Interacting Dynamic Processes for Social Network Data

Katherine Heller
Center for Cognitive Neuroscience
Department of Statistical Science

Department of Computer Science
Department of Neurobiology
Department of Neurology

My Research Areas

- Machine Learning for Healthcare
- Social Network Modeling
- Jointly Modeling Diverse Data Sets
  - Neuroscience
  - Chronic Kidney Disease
- Bayesian time series models

- Cognitive Science Applications with Bayesian Nonparametrics
- Bayesian Clustering
- Topic/language modeling

Social Interactions

- Fundamental to understanding human behavior.
- Very complex.
- Ideal for interdisciplinary research.

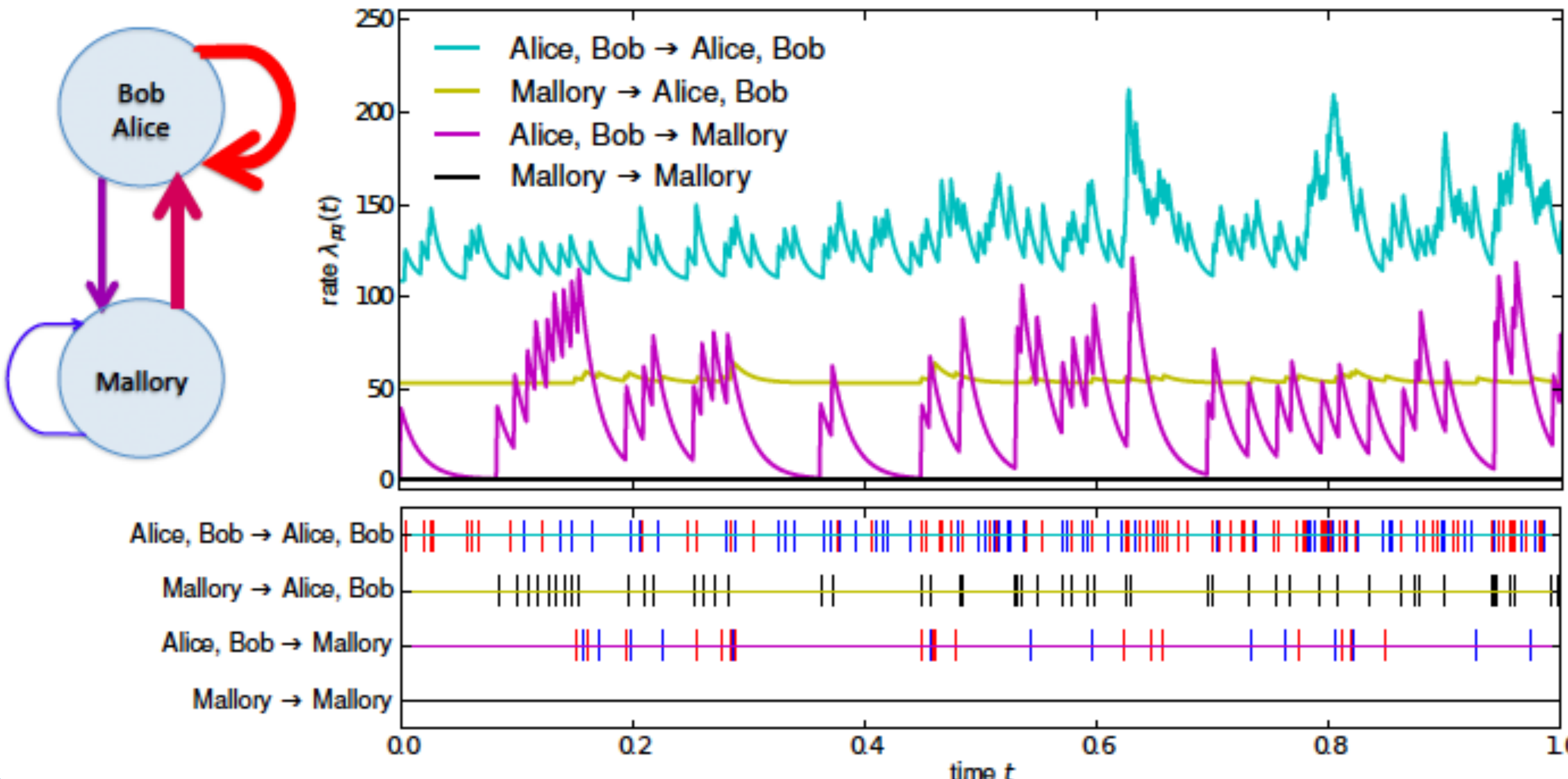
Outline

- Hawkes Processes and the Bayesian Echo Chamber
- Modeling the Spread of Disease with Graph-coupled HMMs
- Other Machine Learning for Healthcare Work

Reciprocating Relationships

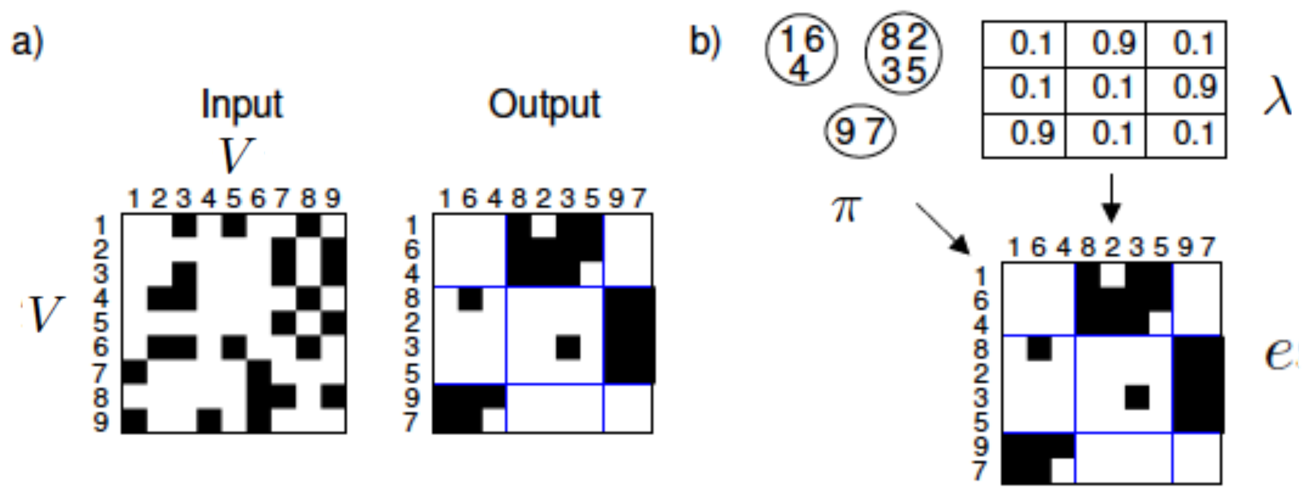
- People organize into social groups.
- We'd like to know the dynamics of these groups.
- But these groups are not explicitly defined.
  - Can use declared relationships (e.g. fb friends) to infer groups.
  - However not always available, truthful, or useful.
- We infer groups from real interactions.
  - Data is a sequence of many events - actions from a sender to a receiver.
  - Leverage reciprocity patterns to discover social groups.

Reciprocating Relationships



Infinite Relational Model

- The IRM is a method for clustering entities based on graphs of declared relationships.



Kemp et al

$$\pi \sim \mathrm{CRP}(\alpha)$$

$$\lambda_{pq} \sim \mathrm{Beta}(\gamma, \gamma) \qquad \forall p, q \in \mathrm{range}(\pi)$$

$$e_{uv} \sim \mathrm{Bernoulli}(\lambda_{\pi(u)\pi(v)}) \qquad \forall u, v \in V$$

Poisson Process IRM

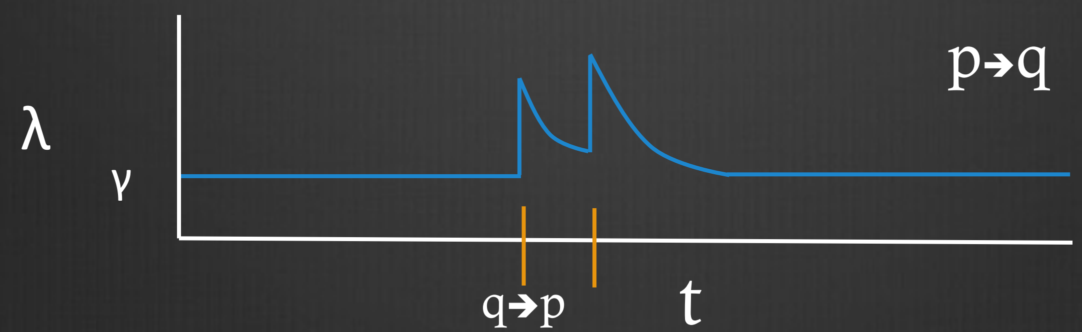
- Often interaction data contains many interactions between the same pair of individuals.
  - Cannot be modeled with a vanilla IRM.
- Can modify to use Gamma-Poisson observation model.
  - But then we cannot predict events into the future.
- Therefore we consider a Poisson process.

$$\pi \sim \mathrm{CRP}(\alpha)$$
$$\lambda_{pq} \sim \mathrm{Gamma}(\delta, \beta)$$
$$N_{uv}(\cdot) \sim \mathrm{PoissonProcess}(\lambda_{\pi(u)\pi(v)})$$

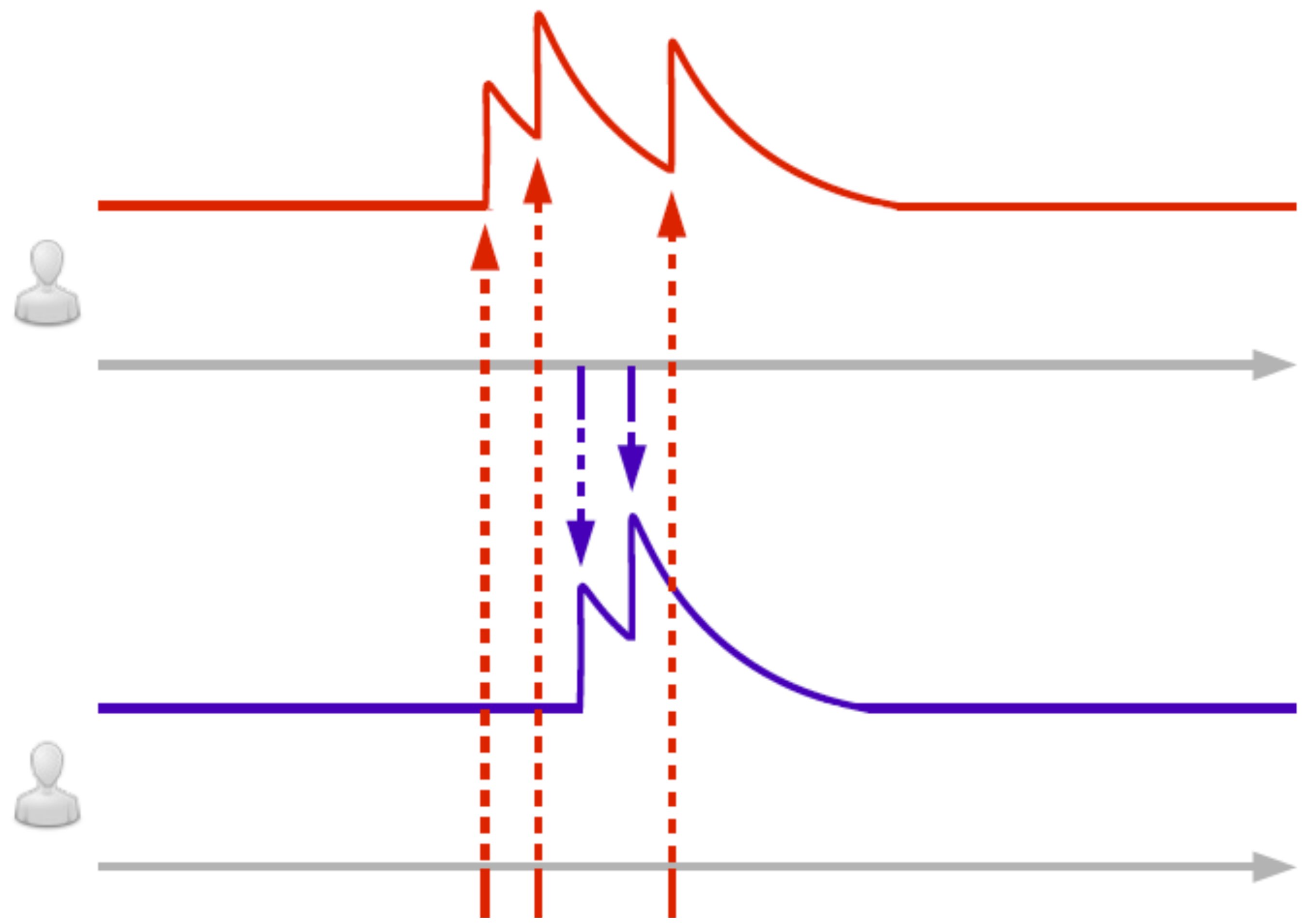
Hawkes Processes

- Drawbacks of using the Poisson Process:
  - Rate of events between pairs of clusters independent of all other pairs.
  - Times of events are uniformly distributed.
- Instead use a mutually-exciting Hawkes process:

$$\lambda_{pq}(t) = \gamma_{pq} + \sum_{i:t_i^{qp}<t} g_{pq}(t - t_i^{qp}), \qquad g_{pq}(\delta) = \beta_{pq} e^{-\frac{\delta}{\tau_{pq}}}$$



Multivariate Hawkes Process



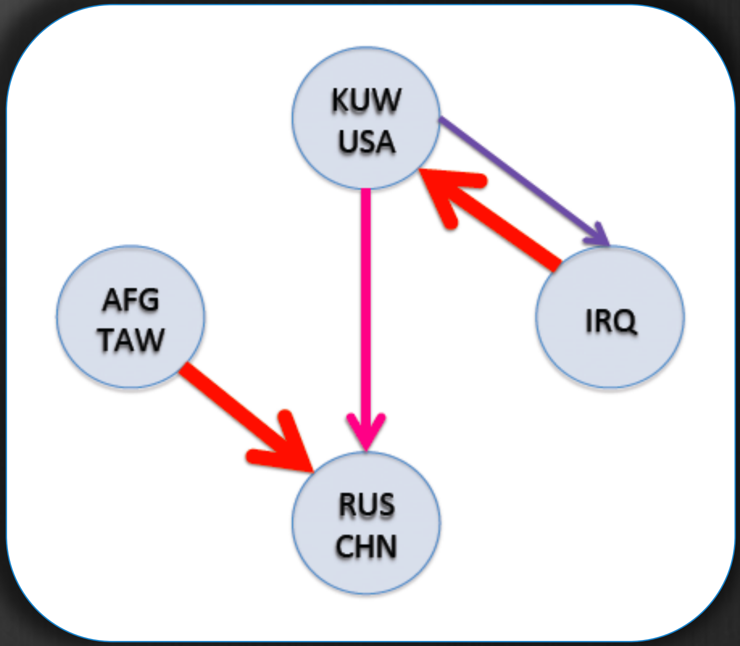
Inference

- Posterior inference is performed using MCMC.
- Unlike previous IRM models there is no conjugate prior for the likelihood.
  - Must sample parameters instead of integrating out.
- We infer the partition of entities, CRP concentration parameter, and Hawkes process parameters using Metropolis within Gibbs and Slice sampling.
  - Basically Neal's algorithm 5 with some additional slice sampling.

Experiments: Correlates of War

- Militarized Interstate Disputes data set which captures correlates of war.
  - Data: Years 1993-2001, all MID incidents and countries involved.
  - Incidents vary from diplomatic threats to deployment of force.

- 3 main conflicts
  - Russia and Afghanistan
  - Taiwan and China
  - USA, Iraq and Kuwait



Blundell et al, NIPS 2012

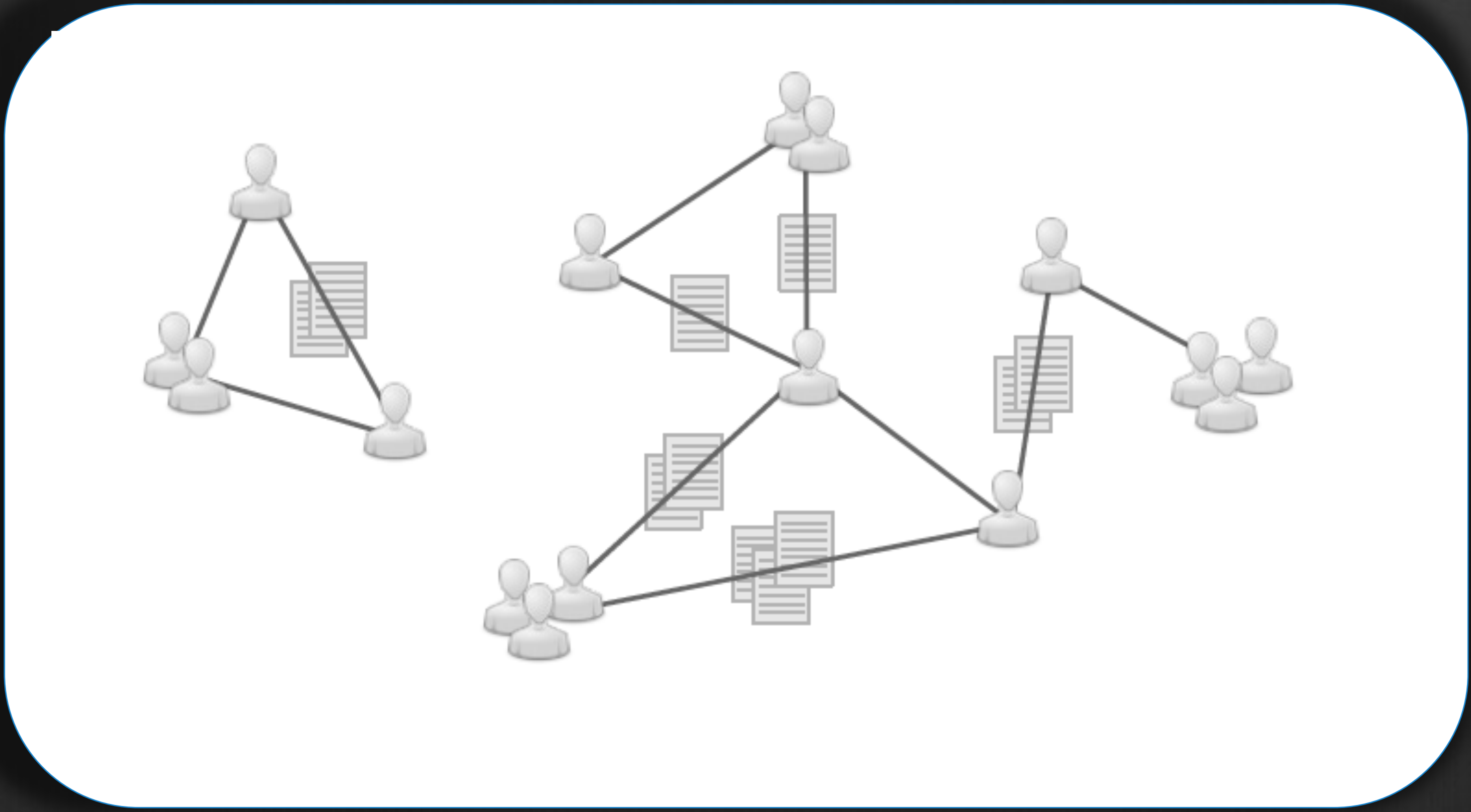
Experiments:

- Log predictive probabilities of events falling in the last 10% of time.
- Also Enron data – 5 longest threads.

	Hawkes IRM	Poisson IRM	Hawkes	Poisson
Synthetic	**23.49±0.03**	20.45±0.04	17.73±0.01	17.76±0.00
Small MID	**10.21±0.26**	-1.81±0.11	0.18±0.08	-2.30±0.07
Full MID	**-127.45±0.45**	-132.71±0.09	-188.29±0.02	-188.44±0.01
Enron 0	**220.16±0.04**	162.44±0.06	194.34±0.02	133.02±0.07
Enron 1	**335.88±0.04**	304.89±0.11	284.71±0.05	256.62±0.11
Enron 2	**108.21±0.02**	104.79±0.05	87.49±0.02	85.57±0.05
Enron 3	**101.32±0.04**	98.90±0.16	82.31±0.04	81.62±0.05
Enron 4	**121.07±0.07**	113.25±0.08	100.16±0.04	91.70±0.08
SB conv 23	**127.91±0.28**	47.39±0.10	26.00±0.01	19.63±0.02
SB conv 26	**-6.19±0.04**	-12.38±0.03	-7.15±0.03	-10.47±0.03
SB conv 12	**53.23±0.38**	45.67±0.12	17.44±0.06	-8.69±0.12
SB conv 49	**199.27±0.30**	128.05±0.13	125.30±0.02	122.44±0.03
SB conv 33	**31.86±0.27**	2.39±0.08	21.75±0.01	21.63±0.01

Bayesian Echo Chamber

- Look at how people influence each other.



Temporal Influence

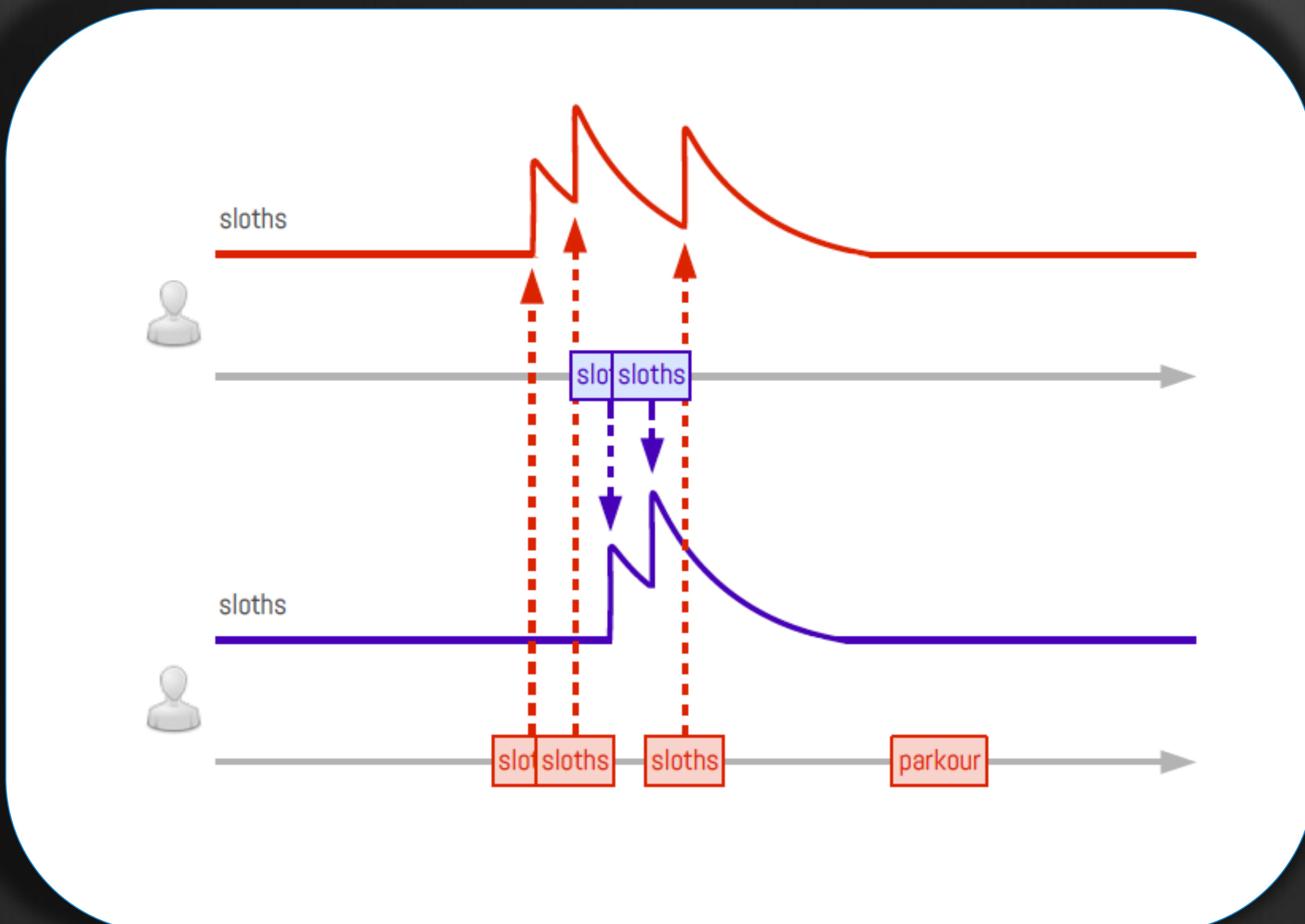
pairwise
influence

$$N^p(a,b] \sim \mathrm{Pois}\left(\int_a^b \lambda_0^p + \sum_{q \neq p} \nu^{qp} \sum_{t_n^q < t} \exp\left(\frac{-(t - t_n^q)}{\tau^p}\right) \mathrm{d}t\right)$$

Linguistic Accommodation

> "During communication encounters, people will try to accommodate or adjust their style of speaking to others. [...] Convergence occurs when there is a strong need for social approval, frequently from powerless individuals.
>
> — West & Turner, 2010

Mutually Exciting Language Models



For each person…

language profile

pairwise influence

$$\boldsymbol{m}_t^p \propto \boldsymbol{\mu}^p + \sum_{q \neq p} \rho^{qp} \sum_{t_n^q < t} \boldsymbol{N}^{t_n^q} \cdot \exp\left(\frac{-(t - t_n^q)}{\tau^p}\right)$$

sum over all other people

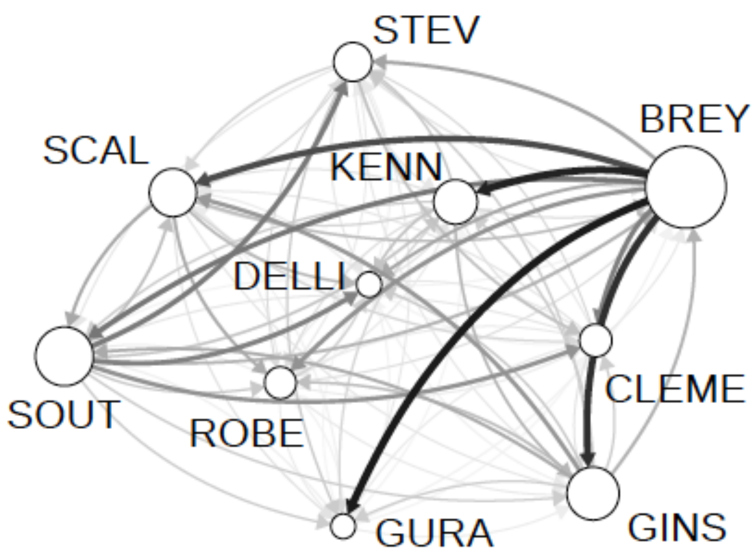
word counts

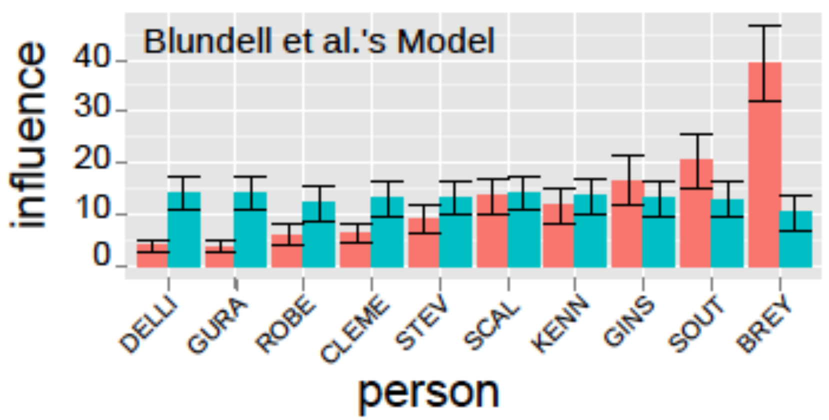
Linguistic Influence

pairwise influence

$$\boldsymbol{m}_t^p \propto \boldsymbol{\mu}^p + \sum_{q \neq p} \rho^{qp} \sum_{t_n^q < t} \boldsymbol{N}^{t_n^q} \cdot \exp\left(\frac{-(t - t_n^q)}{\tau^p}\right)$$

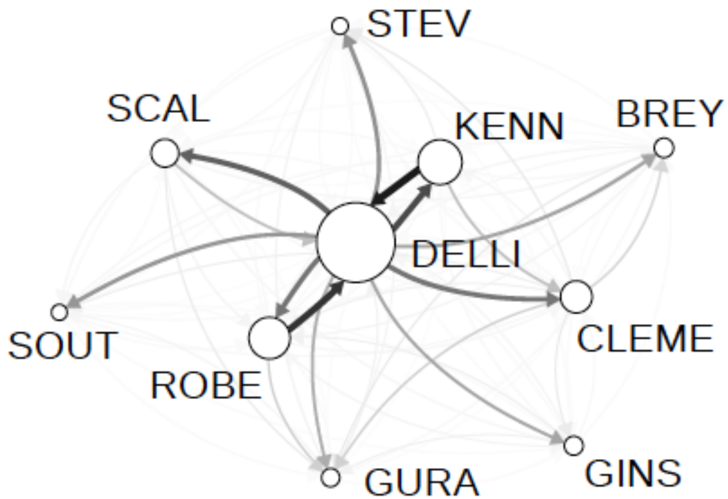
DC v Heller: Temporal Influence

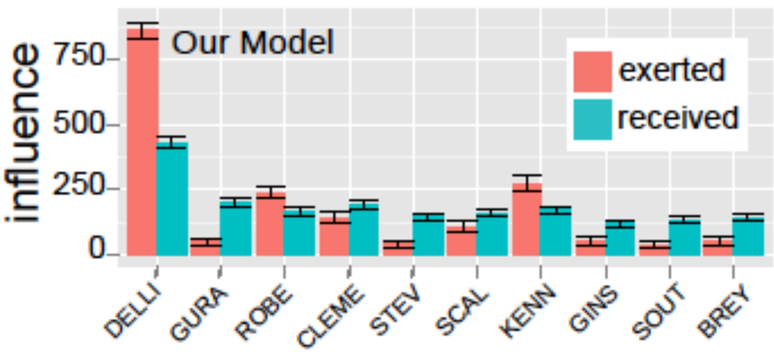




Guo et al, AISTATS 2015

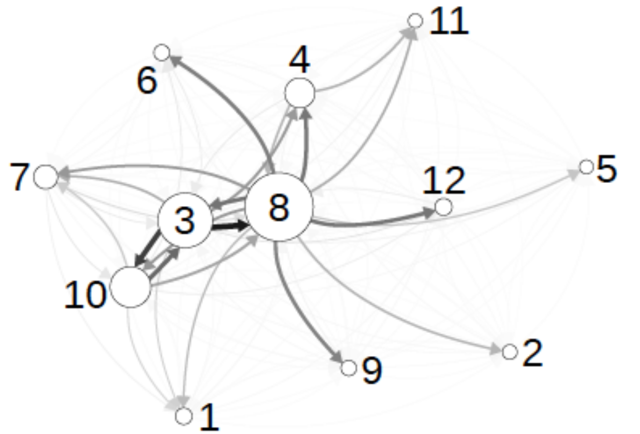
DC v Heller: Linguistic Influence

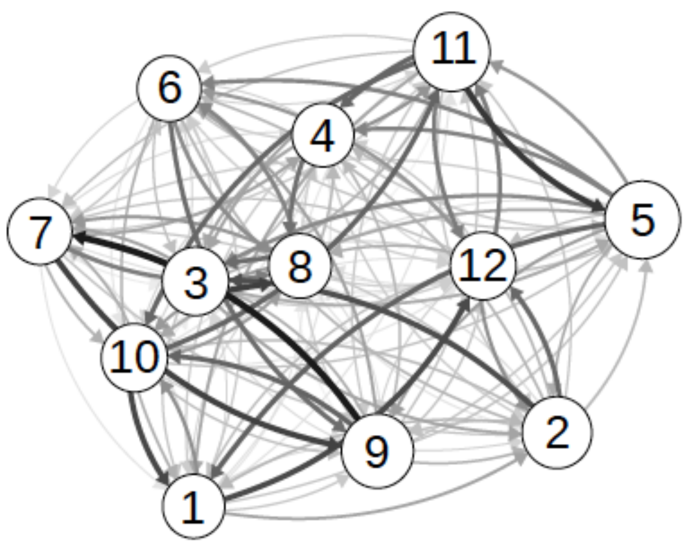


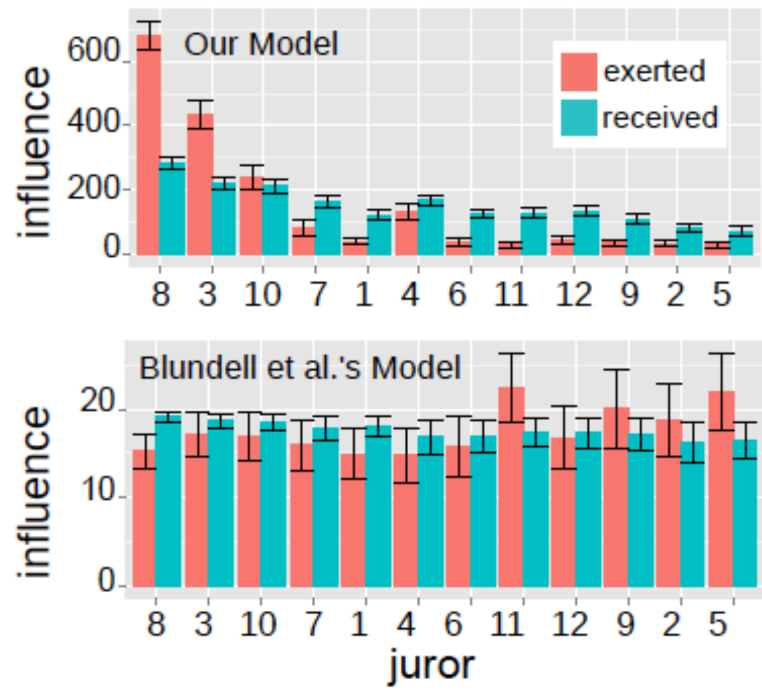


Guo et al, AISTATS 2015

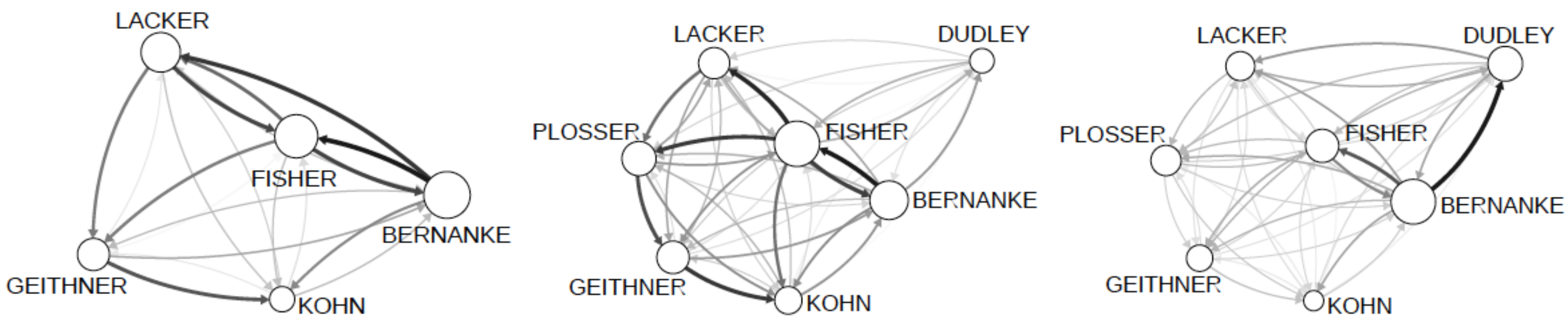
12 Angry Men


11
4
6
7
5
12
3
8
10
2
9
1


11
6
4
7
5
3
8
12
10
2
9
1


Our Model
exerted
received
influence
600
400
200
0
8 3 10 7 1 4 6 11 12 9 2 5
Blundell et al.'s Model
influence
20
10
0
8 3 10 7 1 4 6 11 12 9 2 5
juror

Federal Reserve Board Meetings



Guo et al, AISTATS 2015

Relating Time and Content

- Allow the amount of excitation to be dependent on content in an event based Hawkes process with IRM.

$$\beta_{uv}(s) = e^{r_u(x_{vu}(s)) + s_v(x_{vu}(s))}$$

where

$$r_u(\cdot) \sim \mathcal{GP}(0, k_r)$$
$$s_v(\cdot) \sim \mathcal{GP}(0, k_s)$$

- Results in improved log likelihood on held out test data (enron, santa barbara conversation corpus, citation), and more intuitive clusters.

Tan et al, UAI 2016

Extensions

- Learning clusterings of neurons. Similar processes are used in neuroscience to model the activation and co-activation of neurons. These usually also involve inhibition, but don't directly cluster.
- Scale up inference
- Deal with unknown recipients
- Online bullying

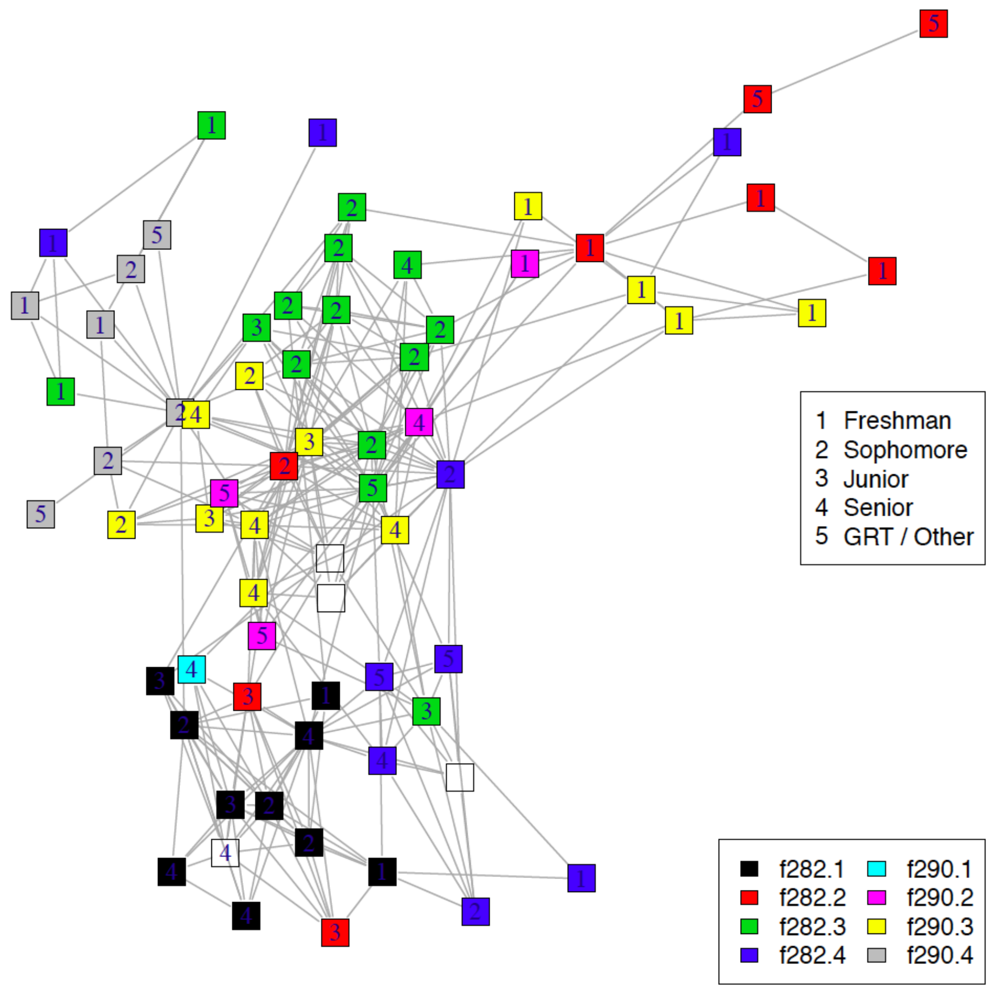
Infection in a Social Network

- Goal: To model dynamical interactions between agents in a social network and apply to inferring the spread of infection.
- Many traditional epidemics models work on a population level, treating each person the same way.
- Contemporary data collection techniques allow us to model the spread of infection on an individual level.
- Being able to make infection predictions on an individual level is enormously beneficial because it allows people to receive more personalized and relevant health advice.

Social Evolution Experiment

- Data collected in the social evolution experiment allows us for the first time to closely track proximities and contagion in an entire community over a substantial period of time.
- Tracked “common cold” symptoms in an MIT residence hall from January to April 2009.
- Monitored over 80% of residents through their cell phones from October 2008 to May 2009, taking daily surveys and tracking their location, proximities and phone calls.
- Monthly surveys on social, health, and political issues taken. Locations taken by having cell phones scan nearby wifi access points and bluetooth devices.

Student Hall Network

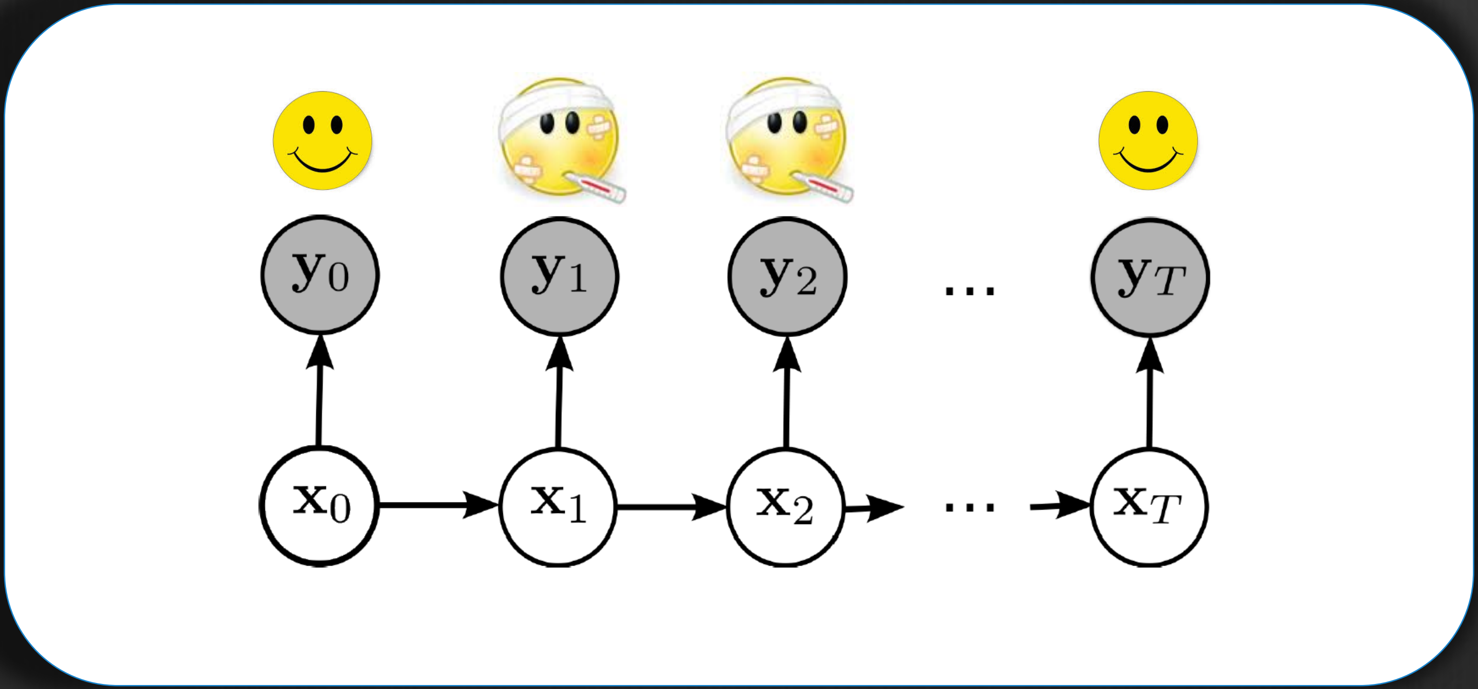


Health Surveys

- In the Social Evolution experiment students were paid $1 a day to answer surveys about contracting infection.
- The surveys asked about symptoms:
  - Runny nose, nasal congestion, sneezing
  - Nausea, vomiting, diarrhea
  - Stress
  - Sadness and depression
  - Fever
- 64 of the 85 residents answered the surveys
- Symptoms dependent on the social network. A student with a symptom had a 3-10x higher odds of having a friend with the same symptom.

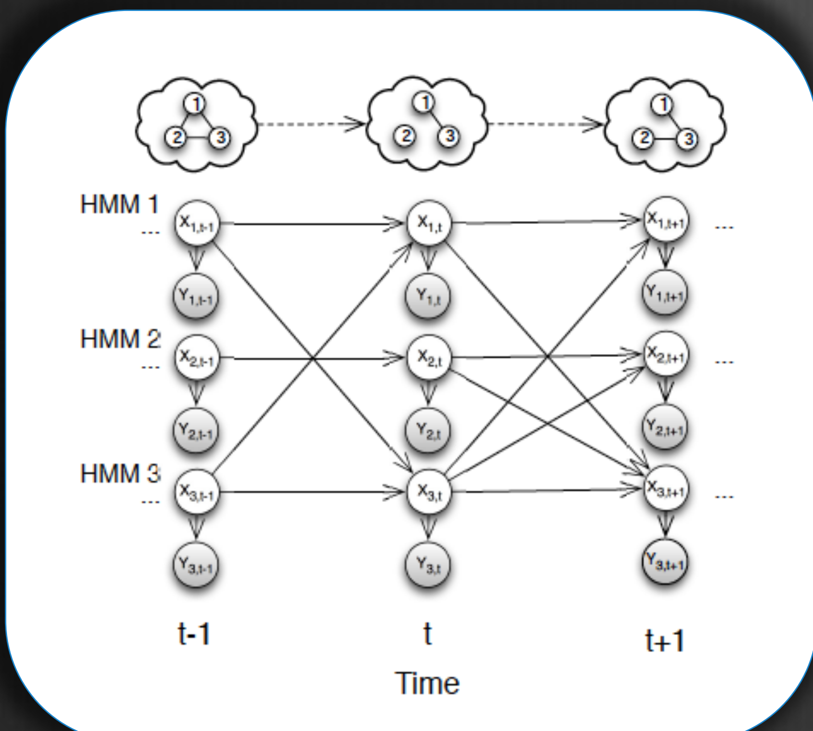
Hidden Markov Models

- We aim to leverage the social evolution data to predict the spread of infection to individuals in our social network using a model related to HMMs.



Graph-coupled HMMs

- Associate each person in the dynamic interaction network with an HMM chain. Let interaction network structure determine the HMM couplings:



$$X_{n,t} \sim \mathrm{Categorical}(\phi_n, X_{e:\{n,\cdot\}\in G_t, t-1})$$

$$Y_{n,t} \sim F(\theta_{X_{n,t}})$$

$$\theta_{X_n} \sim \mathrm{Conj}(\gamma)$$

$$\phi_{n,X_{e:\{n,\cdot\}\in G_t}} \sim H(X_{e:\{n,\cdot\}\in G_t}, \mu)$$

GCHMM Inference

- Inference in the coupled HMM is very hard.
  - Typically ML estimation is done on few chains with few states or another approximation is made.
- In the worst case GCHMM inference is as difficult as CHMM inference.
  - When the graph is fully connected.
- Fortunately, since we're dealing with social networks we can leverage a couple of properties:
  - Social networks are usually sparse
  - For many applications the influence of interactions can be modeled in a fairly simple way via a small number of parameters.

GCHMMs for Modeling Infection

- In the case of the social evolution data the influence of other HMMs can be summarized by counts of interactions in the infectious state.

- The GCHMM can provide an individual level version of the susceptible-infectious-susceptible (SIS) epidemiology model:

$$\dot{S} = -\beta \cdot SI + \gamma \cdot I \qquad \dot{I} = \beta \cdot SI - \gamma \cdot I$$

- GCHMM for infection:

$$X_{n,t} \sim \text{Bernoulli}(\phi_{n, X_{e:\{n,\cdot\} \in G_t, t-1}})$$

$$Y_{n,t,i} \sim \text{Bernoulli}(\theta_{X_{n,t}})$$

$$\theta_{X_n} \sim \text{Beta}(h)$$

$$p(X_{n,t+1} = 0 | X_{n,t} = 1) = \gamma$$

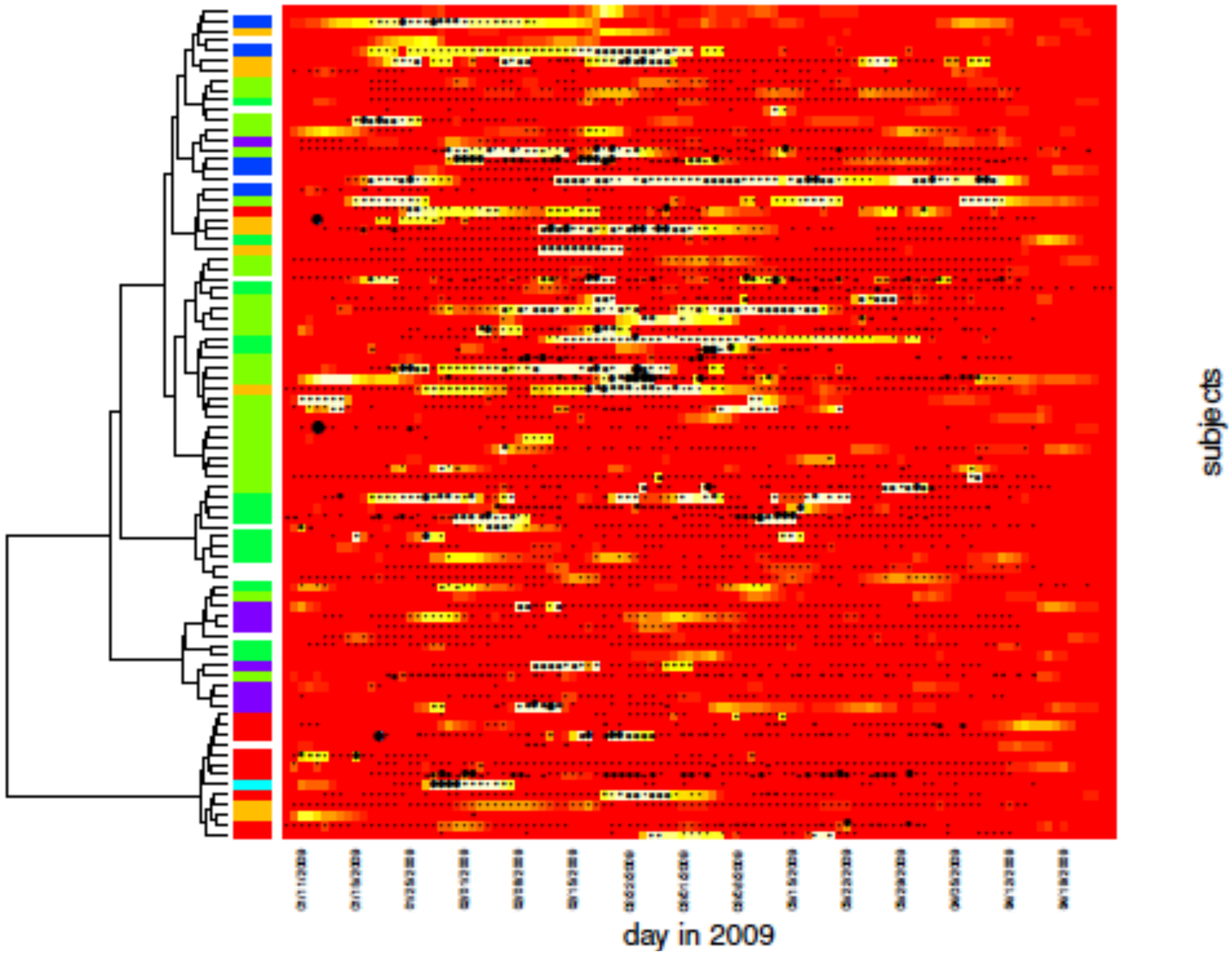
$$p(X_{n,t+1} = 0 | X_{n,t} = 0, X_{e:\{n,\cdot\} \in G_t}) = (1-\alpha)(1-\beta)^{\sum_{e:\{n,\cdot\} \in G_t} X_{ne,t}}$$

$$\alpha \sim \text{Beta}(a, b)$$

$$\beta \sim \text{Beta}(a', b')$$

$$\gamma \sim \text{Beta}(a'', b'')$$

Experimental Results



Dong et al, UAI 2012

Aiello Group Data

- eX-FLU study at University of Michigan
  - 590 students from 6 dorms
  - Chain referral scheme
- A 103 student subset participated in iEpi
  - Smartphone based study where location is tracked and surveys taken
- Unlike MIT study confirmation of interaction was recorded on phones and flu testing was done on students who reported being ill.
- Also an isolation intervention was tested.

Hierarchical GCHMMs

- Add a hierarchical level to where beta distributed infection parameters are learned:

Beta-exponential link

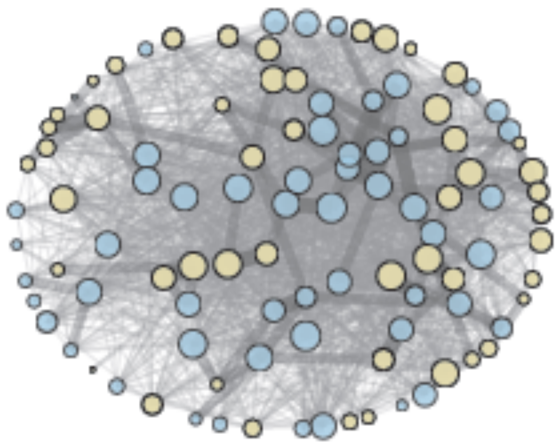
$$
\begin{aligned}
\eta_{\cdot,\cdot} &\sim N(\mu, \Sigma) \\
\gamma_n &\sim \text{Beta}(\exp(z_n^\top \eta_{r,1}), \exp(z_n^\top \eta_{r,2})) \\
\alpha_n &\sim \text{Beta}(\exp(z_n^\top \eta_{a,1}), \exp(z_n^\top \eta_{a,2})) \\
\beta_n &\sim \text{Beta}(\exp(z_n^\top \eta_{b,1}), \exp(z_n^\top \eta_{b,2}))
\end{aligned}
$$

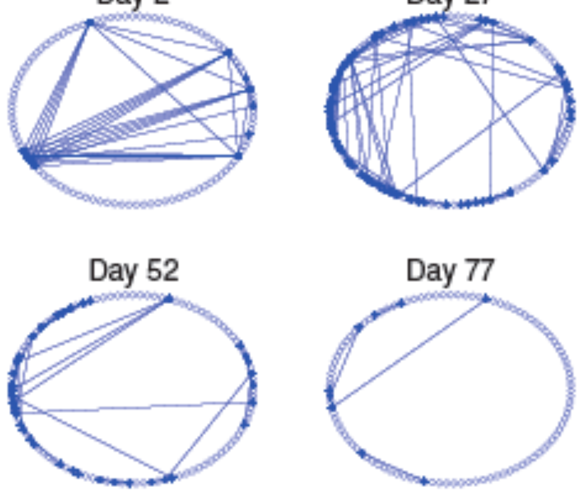
Sigmoid link

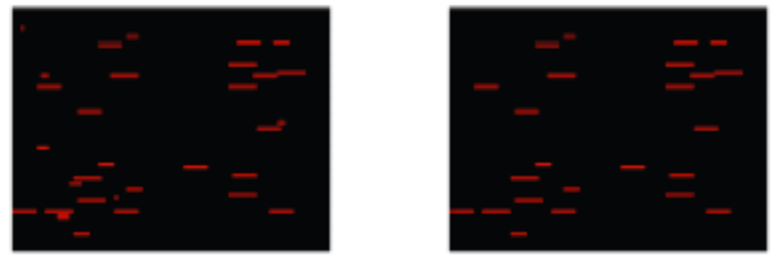
$$
\begin{aligned}
&\eta_{\cdot} \sim N(\mu, \Sigma) \\
&\gamma_n = \sigma(z_n^\top \eta_r), \quad \alpha_n = \sigma(z_n^\top \eta_a), \quad \beta_n = \sigma(z_n^\top \eta_b)
\end{aligned}
$$

- Inference Gibbs-EM algorithm but follow up papers e.g. stochastic VB

Results







Model	Recall	Accuracy
Sigmoid link	0.8974 ± 0.00	0.9978 ± 0.00
Beta-exp link	0.7436 ± 0.00	0.9912 ± 0.00
GCHMMs+LogReg	0.7436 ± 0.00	0.9912 ± 0.00

Table 2: Coefficients Estimation on exFlu Dataset

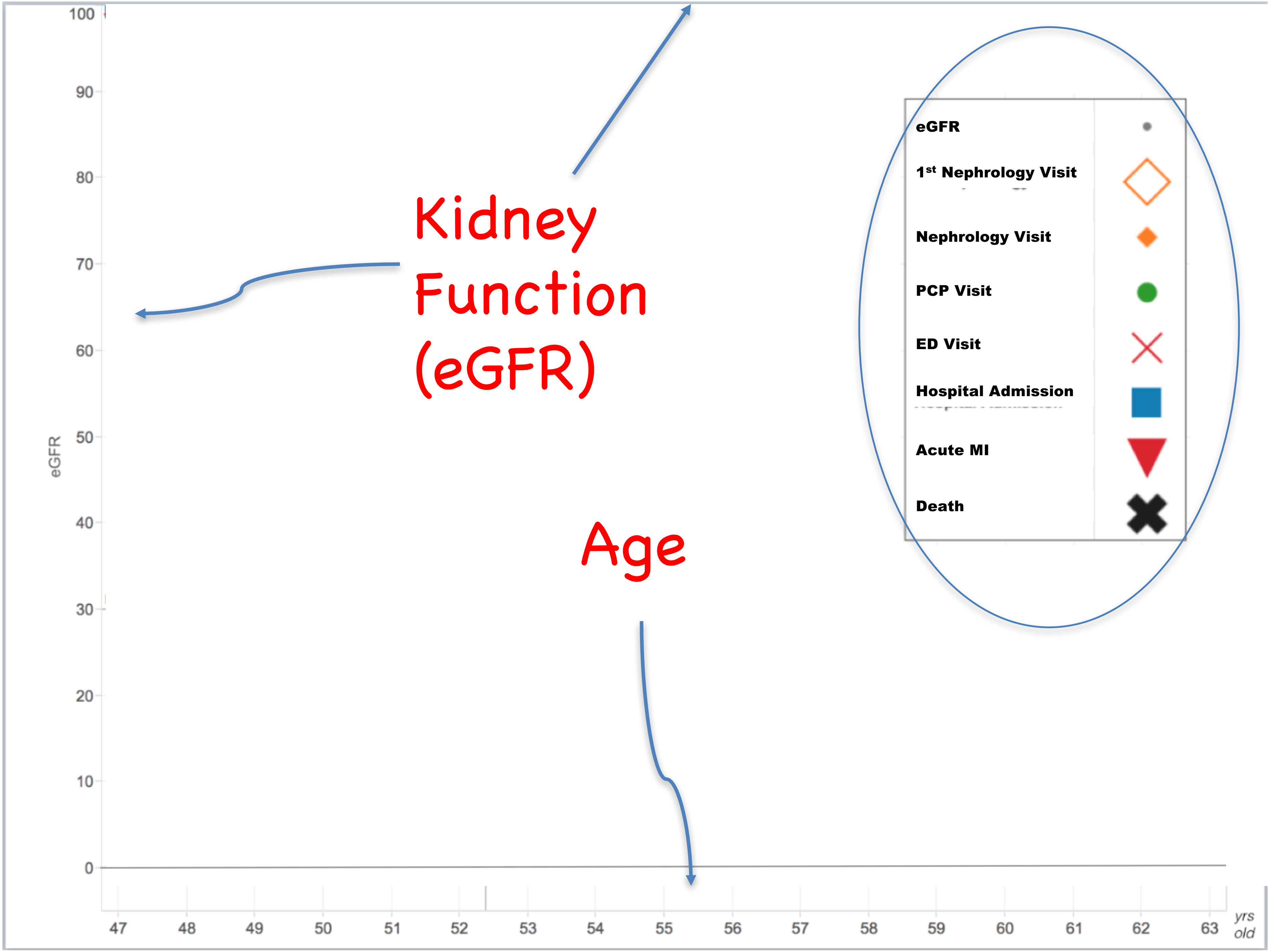
Feature[1]	Recovery η_r	Outside Infect η_a	Inside Infect η_b
Default=1	-1.3022 ± 0.0146	-5.1517 ± 0.0024	-4.1619 ± 0.0281
Gender	-0.1575 ± 0.0118	-0.2428 ± 0.0074	-0.1457 ± 0.0078
Age	0.0074 ± 0.0082	-0.2376 ± 0.0051	-0.0181 ± 0.0017
Alc_Day	0.1090 ± 0.0078	-0.1534 ± 0.0003	-0.0410 ± 0.0018
Vacc_Ever	-0.0698 ± 0.0104	0.1092 ± 0.0095	0.0382 ± 0.0085
Flushot_Yr	0.0769 ± 0.0092	-0.3209 ± 0.0073	0.0837 ± 0.0055
Smoker	-0.1080 ± 0.0029	-0.0536 ± 0.0008	0.0773 ± 0.0021
Drinker	-0.1335 ± 0.0092	0.0628 ± 0.0030	0.1408 ± 0.0029
Act_Days	0.0356 ± 0.0099	0.0054 ± 0.0063	-0.0622 ± 0.0078
Sleep_Qual	0.0225 ± 0.0069	-0.3686 ± 0.0051	-0.0162 ± 0.0077
Wash_Opt	0.0024 ± 0.0103	0.0816 ± 0.0132	-0.0714 ± 0.0048
High_Risk	-0.1274 ± 0.0116	-0.1252 ± 0.0058	-0.0727 ± 0.0007

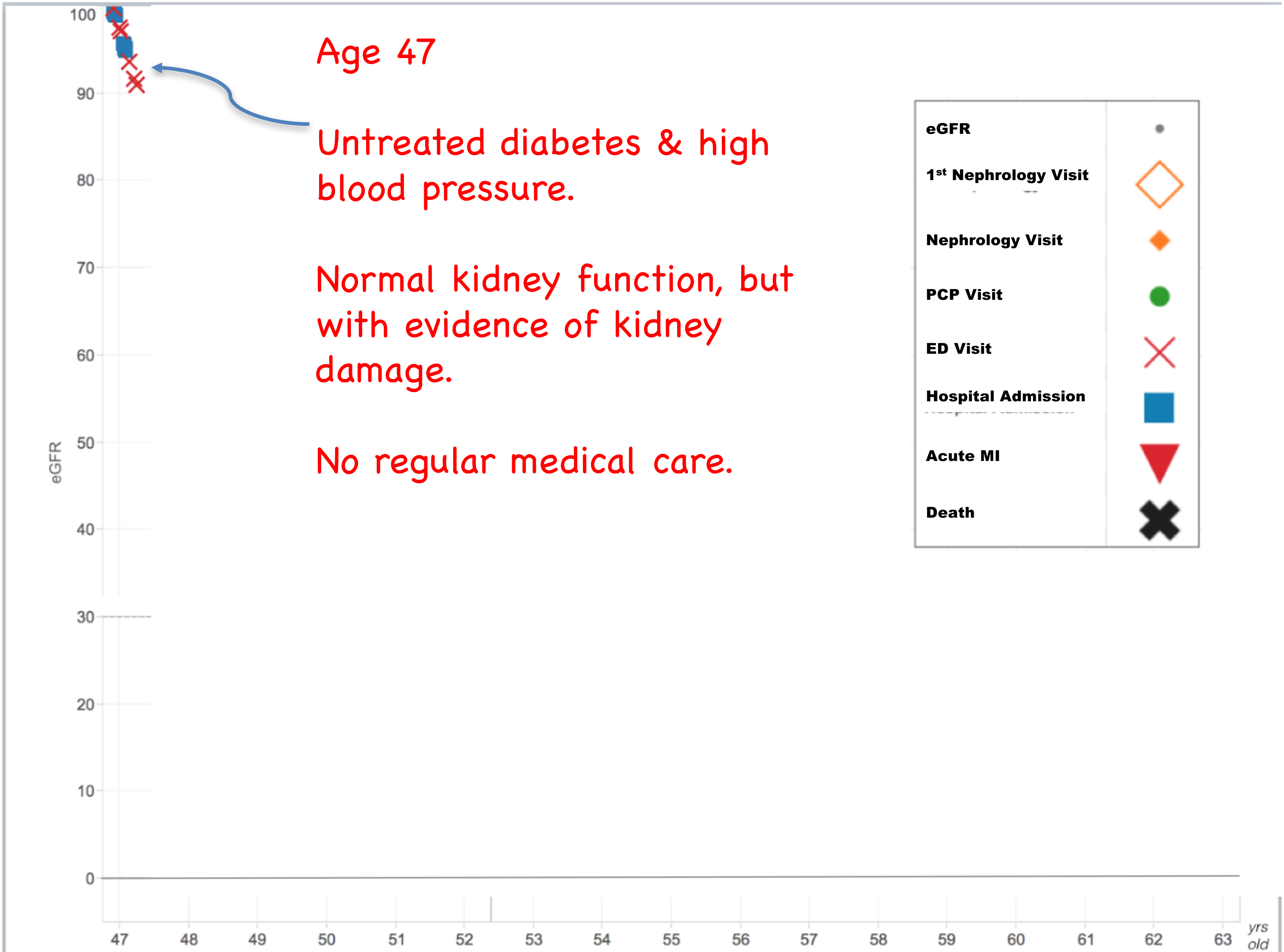
Fan et al, KDD 2016

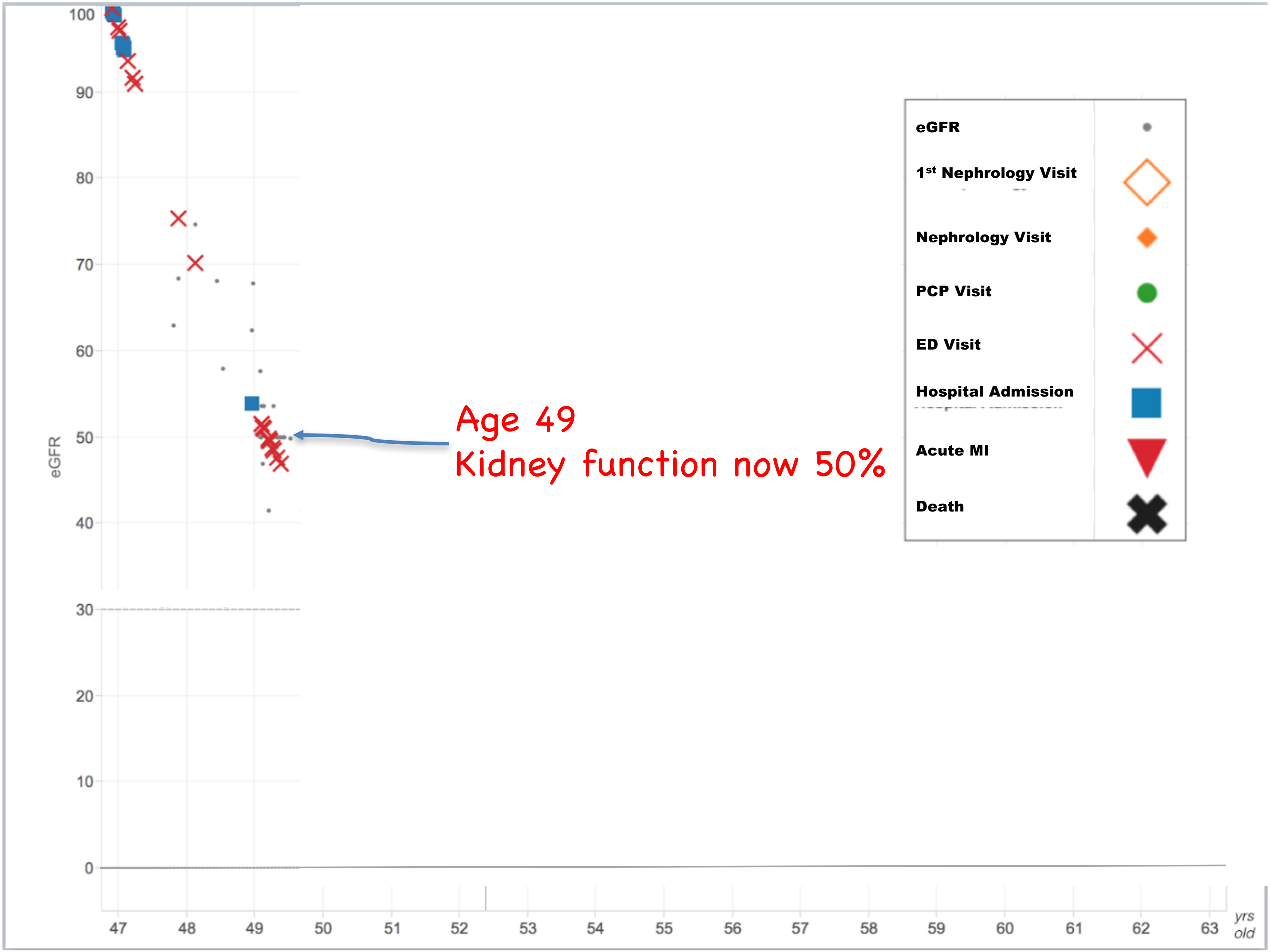
Extensions

- Other applications of the GCHMM to network data, like modeling the influence of opinions.
- Scale up to larger communities (Fan et AAAI 2016)
- Nonparametric Bayesian extensions
- Other factors that may influence contracting an infection
- Learn latent network structure

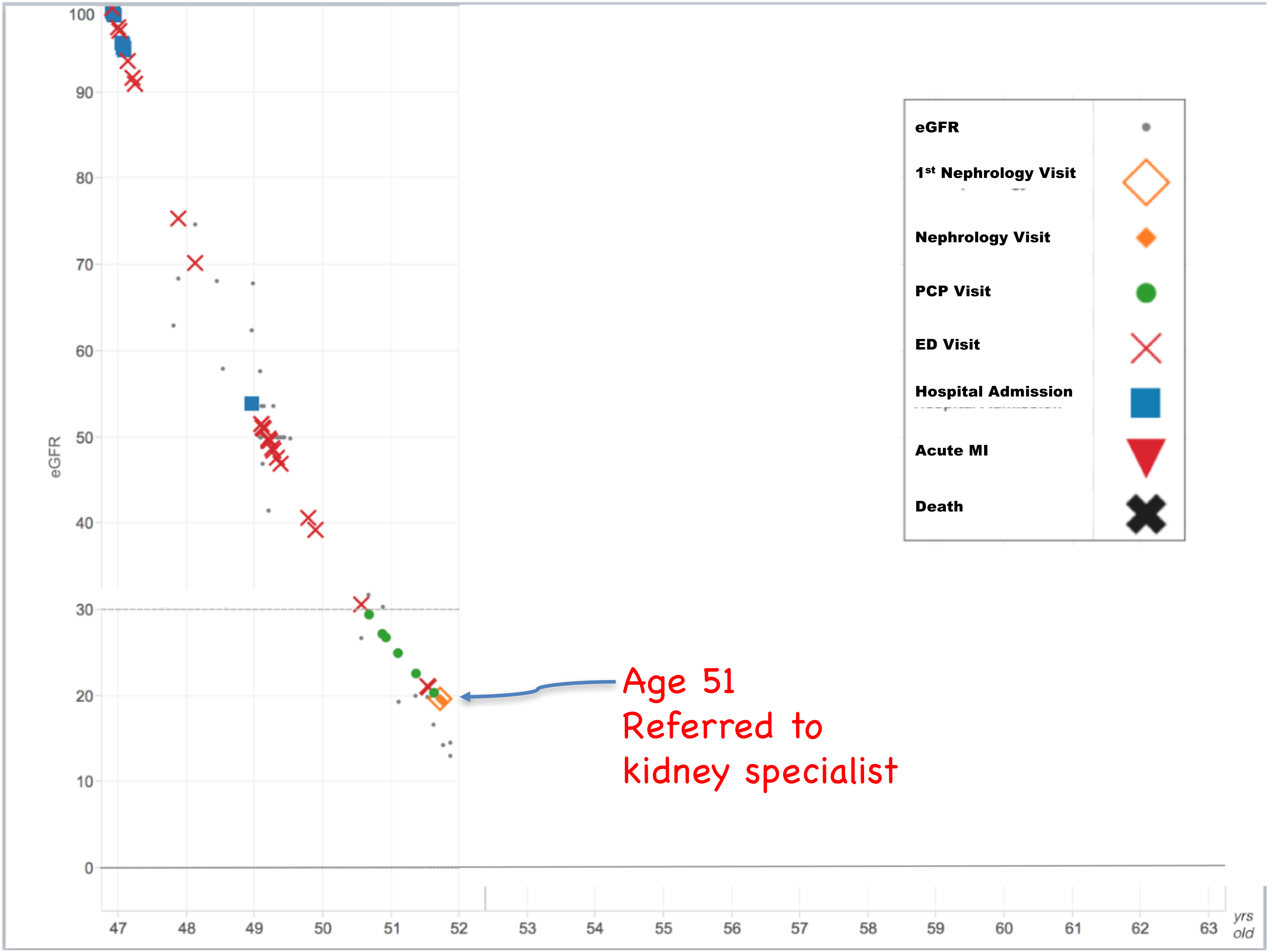
More ML for Healthcare Work


Kidney Function (eGFR)
Age
eGFR
1st Nephrology Visit
Nephrology Visit
PCP Visit
ED Visit
Hospital Admission
Acute MI
Death
100
90
80
70
60
50
40
30
20
10
0
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
yrs old

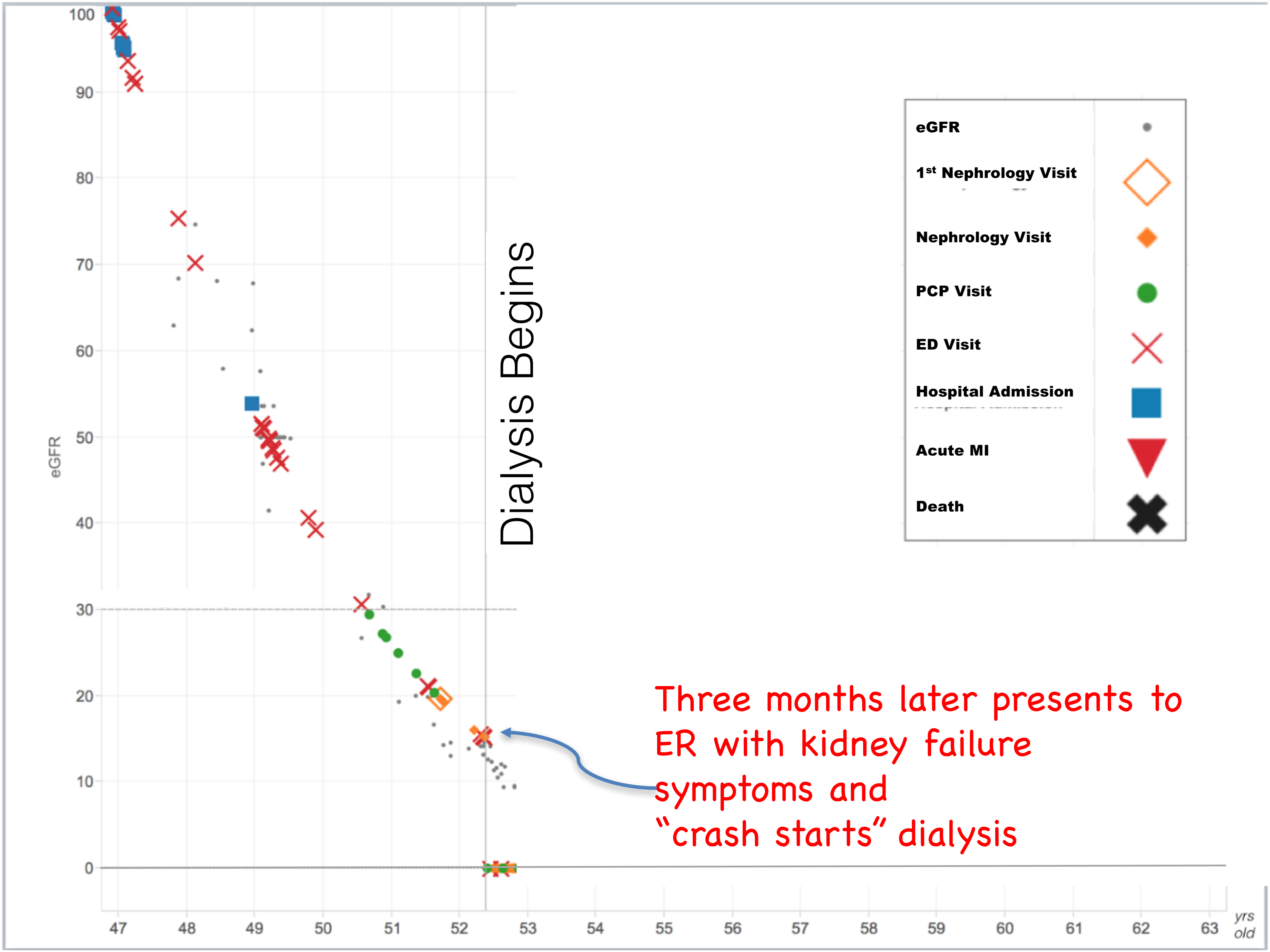

Age 47
Untreated diabetes & high blood pressure.
Normal kidney function, but with evidence of kidney damage.
No regular medical care.
eGFR
100
90
80
70
60
50
40
30
20
10
0
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
yrs old
eGFR
1st Nephrology Visit
Nephrology Visit
PCP Visit
ED Visit
Hospital Admission
Acute MI
Death



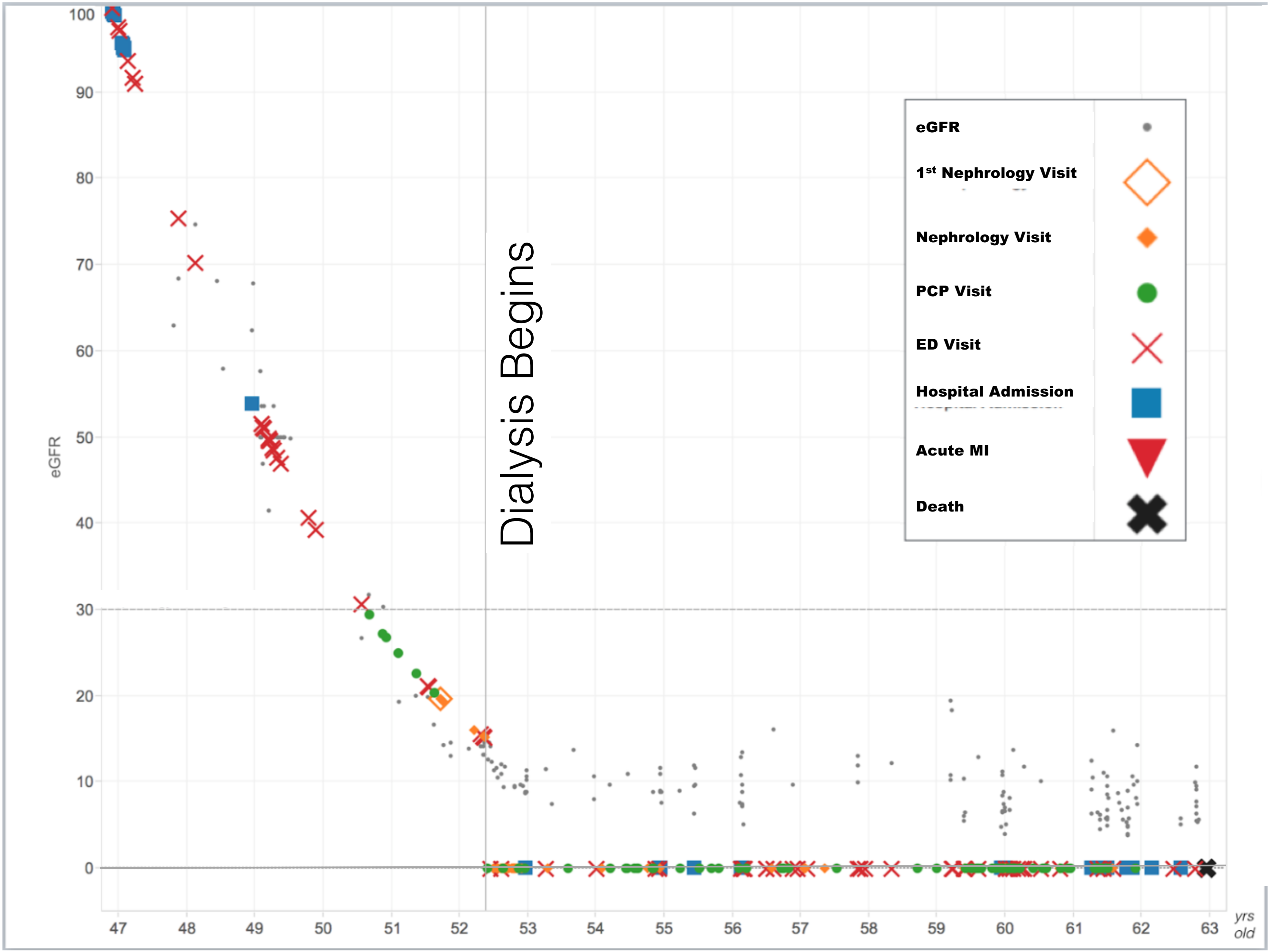
eGFR
100
90
80
70
60
50
40
30
20
10
0
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
yrs old
Age 49
Kidney function now 50%
eGFR
1st Nephrology Visit
Nephrology Visit
PCP Visit
ED Visit
Hospital Admission
Acute MI
Death



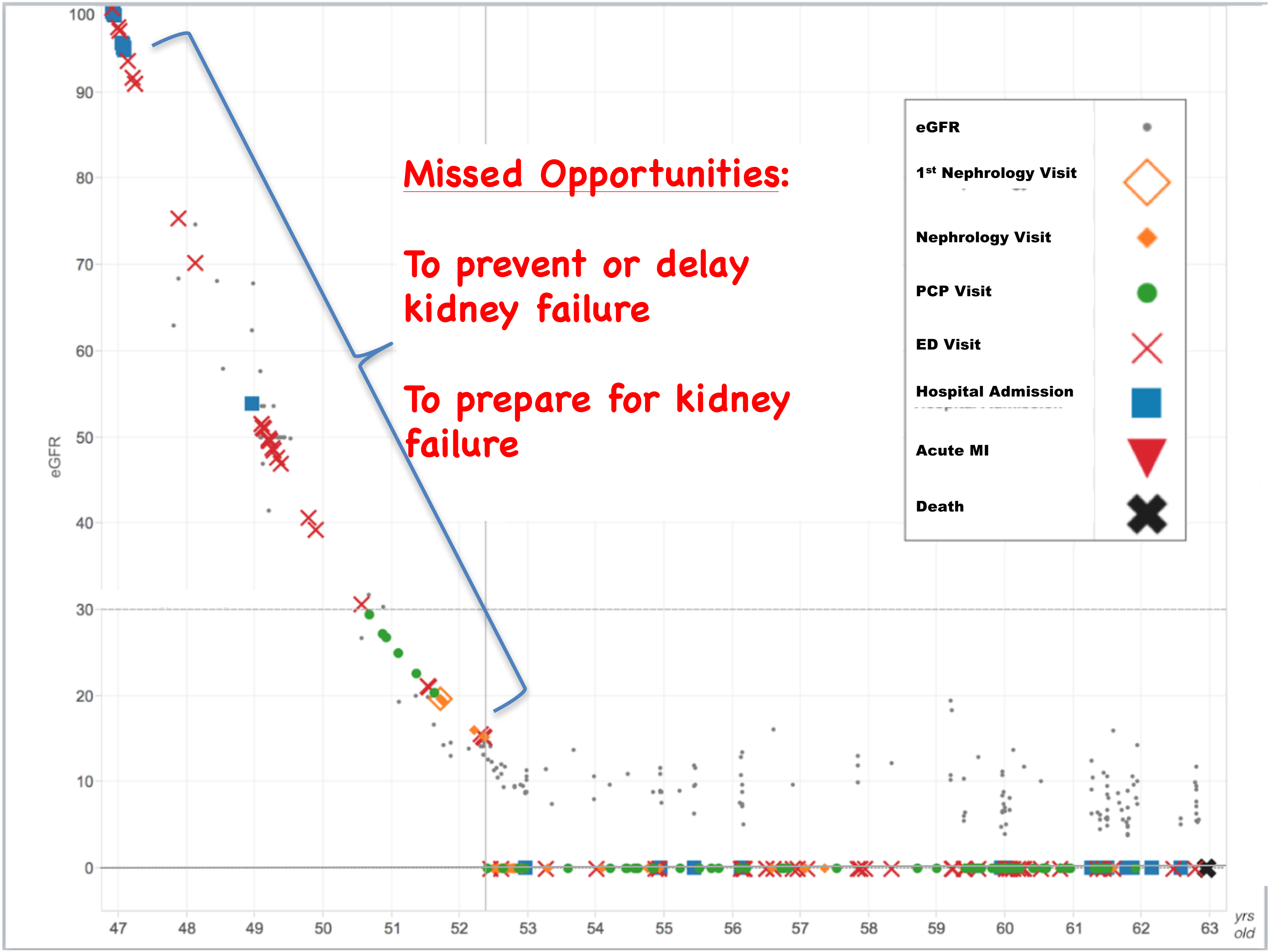
eGFR
100
90
80
70
60
50
40
30
20
10
0
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
yrs old
eGFR
1st Nephrology Visit
Nephrology Visit
PCP Visit
ED Visit
Hospital Admission
Acute MI
Death
Age 51
Referred to
kidney specialist

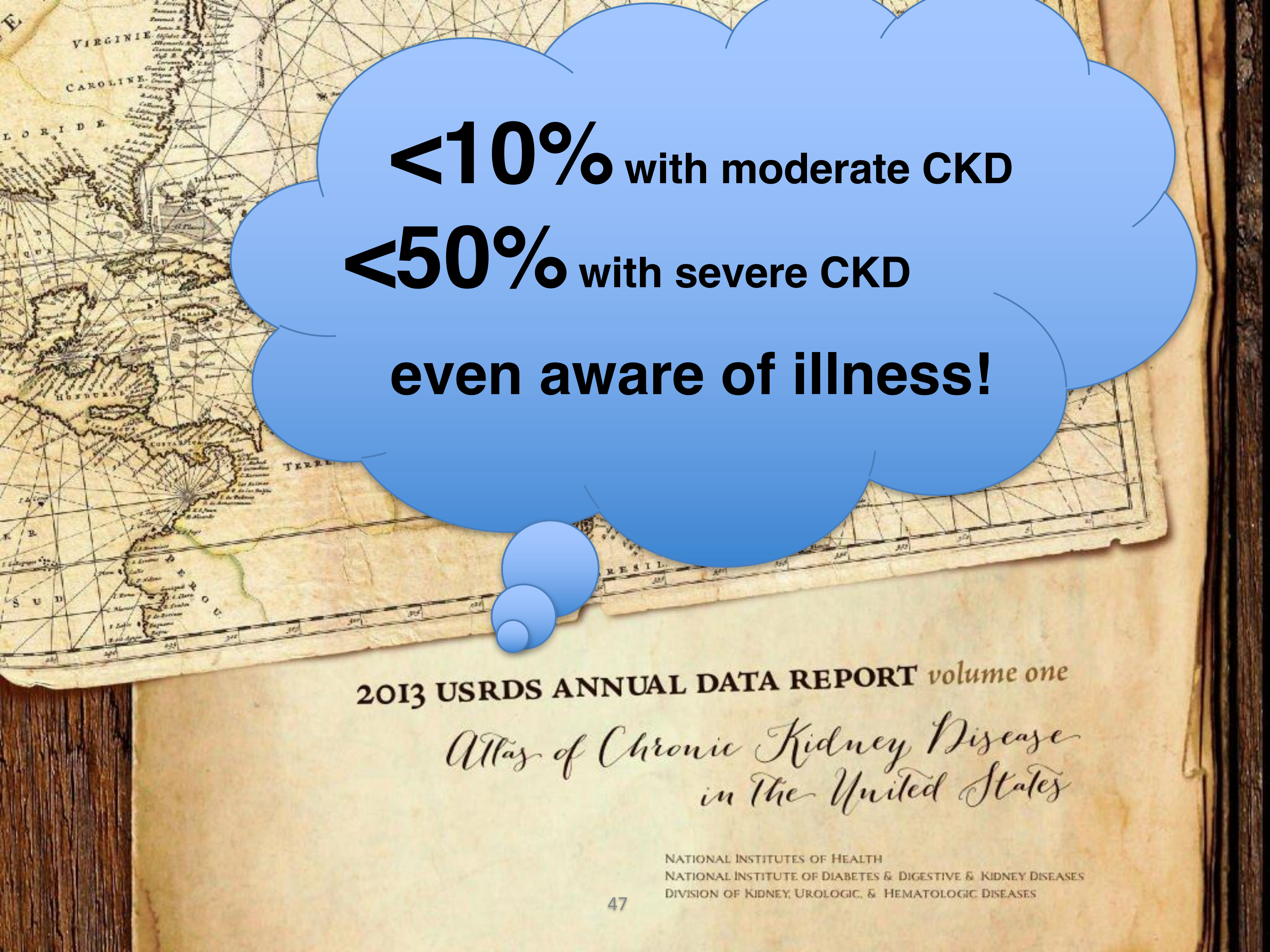


eGFR
100
90
80
70
60
50
40
30
20
10
0
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
yrs old
Dialysis Begins
eGFR
1st Nephrology Visit
Nephrology Visit
PCP Visit
ED Visit
Hospital Admission
Acute MI
Death
Three months later presents to ER with kidney failure symptoms and "crash starts" dialysis



eGFR
100
90
80
70
60
50
40
30
20
10
0
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
yrs old
Dialysis Begins
eGFR
1st Nephrology Visit
Nephrology Visit
PCP Visit
ED Visit
Hospital Admission
Acute MI
Death


Missed Opportunities:
To prevent or delay kidney failure
To prepare for kidney failure
eGFR
1st Nephrology Visit
Nephrology Visit
PCP Visit
ED Visit
Hospital Admission
Acute MI
Death
eGFR
100
90
80
70
60
50
40
30
20
10
0
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
yrs old


<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

Model for a single trajectory

Conditional likelihood factorizes across *P* labs:

$$p(\vec{y_i}|z_i, b_i, c_i; x_i) = \prod_{p=1}^{P} p(\vec{y_{ip}}|z_i, b_i, c_i; x_i)$$

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

$$\mu_{ip}(t) \sim \mathcal{GP}(\Lambda^{(p)} x_i$$

Population effect

$$+ \Phi_z(t)^\top \beta^{(p)}_{\boxed{z_{ip}}}$$

Latent subpopulation curve

$$+ \Phi_l(t)^\top b_{ip},$$

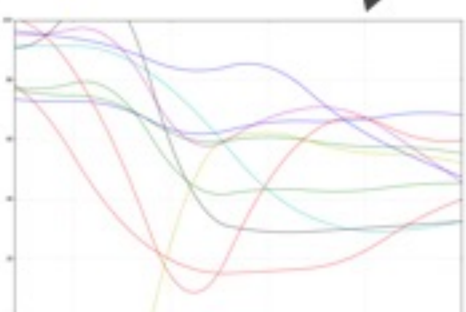
Individual long-term deviations

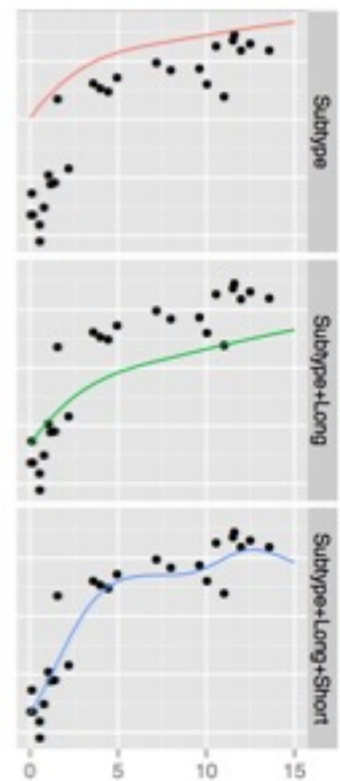
$$K_p)$$

Individual transient deviations (GP)

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

Curves per subtype





Futoma et al, MLHC 2016

Chronic Kidney Disease (CKD)

Heart disease

Diabetes

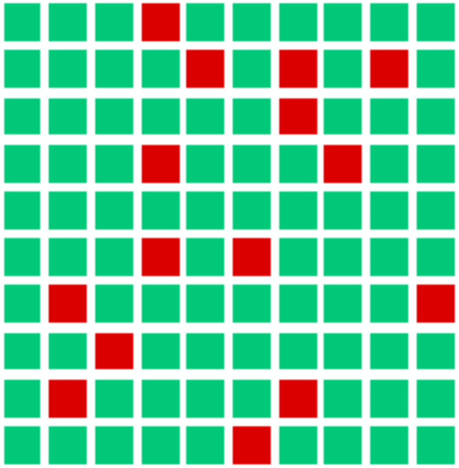
Proposed Joint Model

- 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:

$$p(\vec{y_i}, \vec{u_i} | z_i, b_i, f_i, v_i; x_i) = p(\vec{y_i} | z_i, b_i, f_i; x_i) p(\vec{u_i} | z_i, b_i, f_i, v_i; x_i)$$

Futoma et al, UAI 2016

surgical complications

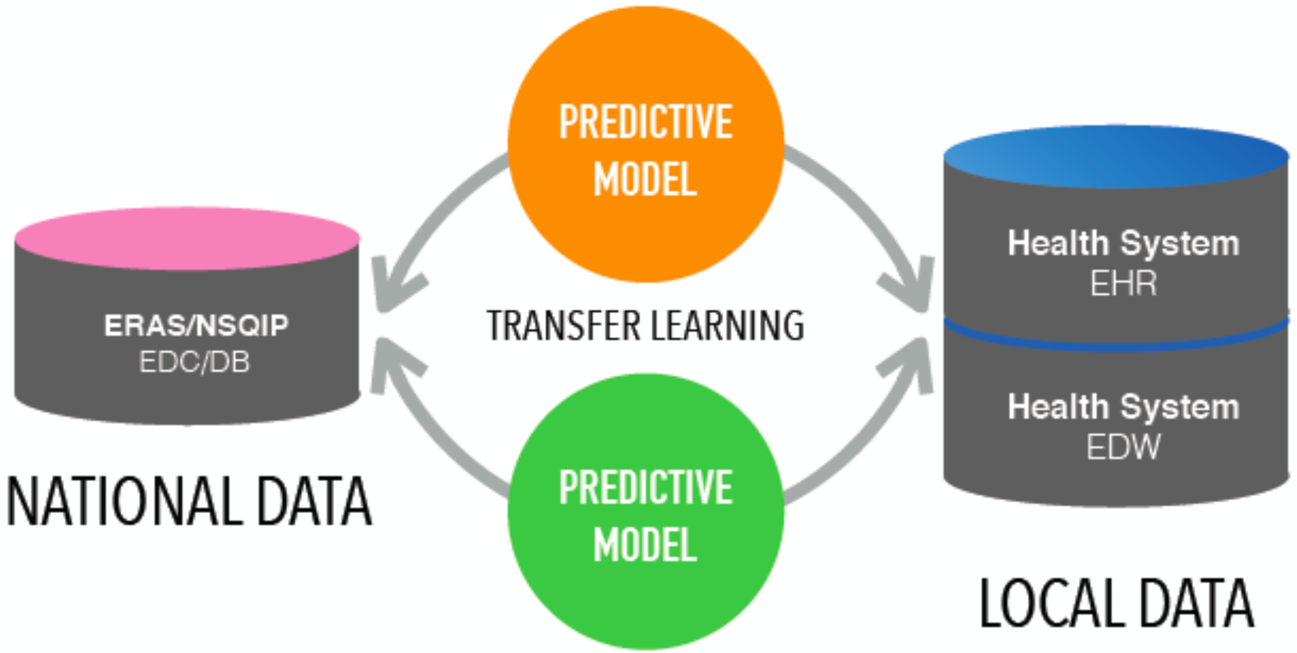


approximately 15 out of every 100 surgical procedures performed results in a complication

American College of Surgeons
National Surgical
Quality Improvement Program
(ACS NSQIP®)

prediction: learning relationship between predictors & outcomes

OUTCOME	most predictive variables	% risk increase
30-day Mortality	ASA Class 5	55%
	Totally dependent functional status	19%
	Preoperative septic shock	18%
	DNR status	17%
	Preoperative ventilator dependence	9%
	Liver disease (varices or ascites)	9%
30-day Any Morbidity	Dx-Esophageal cancer	27%
	Totally dependent functional status	25%
	Preoperative septic shock	24%
	Dx-Nutritional deficiency	21%
	Dx-Injury	21%
	ASA Class 4	19%



CONTINUOUS LEARNING

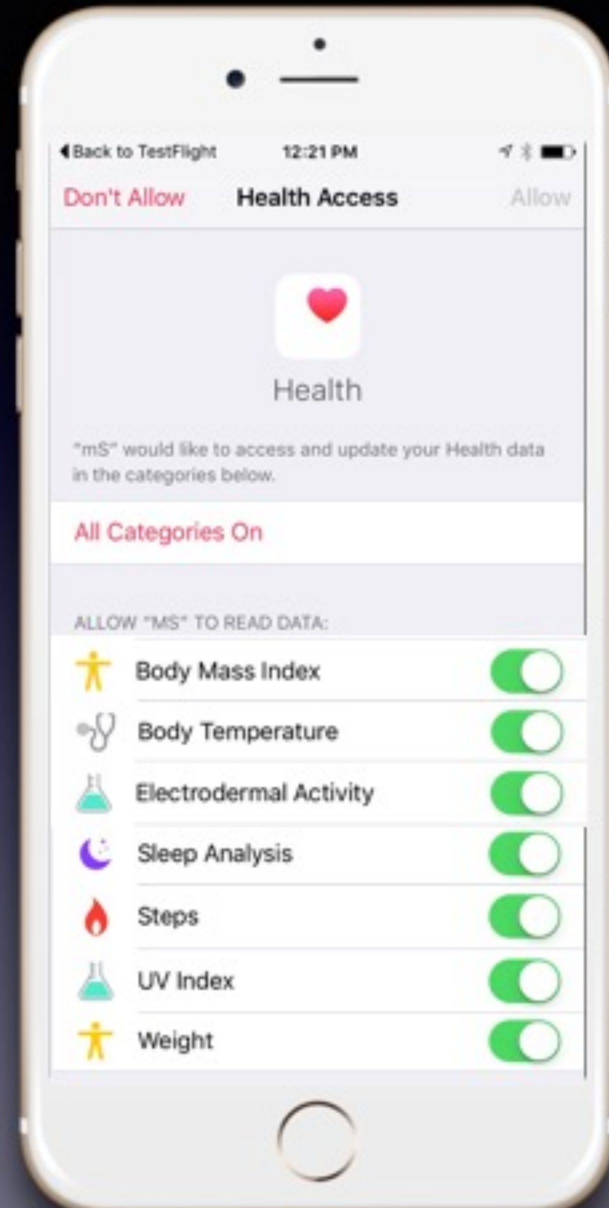
Outcomes (2 of 8)	AUC - NoTransferLearning	AUC- TransferLearning
Pneumonia	0.832	**0.848**
Cardiac	0.909	**0.920**

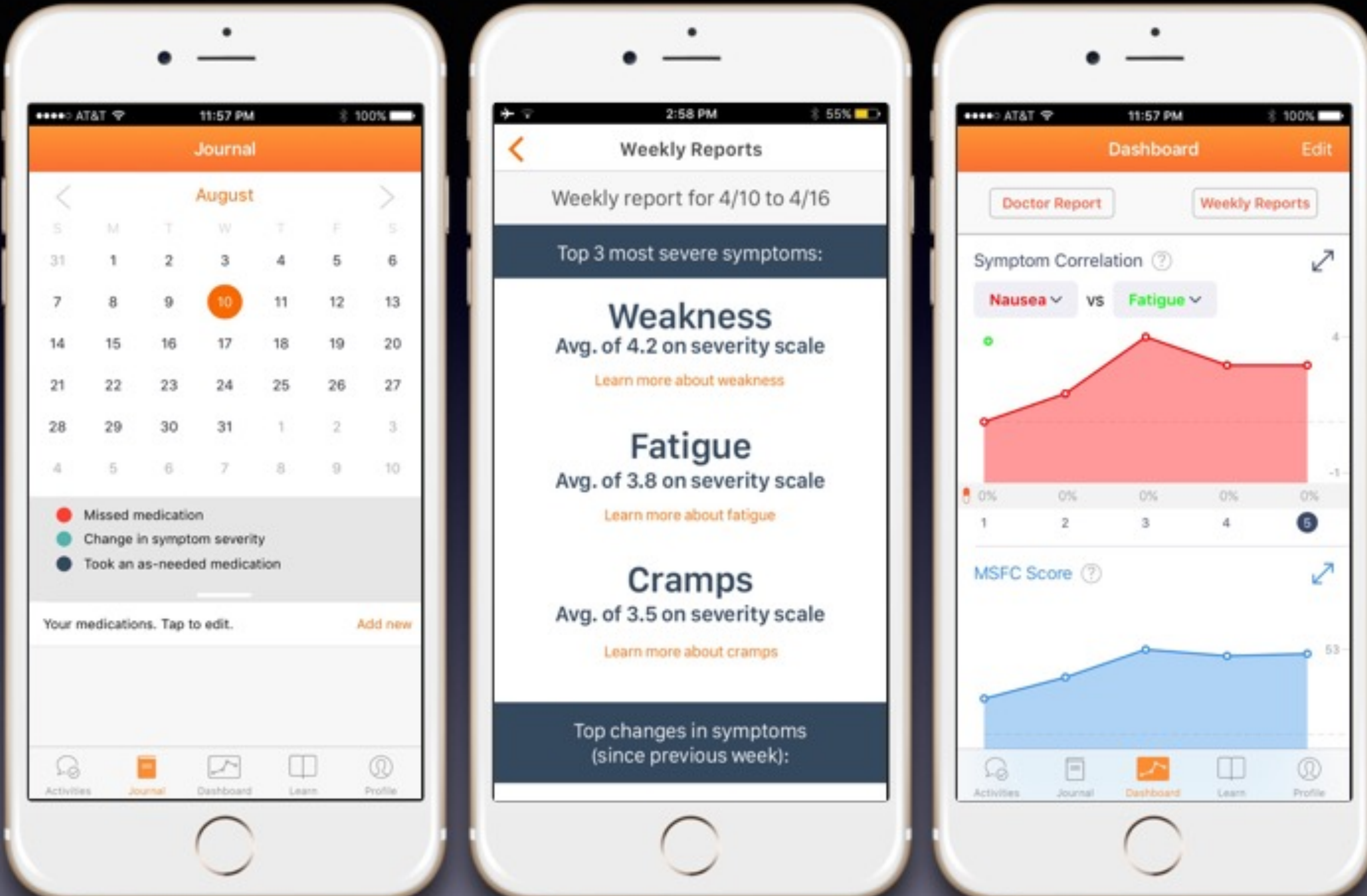
MS Mosaic App

Passively collects HealthKit data

82 Different Data Types (>10 thought to affect MS)

Tagged with date, time, and provenance




AT&T
11:57 PM
100%
Journal
August
31 1 2 3 4 5 6
7 8 9 10 11 12 13
14 15 16 17 18 19 20
21 22 23 24 25 26 27
28 29 30 31 1 2 3
4 5 6 7 8 9 10
Missed medication
Change in symptom severity
Took an as-needed medication
Your medications. Tap to edit.
Add new
Activities
Journal
Dashboard
Learn
Profile
2:58 PM
55%
Weekly Reports
Weekly report for 4/10 to 4/16
Top 3 most severe symptoms:
Weakness
Avg. of 4.2 on severity scale
Learn more about weakness
Fatigue
Avg. of 3.8 on severity scale
Learn more about fatigue
Cramps
Avg. of 3.5 on severity scale
Learn more about cramps
Top changes in symptoms
(since previous week):
AT&T
11:57 PM
100%
Dashboard
Edit
Doctor Report
Weekly Reports
Symptom Correlation
Nausea
vs
Fatigue
0% 0% 0% 0% 0%
1 2 3 4 5
MSFC Score
53
Activities
Journal
Dashboard
Learn
Profile

Initial Analyses

Develop a sparse logistic regression model for predicting the likelihood of each symptom experience

Incorporate a hierarchical layer based on Gaussian processes for modeling time series data (e.g. sleep)

Discover hidden subpopulations within symptoms (using clustering methodology, such as Dirichlet Process mixture models)

Evaluate the efficacy of symptom interventions using longitudinal models and clinical trials

Future Work

- Further work in machine learning for health care
  - Sepsis
  - Congestive Heart Failure
  - Surgical Transplants
- Implement infectious disease work locally
  - Develop models for data from more diverse sensors
  - Causal Inference
- Additional sensors on other groups related to physiometrics
  - Does heart rate impact basketball players shot percentage
- Other social network applications – like online bullying
- Other joint modeling applications – measuring diverse data sets from e.g. different sensory modalities (and animals) to infer functional neural networks in the brain.

Future Work

- Role of other media in ability to make health predictions
  - Xbox
  - Cortana
  - Bing
- Incorporation of Microsoft Health devices into research
  - Microsoft Band
  - Development of machine learning for Microsoft apps
  - HealthVault Insights
- Incorporation of behavior and choice into patient feedback
  - Medication Adherence
  - Exercise
  - Health bot
- Joint modeling of Microsoft data with other data (e.g. NIH)

Collaborators

- Charles Blundell
- Hanna Wallach
- Richard Guo
- Alan Qi
- Vinayak Rao
- Xi Tan
- Syed Naqvi
- Jerry Zhu
- Wei Zhang
- Fan Bu

Kai Fan
Allison Aiello
Marissa Eisenberg
Alison Walsh
Sandy Pentland
Wen Dong

Joseph Futoma
Elizabeth Lorenzi
Blake Cameron
Mark Sendak
Lee Hartsell
Erich Huang
Jeff Sun
Sanjay Hariharan