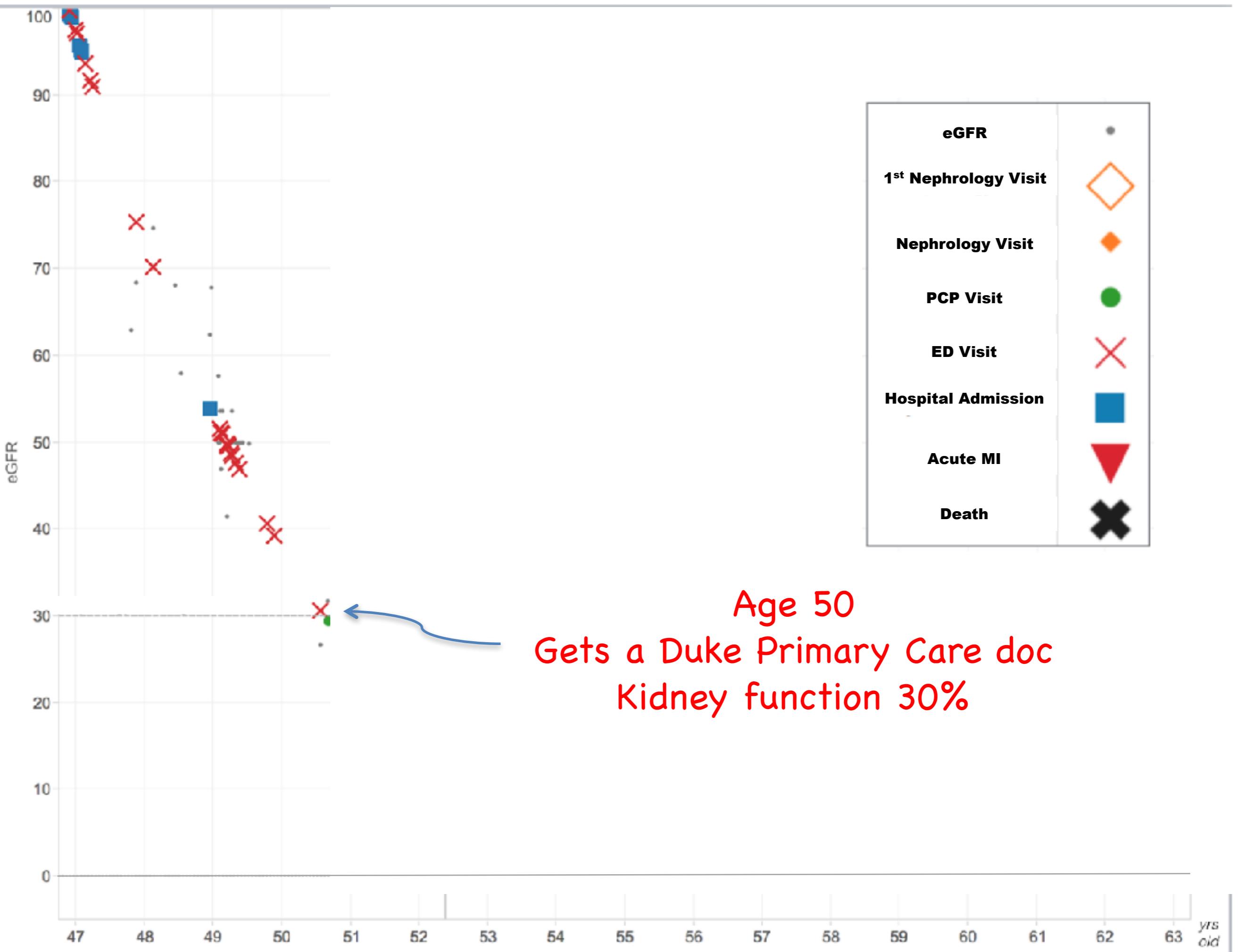
Age

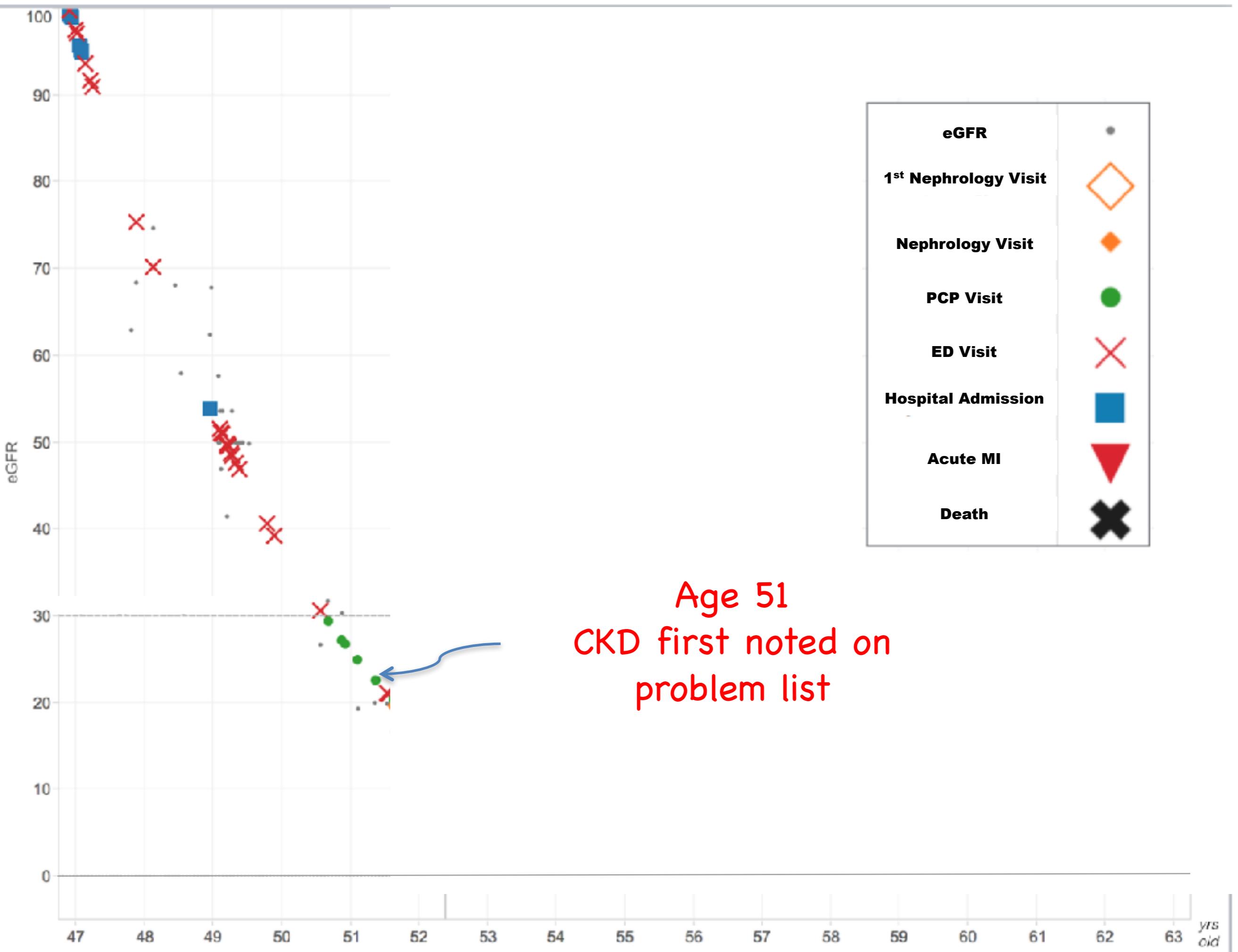


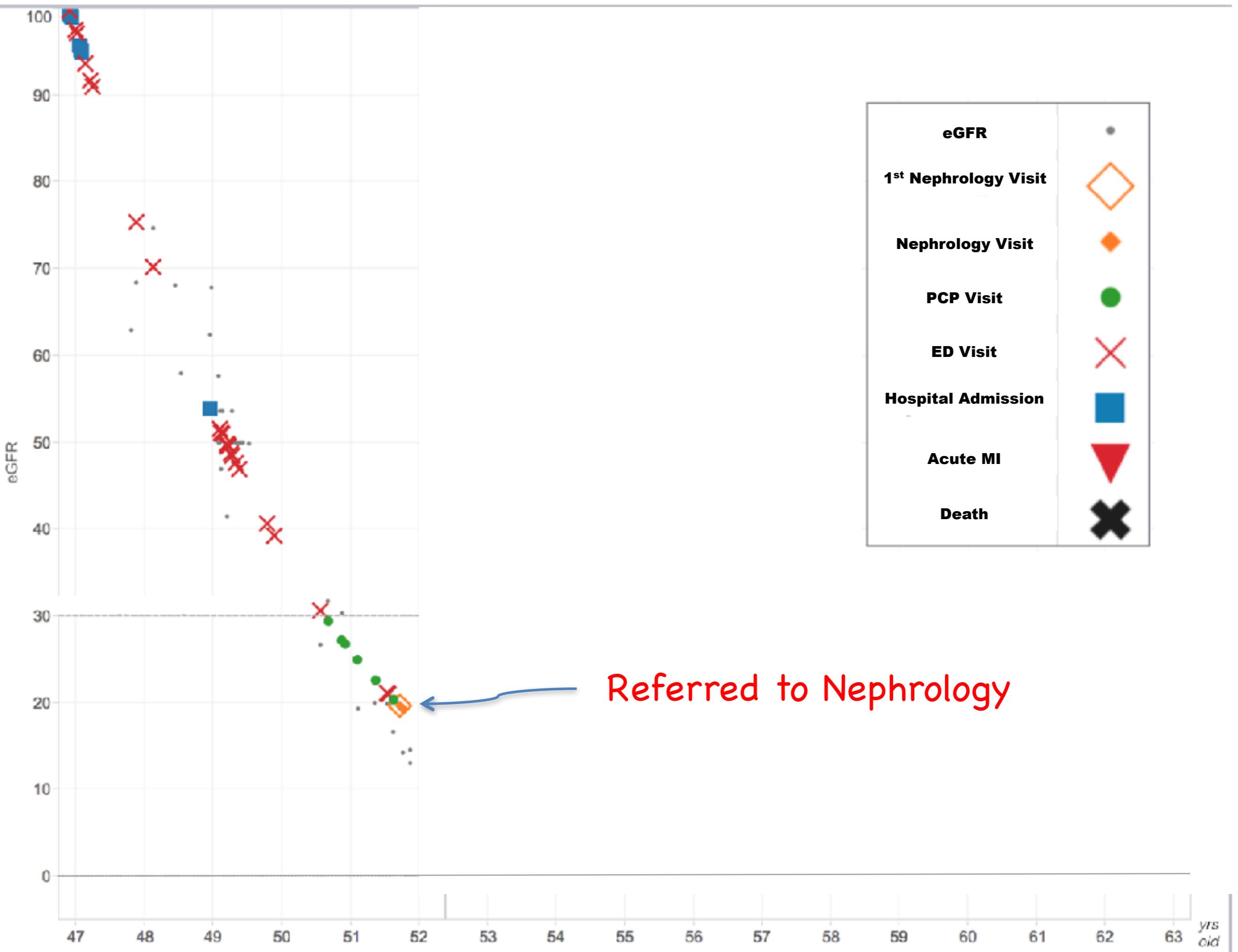


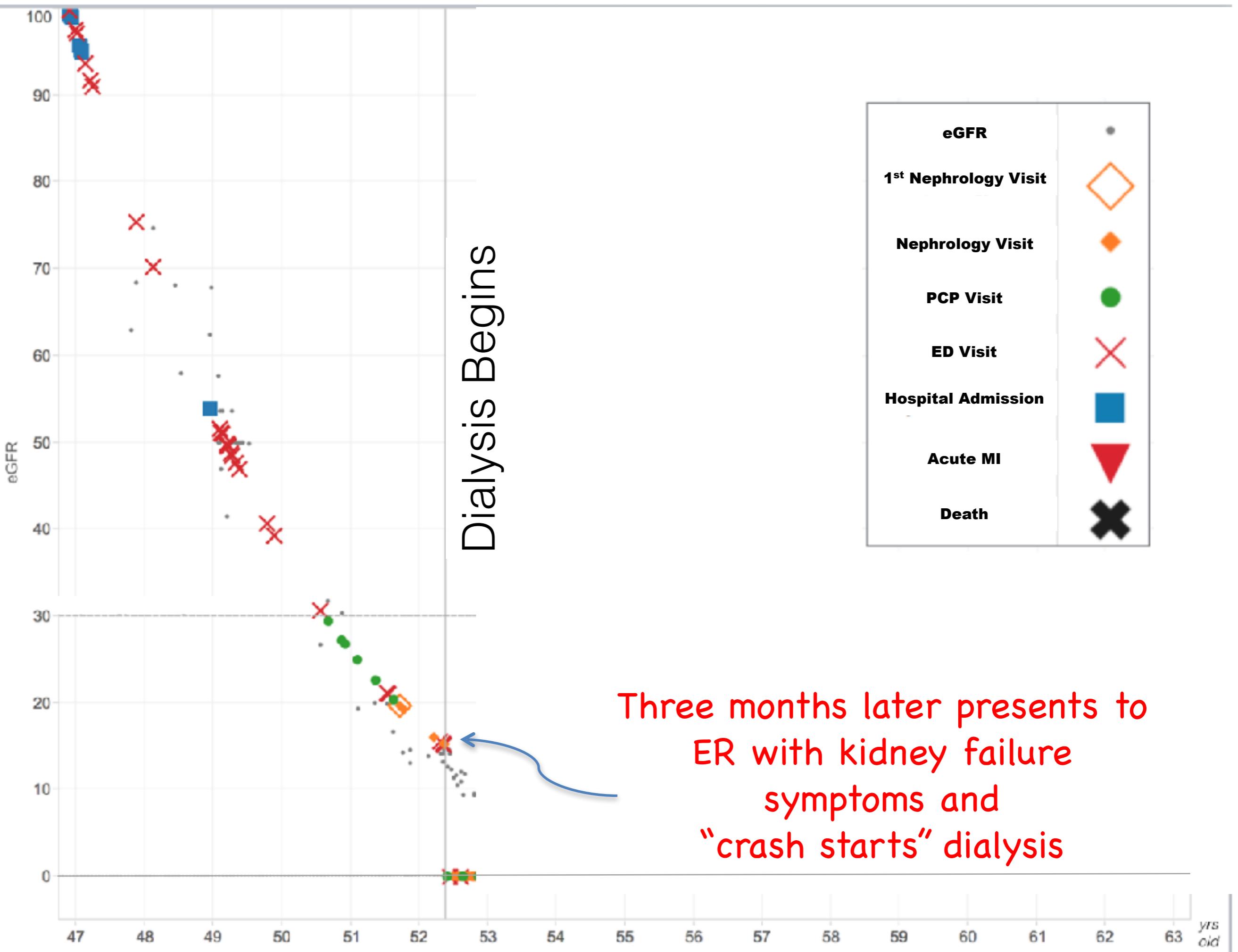


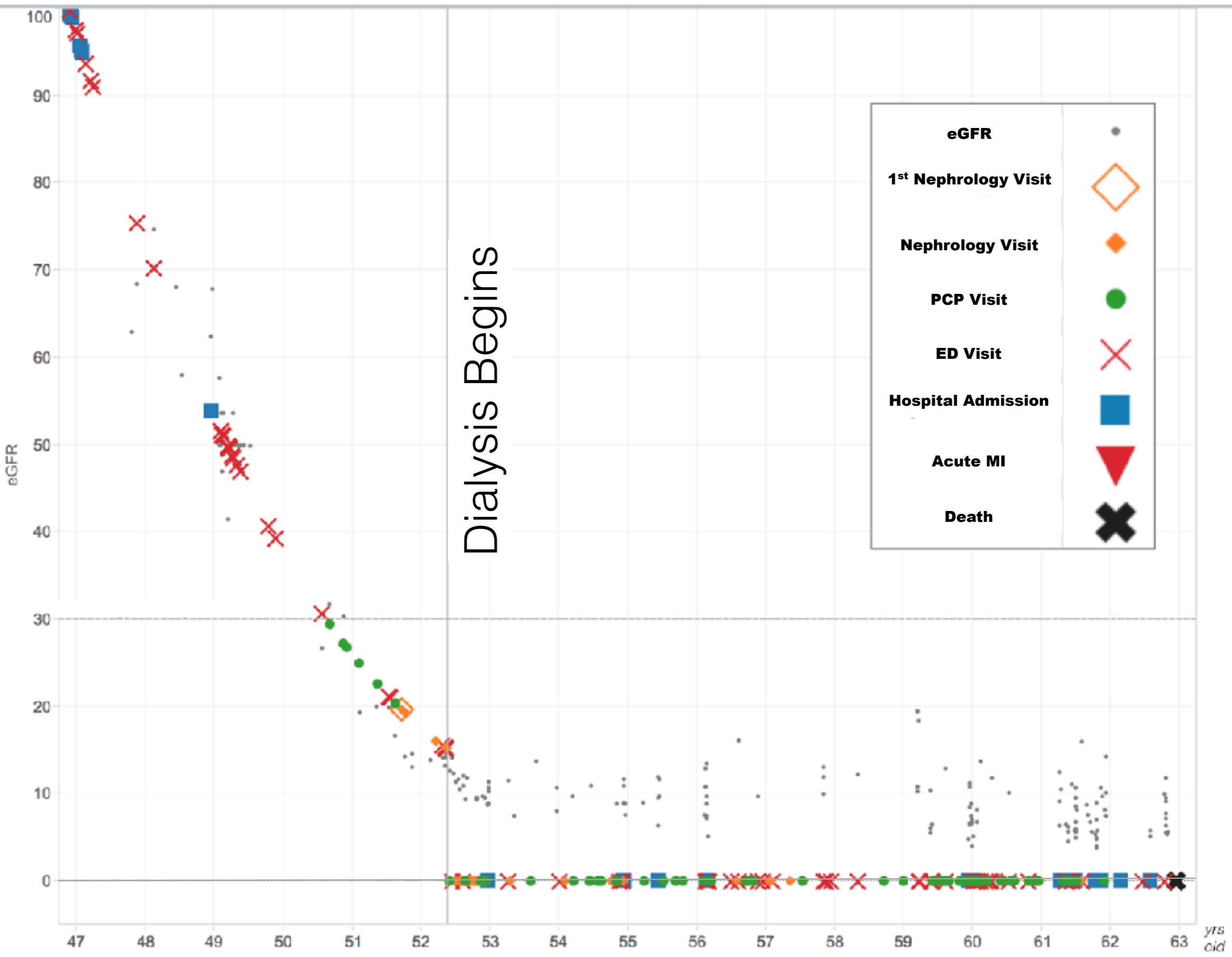


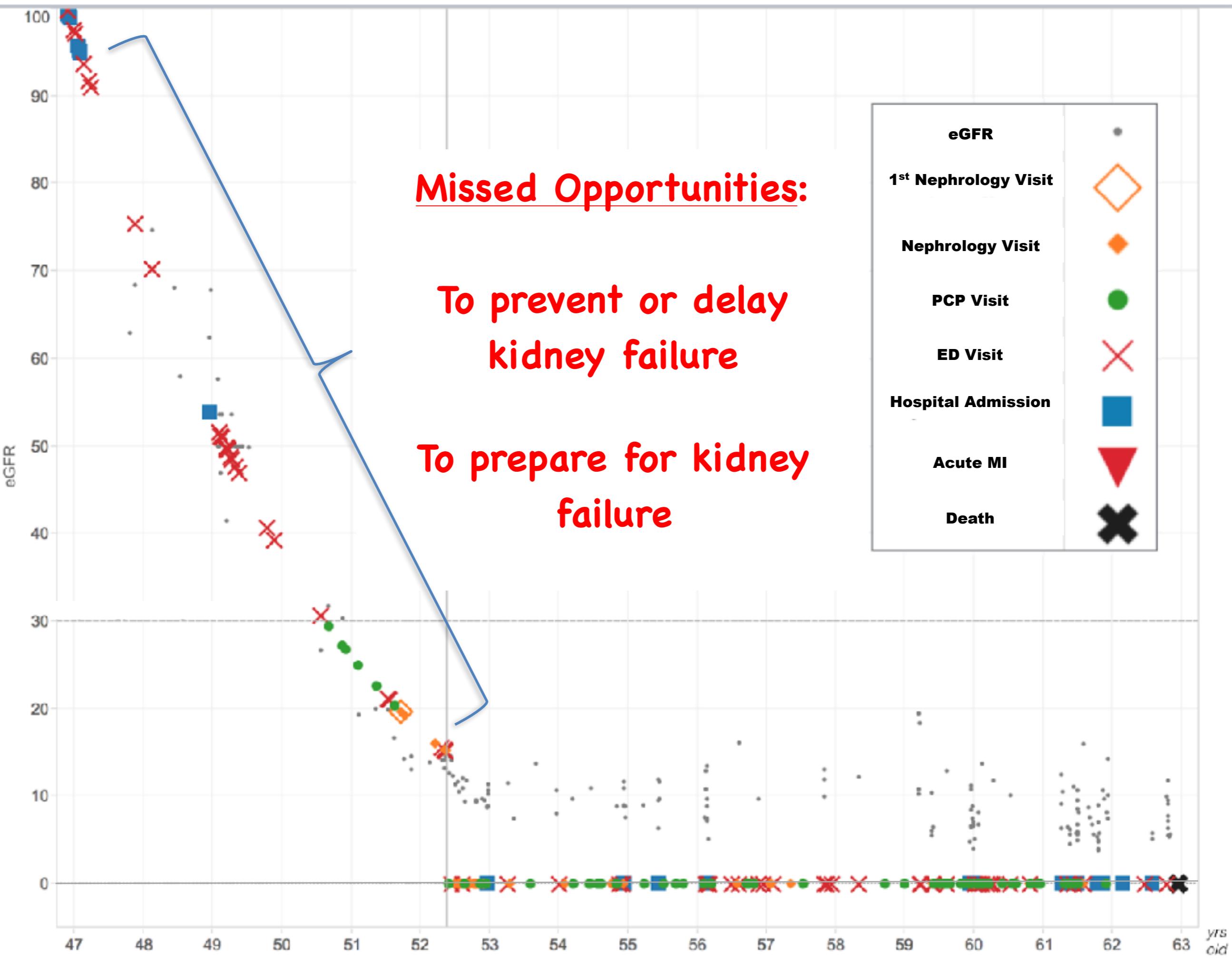


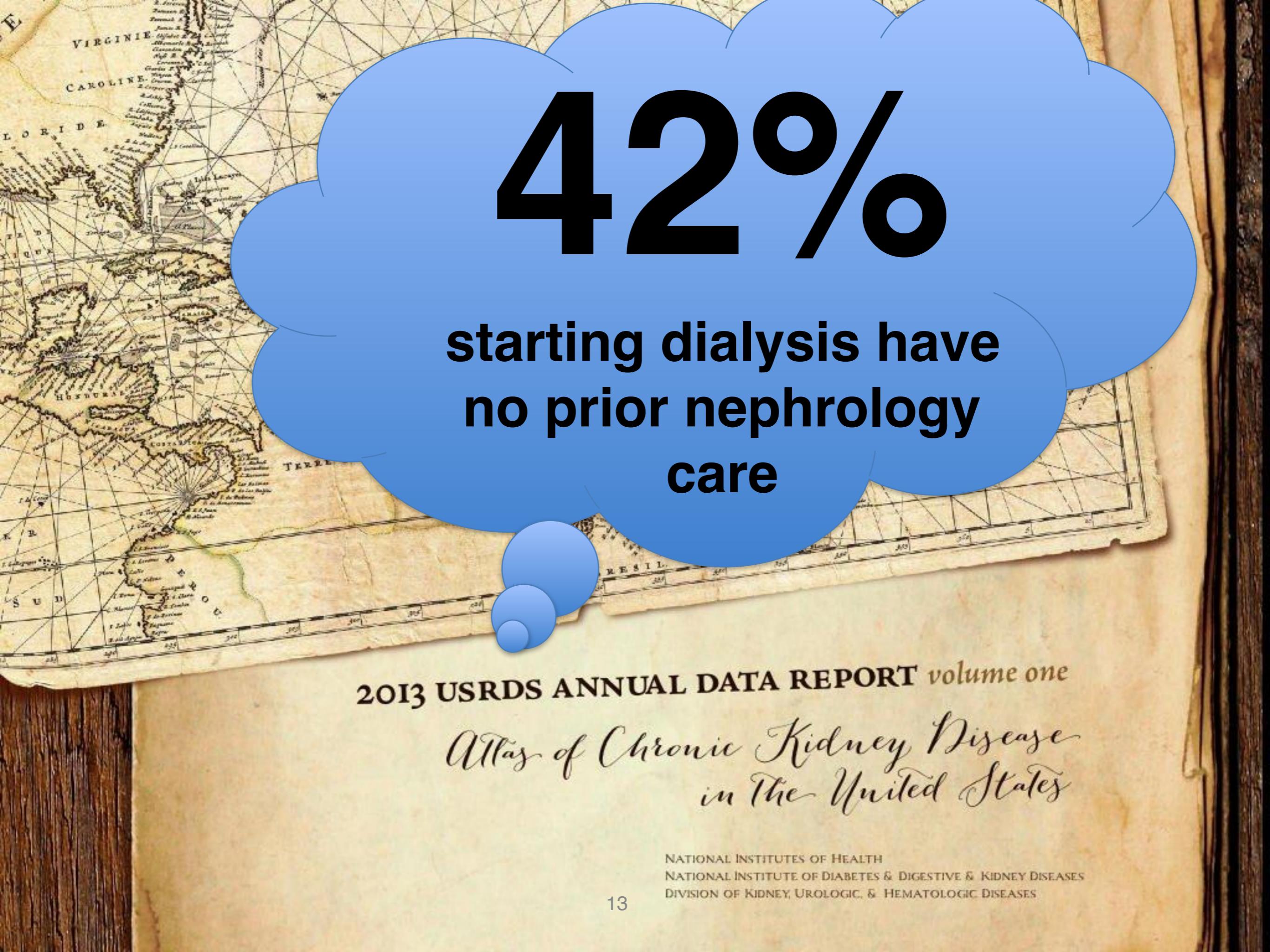












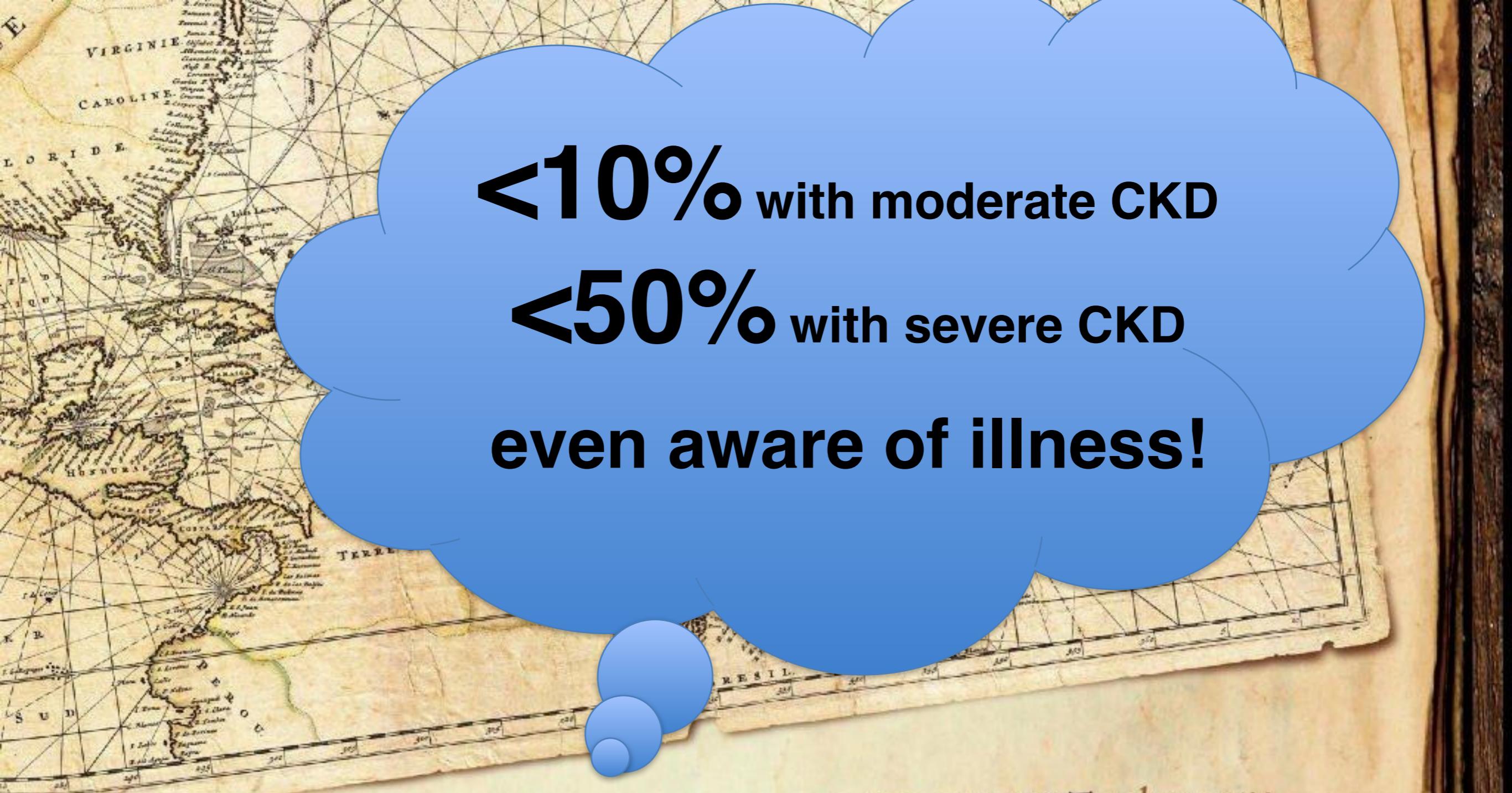
42%

starting dialysis have
no prior nephrology
care

2013 USRDS ANNUAL DATA REPORT volume one

*Atlas of Chronic Kidney Disease
in the United States*

NATIONAL INSTITUTES OF HEALTH
NATIONAL INSTITUTE OF DIABETES & DIGESTIVE & KIDNEY DISEASES
DIVISION OF KIDNEY, UROLOGIC, & HEMATOLOGIC DISEASES



<10% with moderate CKD

<50% with severe CKD

even aware of illness!

2013 USRDS ANNUAL DATA REPORT volume one

*Atlas of Chronic Kidney Disease
in the United States*

NATIONAL INSTITUTES OF HEALTH

NATIONAL INSTITUTE OF DIABETES & DIGESTIVE & KIDNEY DISEASES

DIVISION OF KIDNEY, UROLOGIC, & HEMATOLOGIC DISEASES

Chronic Kidney Disease (CKD)

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

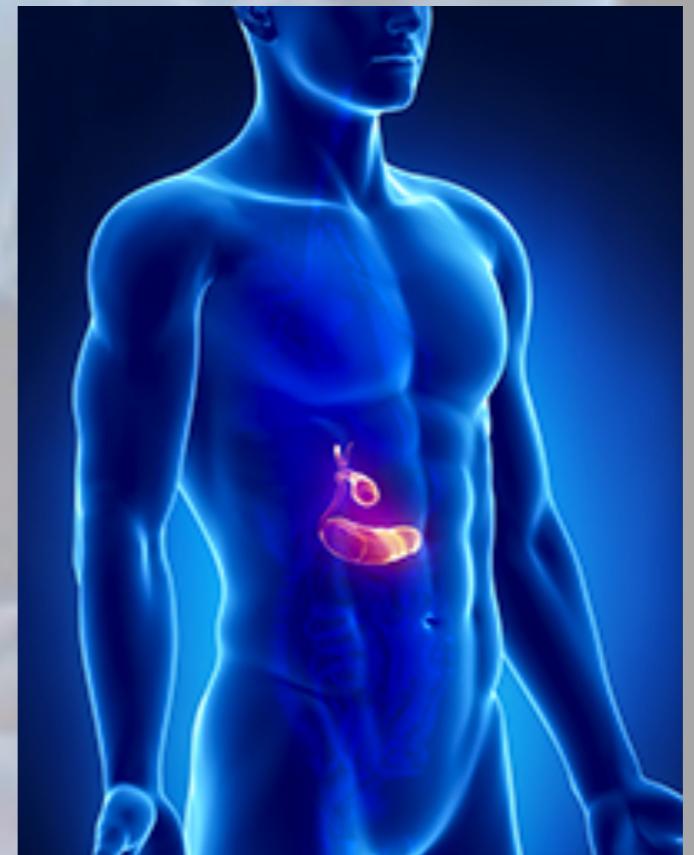
Chronic Kidney Disease (CKD)



Heart disease

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

Chronic Kidney Disease (CKD)



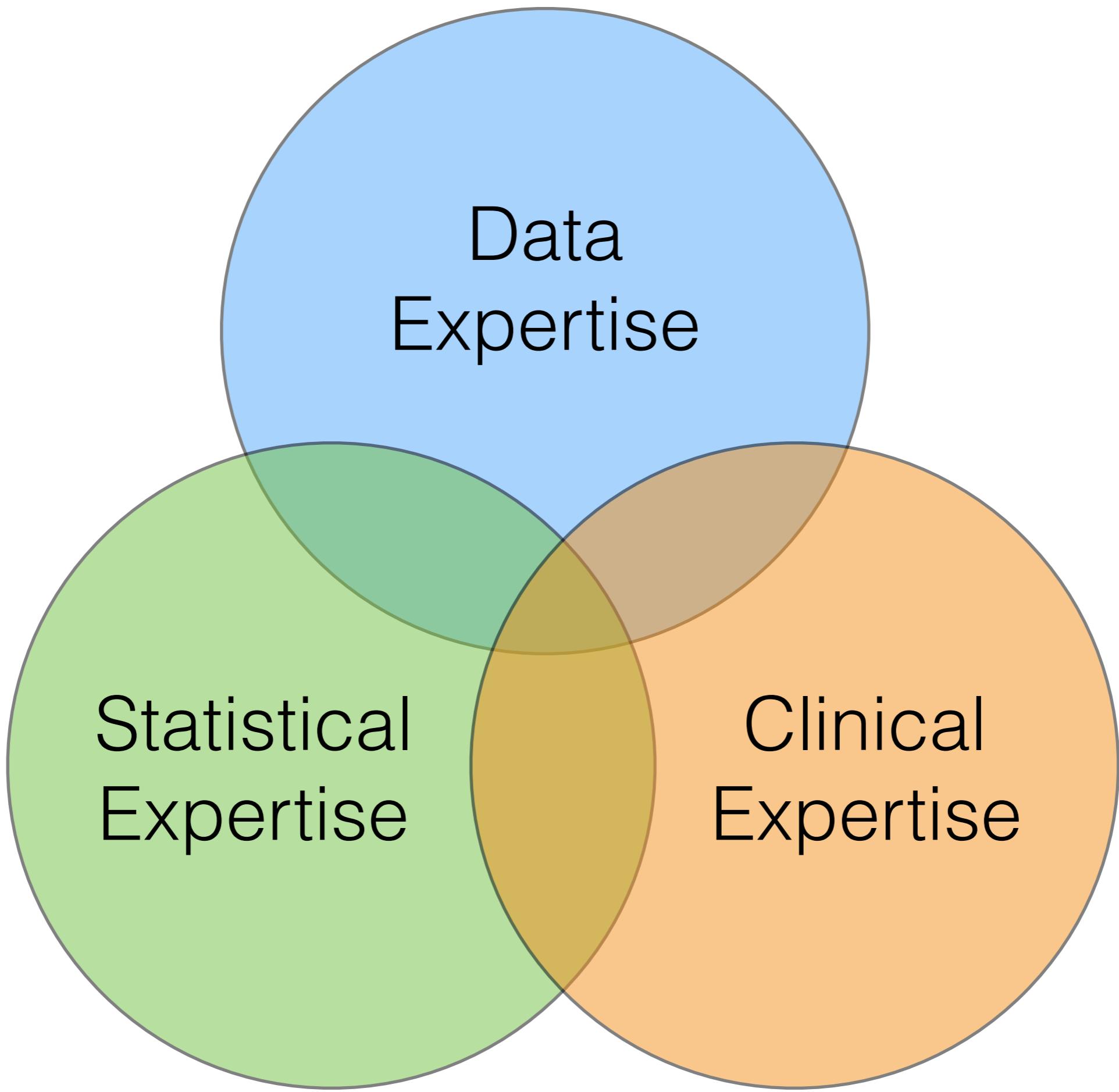
Heart disease

Diabetes

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

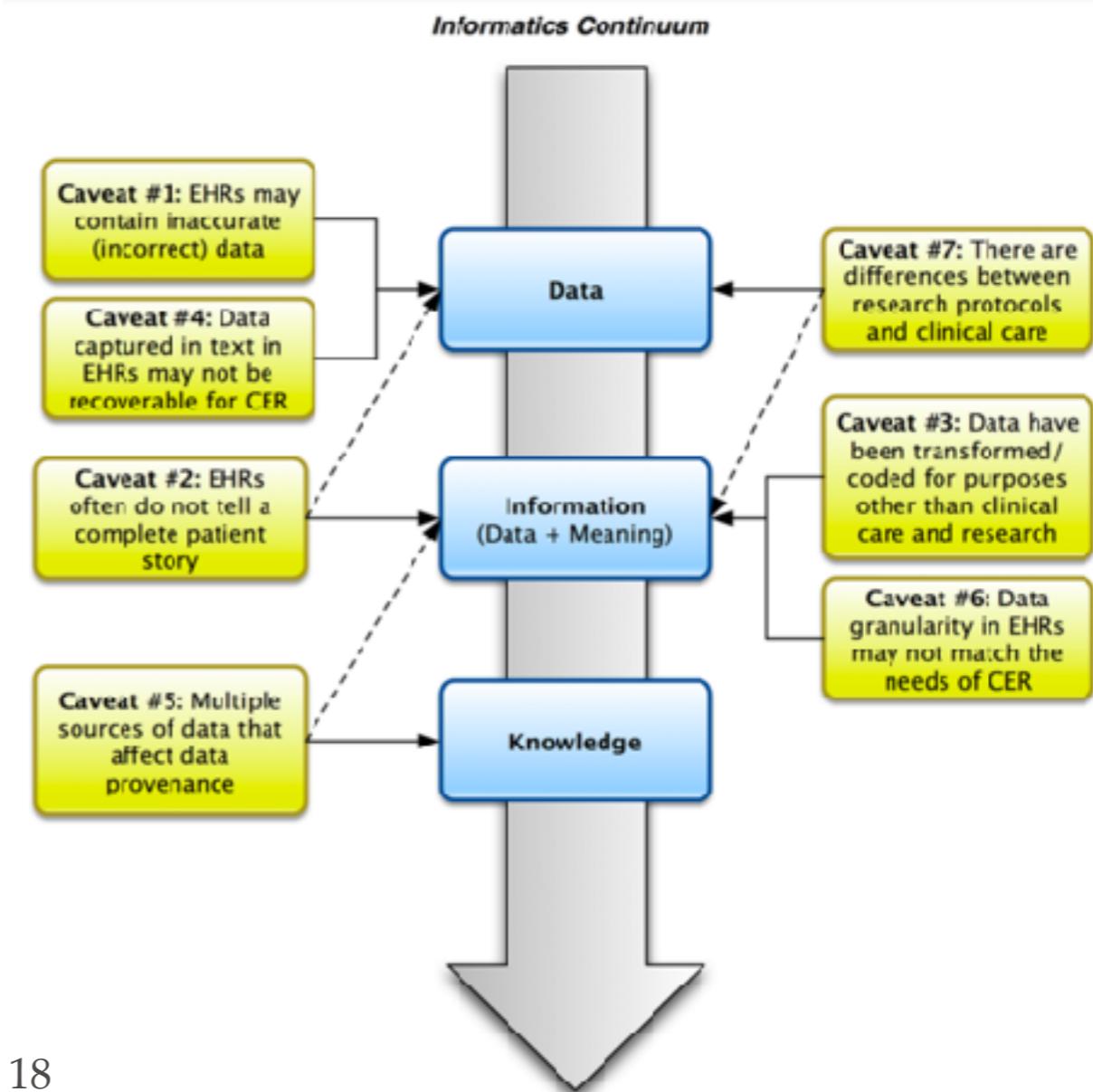
Outline

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



Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research

William R. Hersh, MD,* Mark G. Weiner, MD, † Peter J. Embi, MD, MS,‡ Judith R. Logan, MD, MS,*
Philip R.O. Payne, PhD,‡ Elmer V. Bernstam, MD, MSE,§ Harold P. Lehmann, MD, PhD,||
George Hripcsak, MD, MS,¶ Timothy H. Hartzog, MD, MS,# James J. Cimino, MD,**
and Joel H. Saltz, MD, PhD††





Cohort Manager

Project CKD - 80k

Recent Lines

Line ID	Cohort Name	Line / Line Groups Source	Updated On	Action
1	Duke CKD	70,702	7/10/2012 2:09:11 PM	Basic Detail Delete Print Map It...

Data Selected

Filter by Lines **Copy Selected Columns** **Copy Selected Rows**

Refine cohorts using Boolean logic:
Advanced cohort using either **And** or **Or**

Enter join expression using Line IDs from table above:
Add Join Expression **Join Expression List**

Provide an optional name for this export:

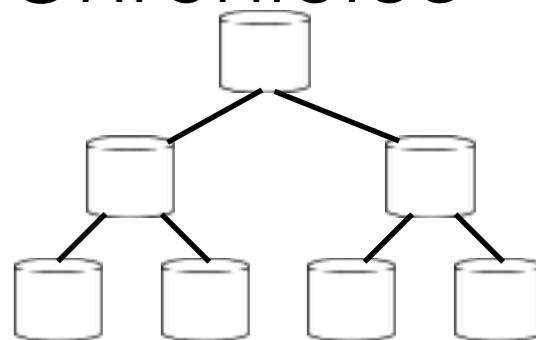
Browse Search

Data Element Name

ICD Procedure Codes	Description
ICD Procedure Codes	The International Classification of Diseases (ICD) code that represents a specific medical procedure
ICD Procedure Short Description	A short textual description of a medical procedure
ICD Procedure Long Description	A long textual description of a medical procedure
ICD Procedure Category Description	Textual description of the low-level grouping of related procedures that the procedure belongs to
ICD Procedure Level 1 Item	The description of the category to which this ICD procedure belongs
AHQ Procedure Level 1 Item	Description for first level of bundling provided by AHRQ



Chronicles



Extract, Transform,
and Load

Clarity

A	B	C	D	E	F
A	B	C	D	E	F
A	B	C	D	E	F
A	B	C	D	E	F
A	B	C	D	E	F
1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6

Export to File **Export to Data Mart**

File Type Export to Delimited Export to SAS

Delimiter Comma Separated Tab Separated Other (Please Specify)

Subject Area Combining

- Separate Subject Areas
- Partially Combine Subject Areas
- Combine Subject Areas

Zip File? (combine multiple files into one & improve download performance)

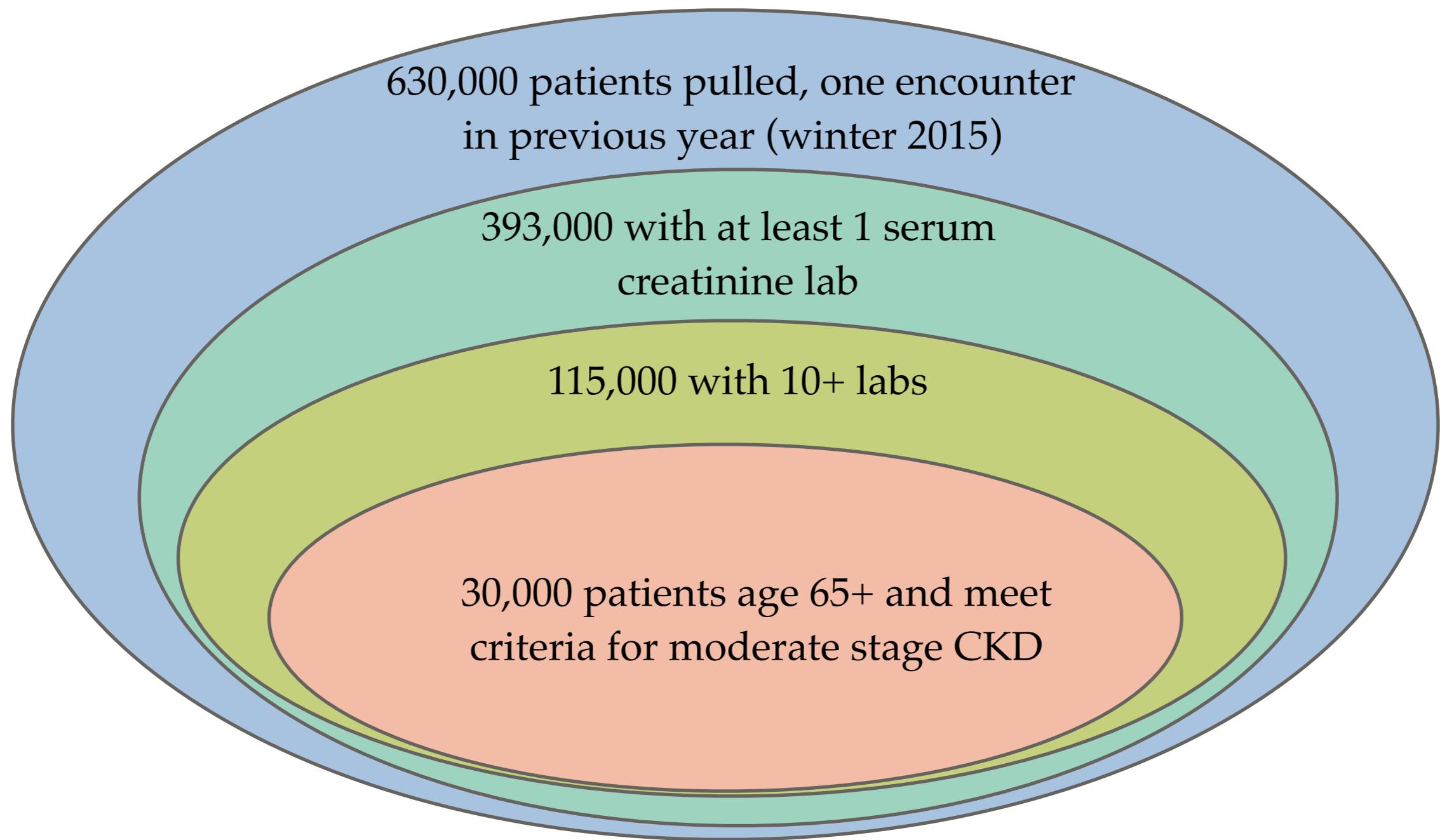
Demographics	Encounters	Diagnoses (13,000 ICD9)
Procedures (10,000 CPT)	Medications	Lab Results
Physician Orders	Culture Results	Radiation/Oncology Notes
Vitals	Social History	Problem Lists
Allergies	Nursing notes	Geography

Decision Support Repository

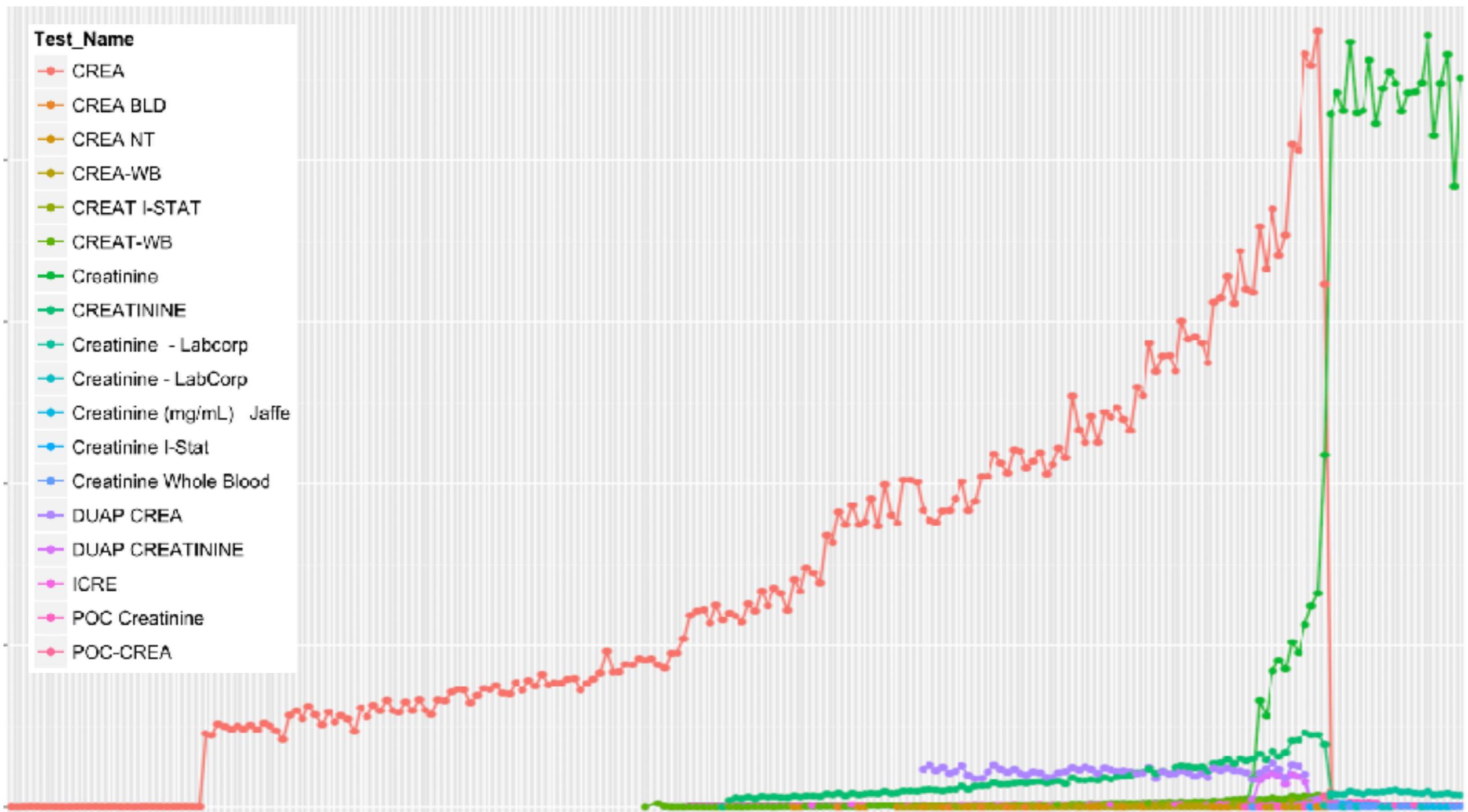


Extract, Transform,
and Load

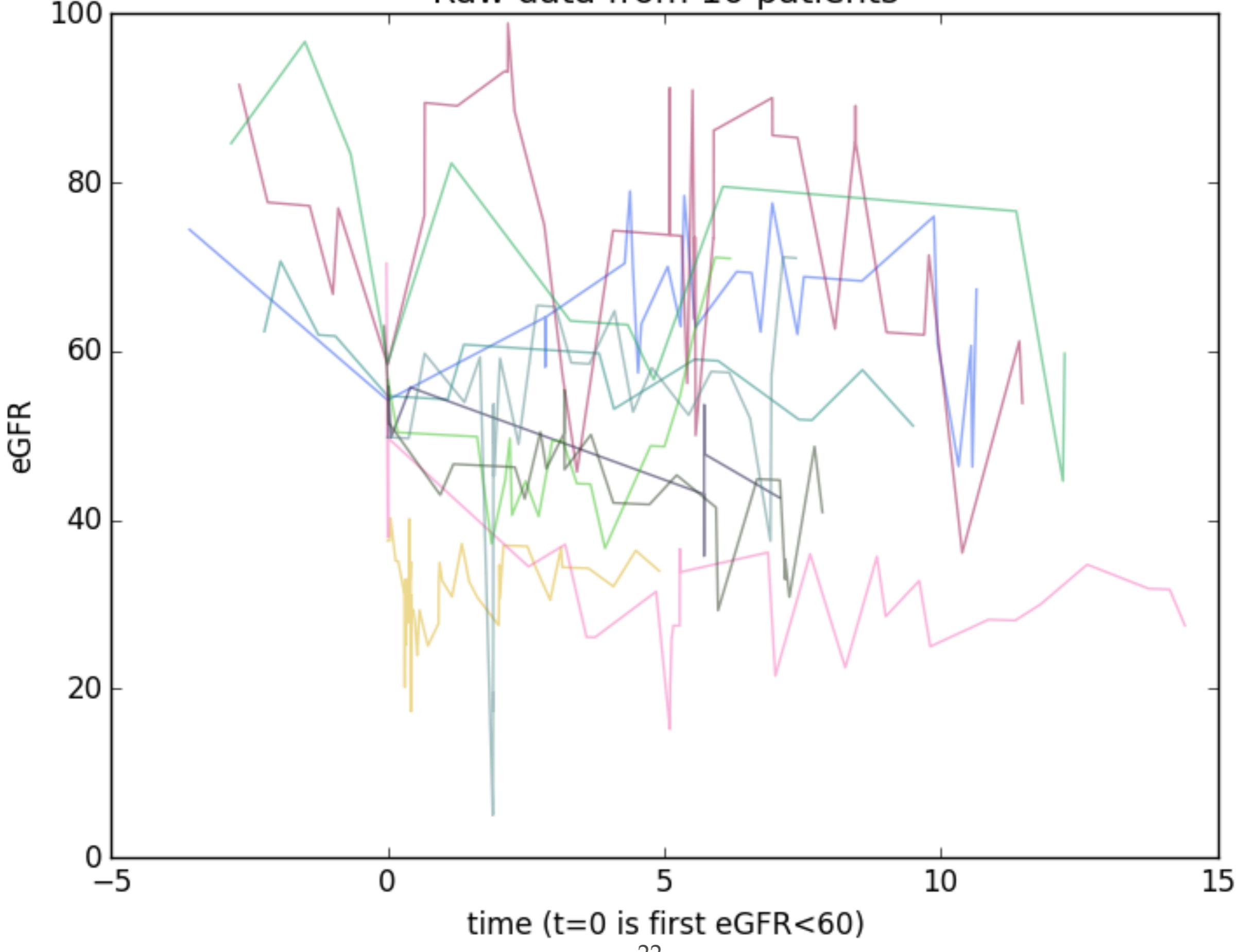
Data acquisition from DEDUCE



Transforming Data



Raw data from 10 patients



Quantifying CKD Progression

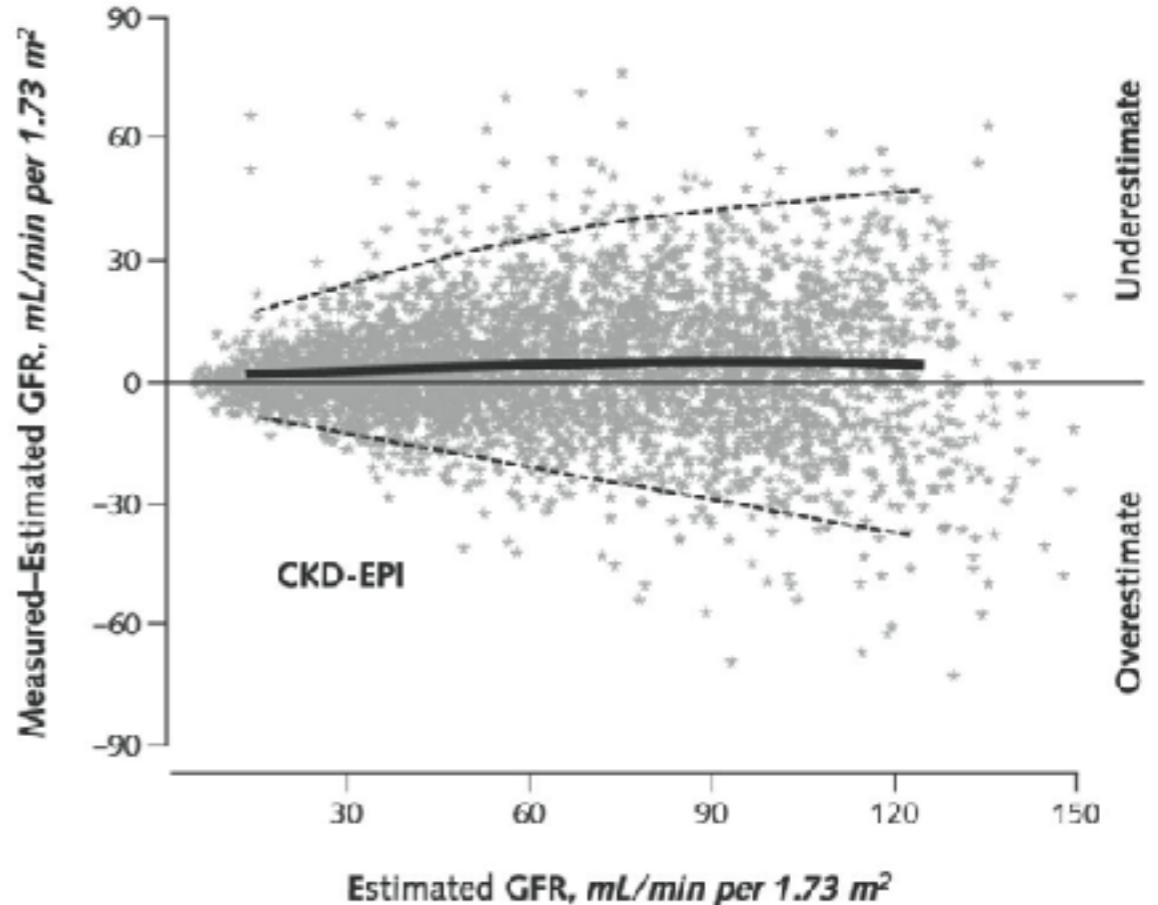
- Estimated glomerular filtration rate (eGFR) is an **extremely noisy** estimate of kidney function. Most common validated equation:

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

S_{cr} : Serum creatinine (mg/dL)

κ : 0.7 (female), 0.9 (male)

α : -0.329 (female), -0.411 (male)



Quantifying CKD Progression

- ❖ Estimated glomerular filtration rate (eGFR) is an **extremely noisy**

American Journal of
Nephrology

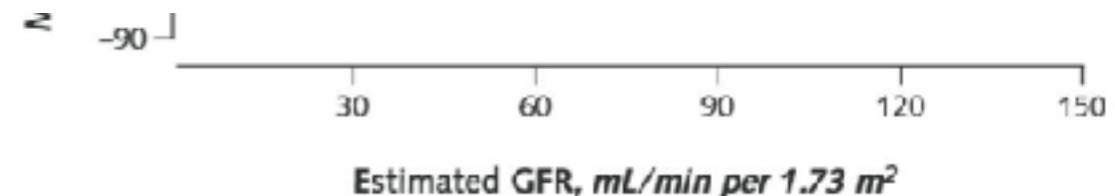
Am J Nephrol 2012;36:1–10
DOI: [10.1159/000339327](https://doi.org/10.1159/000339327)

Received: April 2, 2012
Accepted: May 7, 2012
Published online: June 13, 2012

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
- ❖ \vec{y}_i : eGFRs at times \vec{t}_i ; \vec{u}_i : event times (may be none); x_i covariates
- ❖ Conditional independence in joint likelihood:

$$p(\vec{y}_i, \vec{u}_i | z_i, b_i, f_i, v_i; x_i) = p(\vec{y}_i | z_i, b_i, f_i; x_i)p(\vec{u}_i | z_i, b_i, f_i, v_i; x_i)$$

Longitudinal Submodel

Longitudinal Submodel

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

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

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$

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

Longitudinal Submodel

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

$$\begin{aligned} & \text{iid random noise} \\ & y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2) \\ & m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t). \end{aligned}$$

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$
$$m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t).$$

Longitudinal Submodel

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$
$$m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t).$$

Population component (fixed intercept and slope)

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

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

basis expansion,
in practice [1,t] coefficient matrix baseline covariates

Population component (fixed intercept and slope)

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$
$$m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t).$$

Longitudinal Submodel

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$
$$m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t).$$

Subpopulation component (unique disease trajectory with splines)

Longitudinal Submodel

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$
$$m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t).$$

fixed B-spline
expansion of time coefficient subpopulation
vector assignment (1-G),
 $p(z_i = g) \propto \exp\{w_g^\top x_{iz}\}$

Subpopulation component (unique disease trajectory with splines)

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$
$$m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t).$$

Longitudinal Submodel

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$
$$m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t).$$

Individual component (random intercept and slope)

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$

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

basis expansion,
in practice [1,t]

random effect,
 $b_i \sim N(0, \Sigma_b)$

Individual component (random intercept and slope)

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$
$$m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t).$$

Longitudinal Submodel

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$
$$m_i(t) = \Phi_p(t)^\top \Lambda x_{ip} + \Phi_z(t)^\top \beta_{z_i} + \Phi_l(t)^\top b_i + f_i(t).$$

Structured noise process (GP noise, transient trends)

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$

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

zero mean GP,

OU kernel: $K_{OU}(t_1, t_2) = \sigma_f^2 \exp\left\{-\frac{|t_1 - t_2|}{l}\right\}$

Structured noise process (GP noise, transient trends)

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

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$$

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

Point Process Submodel

Point Process Submodel

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

Point Process Submodel

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

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

Point Process Submodel

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

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

- ❖ 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 $\{u_{ik}\}_{k=1}^{K_i}$:

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

- ❖ 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 $\{u_{ik}\}_{k=1}^{K_i}$:

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

- ❖ 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 baseline
 covariates

Point Process Submodel

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

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

- ❖ 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

baseline
covariates

association between
event risk and
expected mean/slope
of eGFR

Point Process Submodel

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

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

- ❖ 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

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. $\lambda_i = \{\nu_{z_i}, \mu_{b_i}, \Sigma_{b_i}, \mu_{v_i}, \sigma_{v_i}^2, \mu_{f_i}, \Sigma_{f_i}\}$, pt. est. $\hat{\Theta}$
- ❖ ELBO has closed form, exact gradients with automatic differentiation
- ❖ Stochastic optimization, **subsample** observations for noisy unbiased gradients wrt Θ

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 $< 60\text{mL/min}$, 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 $< 60\text{mL/min}$
- ❖ 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+ CVA code (mean w/ 1+: 6.4, std dev: 13.3, median: 3.0)
- ❖ Baseline covariates: baseline age, race, gender; hypertension, diabetes

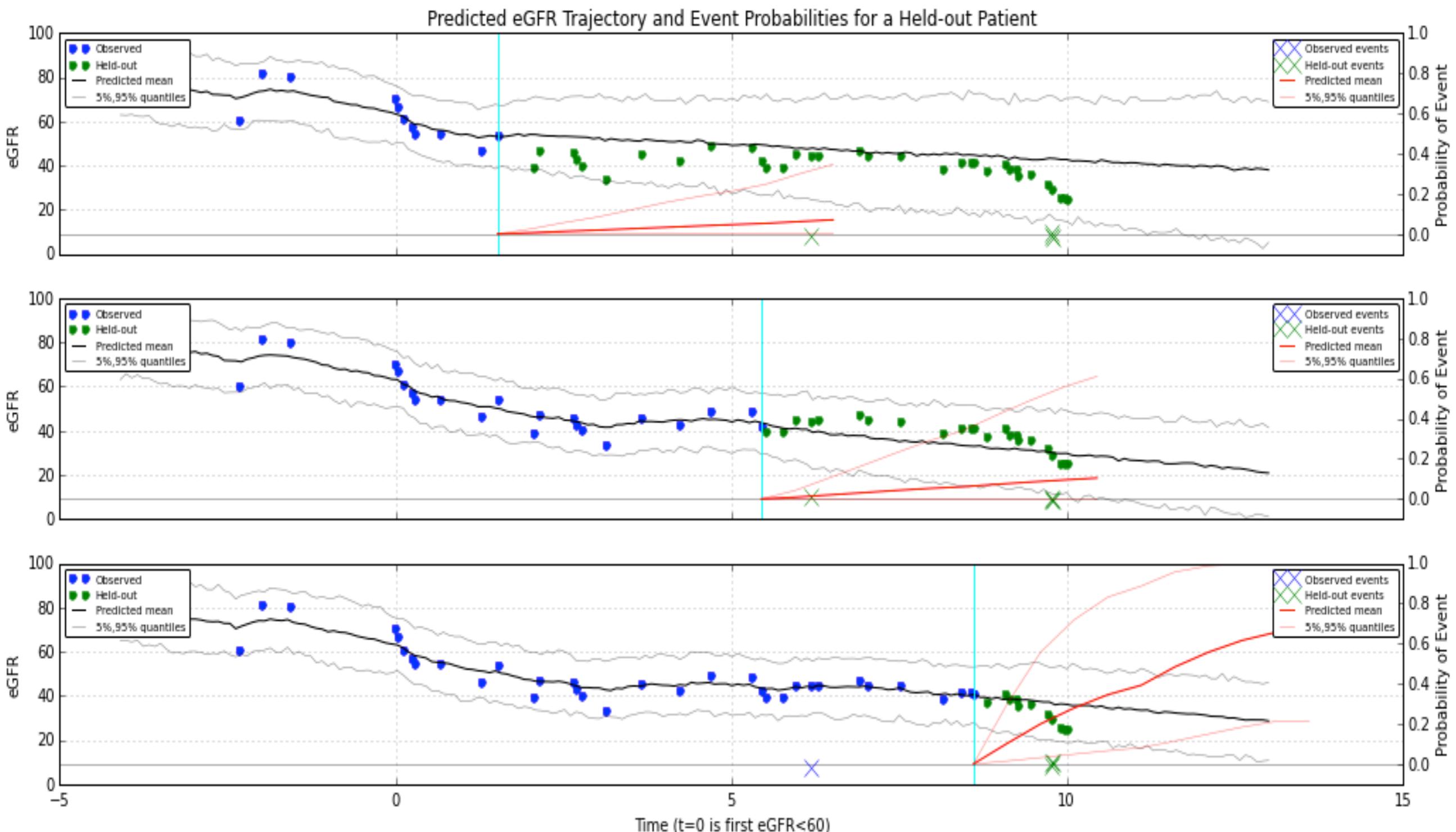
Experimental Setup

- ❖ 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

<u>Longitudinal Submodels</u>	MSE		MAE	
[Schulam & Saria, 2015]		155.80		9.27
Joint Model (CVA)		147.31		9.01
Joint Model (AMI)		152.78		9.15
<u>CVA: AUROCs</u>				
Joint Model	0.786	0.746	0.727	0.742
Cox	0.574	0.597	0.602	0.606
Time-varying Cox	0.576	0.557	0.563	0.593
<u>AMI: AUROCs</u>				
Joint Model	0.755	0.704	0.737	0.654
Cox	0.704	0.676	0.617	0.599
Time-varying Cox	0.640	0.652	0.647	0.663
<u>CVA: AUPRs</u>				
Joint Model	0.423	0.389	0.370	0.405
Cox	0.065	0.101	0.123	0.137
Time-varying Cox	0.062	0.086	0.114	0.157
<u>AMI: AUPRs</u>				
Joint Model	0.163	0.128	0.172	0.166
Cox	0.052	0.059	0.051	0.065
Time-varying Cox	0.048	0.057	0.067	0.088

Joint Model Results

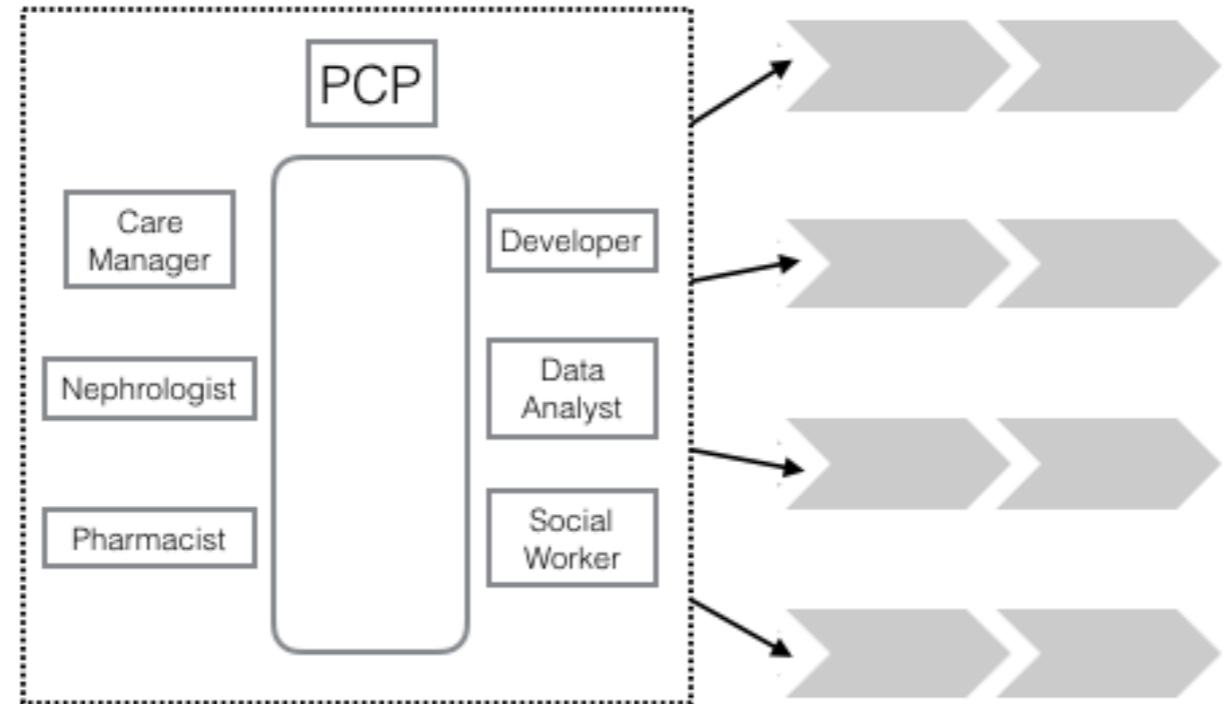


Outline

- ❖ 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



Patients

Filters

Name	MRN	$\Delta eGFR$ range	- ∞	-5	Tangri range	15	∞
Flags/Diagnoses	Without Flags/Diagnoses						

Reset Apply

Tangri	$\Delta eGFR$	MRN	Name	Age	Flags/Diagnoses
96.2	-16.8	90HWL3M2	Johnpaul Shields	47	CKD Diabetes HTN Rapid Decline Nephrology
61.4	-9.5	FSC8M6H0	Beaulah McKenzie	57	CKD Deceased Diabetes HTN Rapid Decline Nephrology
29.6	-6.9	SH4ER2FW	Aditya Wehner	46	CKD Diabetes HTN Rapid Decline Rounded Dialysis
26.9	-5.7	KTCVWGU	Jovani Lindgren	47	CKD HTN Rapid Decline Rounded Prorounded
19.2	-12.3	SXXY5L9O	Brook Reichert	47	CKD Diabetes HTN Rapid Decline Nephrology

Tracy Morgan, 44 9MSPI33Q

Tangri: 0.3 ΔeGFR: — CKD Stage: 2 Durham, NC, Durham County

Flags and Diagnoses
Diabetes HTN Prerounded

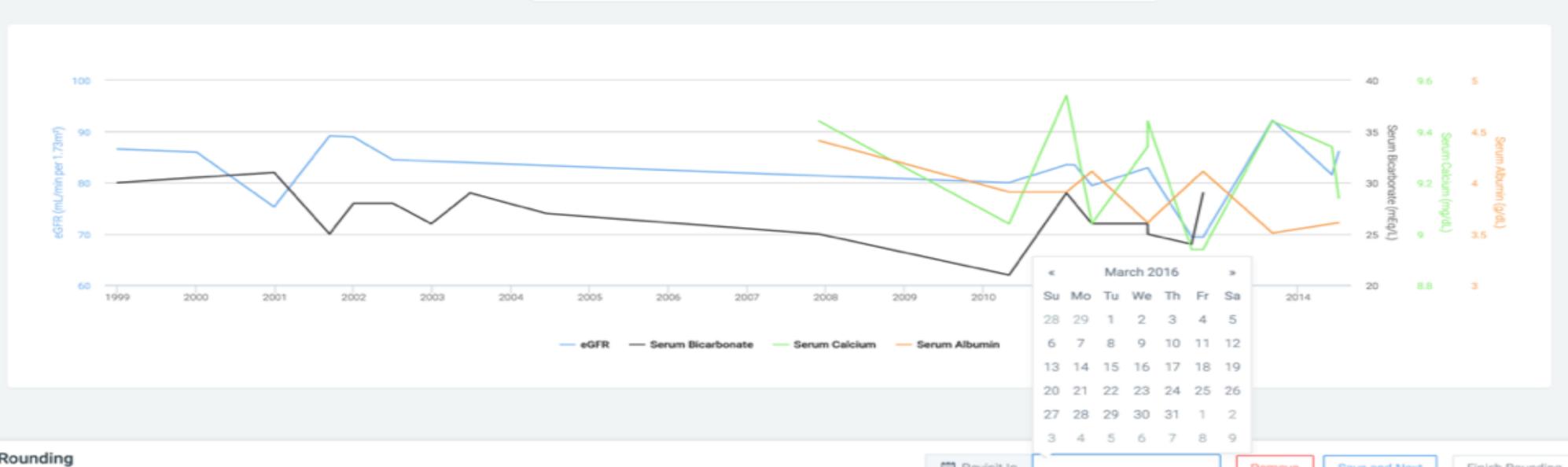
Comments

+ New

Next Actions

+ New

- Make primary care appointment
Seen in Pickens
Assigned to General 9 days ago by Angela Anteround

[Edit](#)

Rounding

12 Rounded, 45 Remaining

[Revisit In](#)[Remove](#)[Save and Next](#)[Finish Rounding](#)

Tasks Assigned to Your Groups

Aditya Wehner, 46 Tangri: 29.6 ΔeGFR: -6.92 5H4ER2FW

Kerry Packer, 49 Tangri: 1.1 ΔeGFR: -2.59 LUYIN558

- schedule PCP appointment
with Dr. Johnson in Hillsborough
Assigned to General 16 days ago by Angela Anteround

[Edit](#)

Janie Abbott, 52 Tangri: 0.5 ΔeGFR: — NLHUWFUD

- Make nephrology appointment
Assigned to General 10 days ago by Angela Anteround

[Edit](#)

Tracy Morgan, 44 Tangri: 0.3 ΔeGFR: — 9MSPI33Q

- Make primary care appointment
Seen in Pickens
Assigned to General 9 days ago by Angela Anteround

[Edit](#)

Conclusion

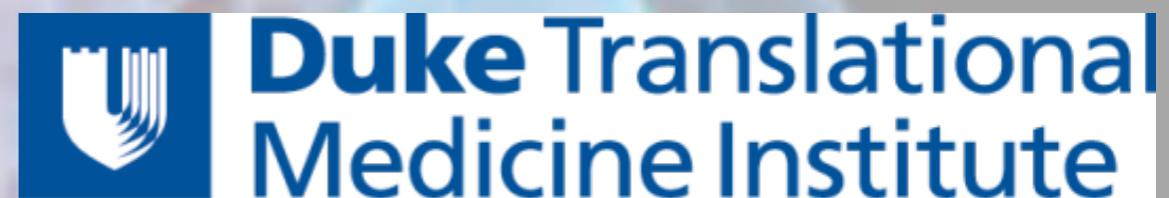
- ❖ Novel joint model for longitudinal, point process data
 - ❖ First stochastic variational inference algorithm for joint models
- ❖ **Actionable** implementation in use for CKD patients!
- ❖ Future work:
 - ❖ Multivariate in both longitudinal variables and events
 - ❖ More flexible models (e.g. beyond 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.



References

- C. Lloyd, T. Gunter, M. A. Osborne, and S. J. Roberts. Variational inference for Gaussian process modulated Poisson processes. *ICML*, 2015.
- A. Perotte, R. Ranganath, J. S. Hirsch, D. Blei, and N. Elhadad. Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis. *Journal of the American Medical Informatics Association*, 2015.
- C. Proust-Lima, M. Sene, J. Taylor, H. Jacqmin-Gadda. Joint latent class models for longitudinal and time-to-event data: A review. *Statistical Methods in Medical Research* 23(1):74-90, 2014.
- D. Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Chapman and Hall / CRC Biostatistics Series, 2012.
- P. Schulam and S. Saria. A Framework for Individualizing Predictions of Disease Trajectories by Exploiting Multi-Resolution Structure. *NIPS*, 2015.
- N. Tangri, L. A. Stevens, et al. A Predictive Model for Progression of Chronic Kidney Disease to Kidney Failure. *JAMA* 305(15):1553-1559, 2011.

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 Θ .

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 Θ .

 Update Θ 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

$$q_i(f_i(\vec{t}_i), f_i(\vec{u}_i), f_i(t_i^{\text{grid}})) = p(f_i(\vec{u}_i), f_i(t_i^{\text{grid}}) | f_i(\vec{t}_i)) q(f_i(\vec{t}_i) | \mu_{f_i}, \Sigma_{f_i})$$

Data: data y, u ; hyperparameters.

Result: point estimate $\hat{\Theta}$, approximate posteriors q_i .

Initialize global parameters Θ .

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 Θ .

 Update Θ using AdaGrad.

until convergence of the ELBO;

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