Scalable Modeling of Multivariate Longitudinal Data

for Prediction of Chronic Kidney Disease Progression

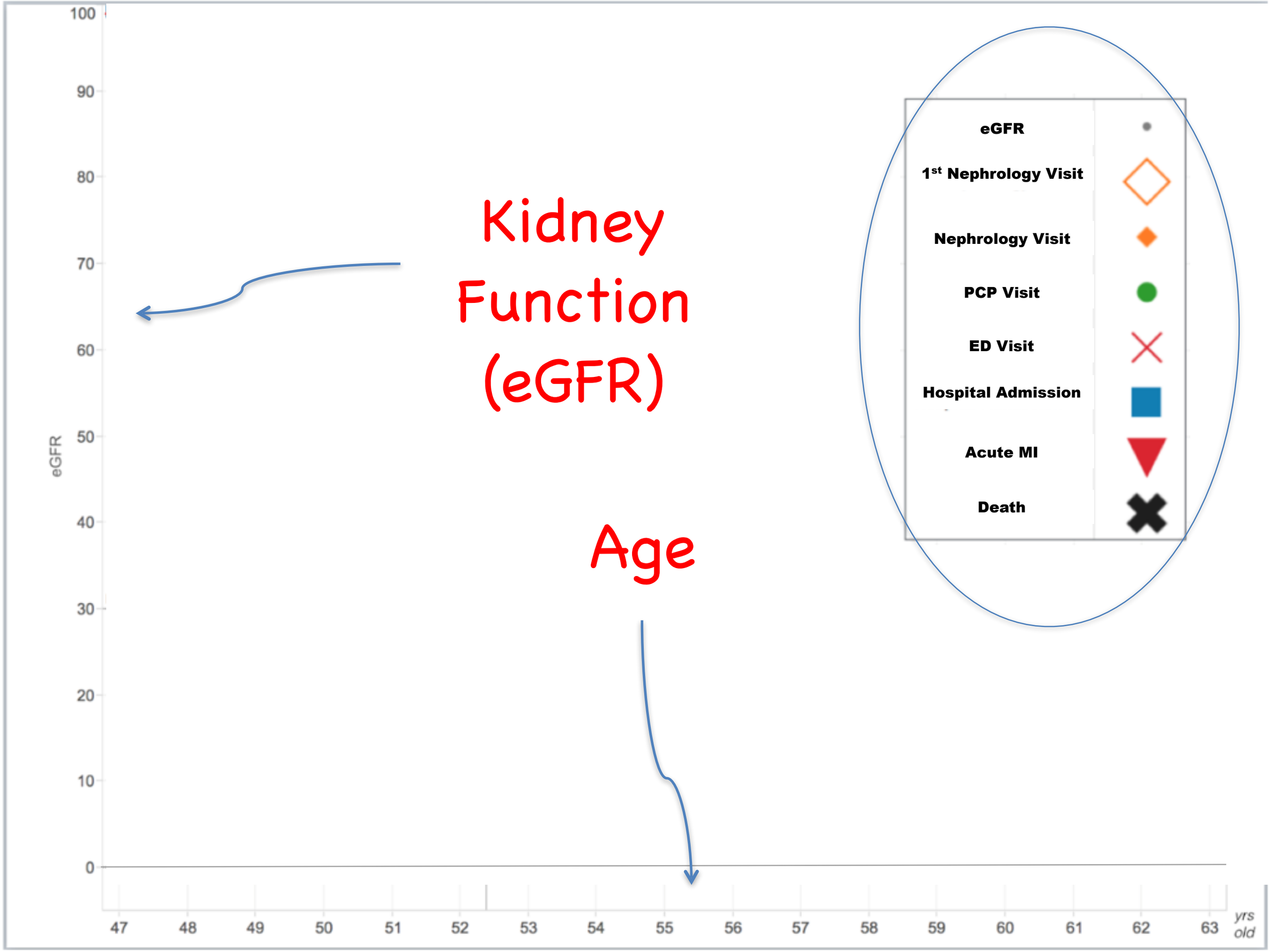
Joe Futoma

Mark Sendak

Blake Cameron

Katherine Heller

Duke UNIVERSITY

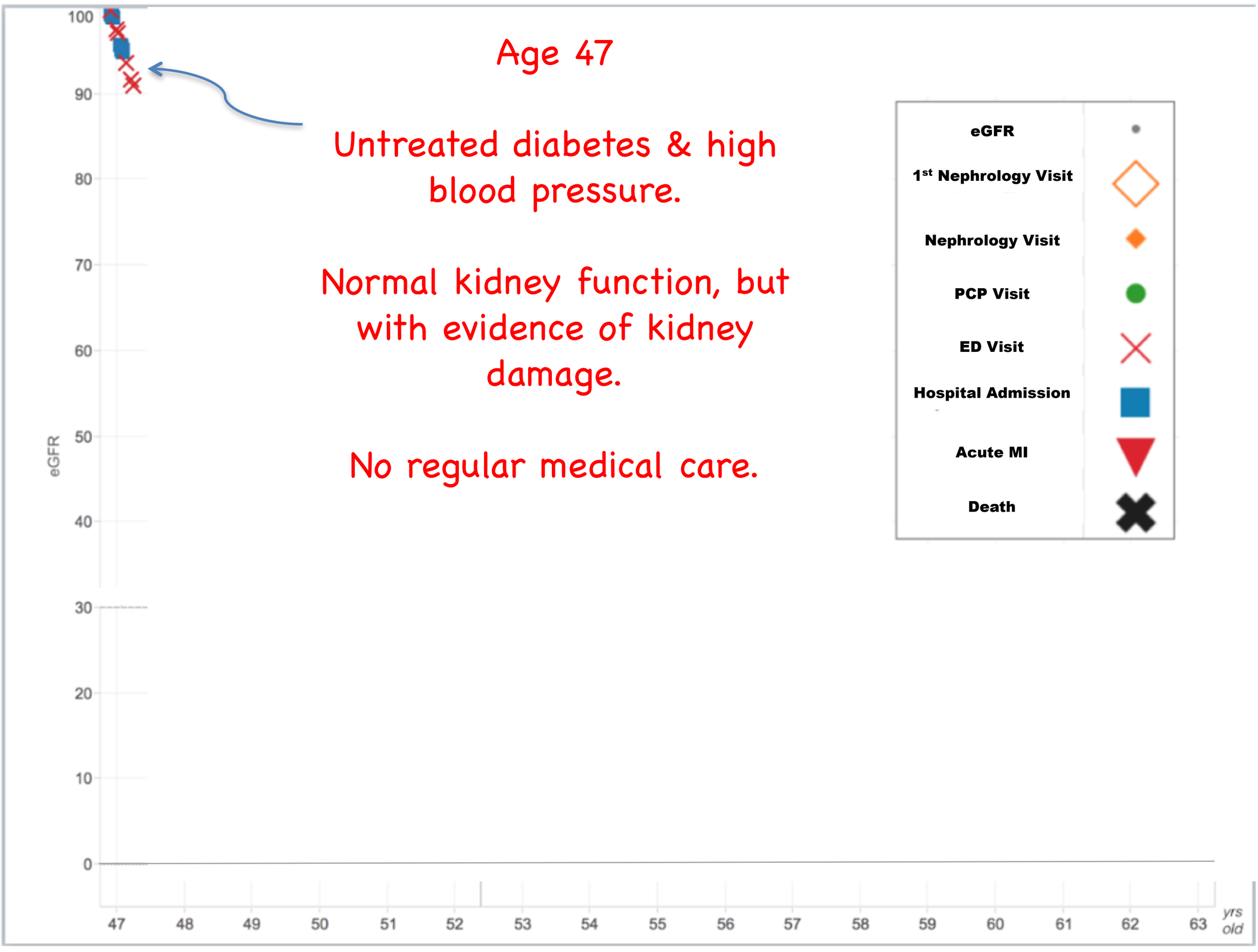

100
90
80
70
60
50
40
30
20
10
0
eGFR
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
yrs old
Kidney Function (eGFR)
Age
eGFR
1st Nephrology Visit
Nephrology Visit
PCP Visit
ED Visit
Hospital Admission
Acute MI
Death

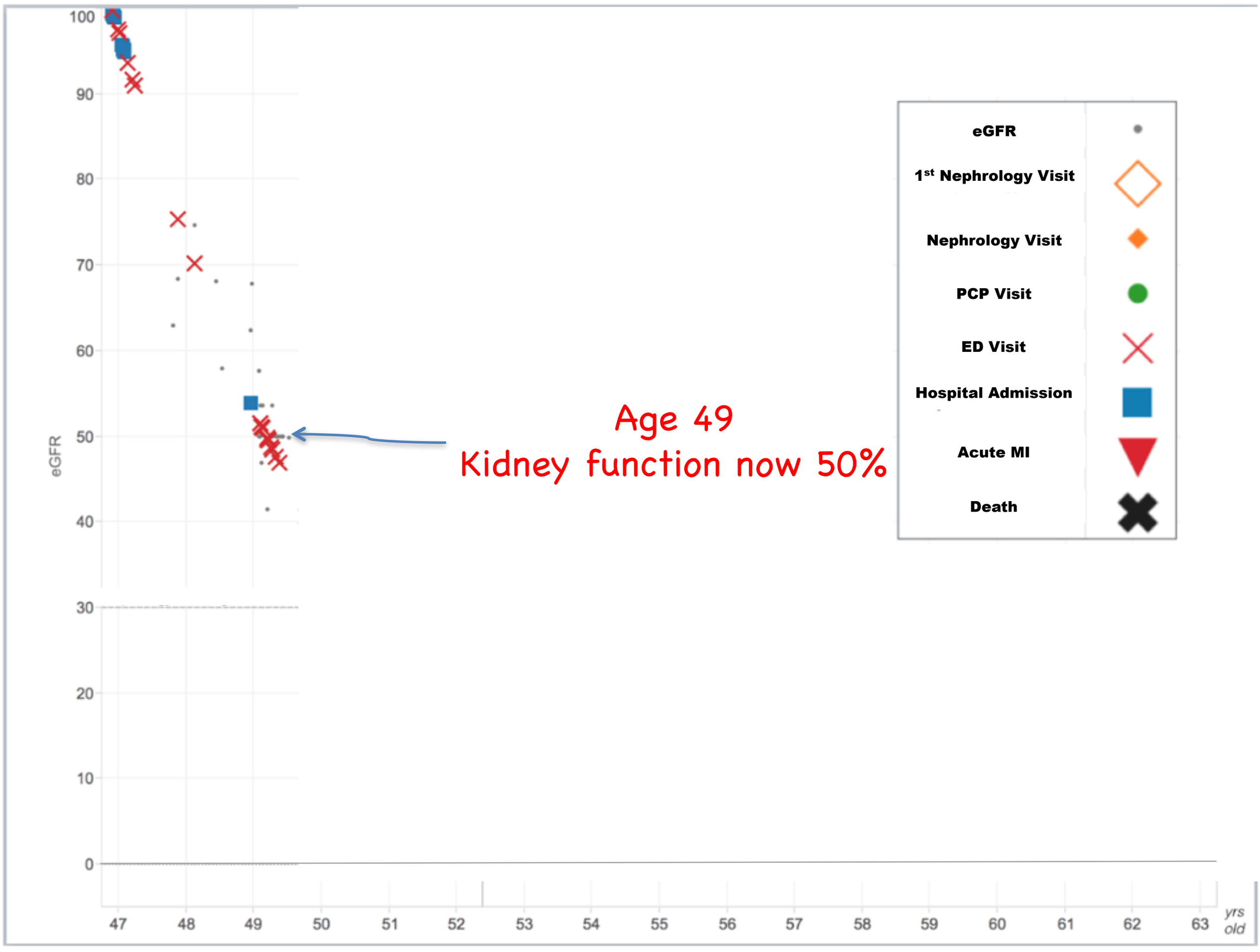
Age 47

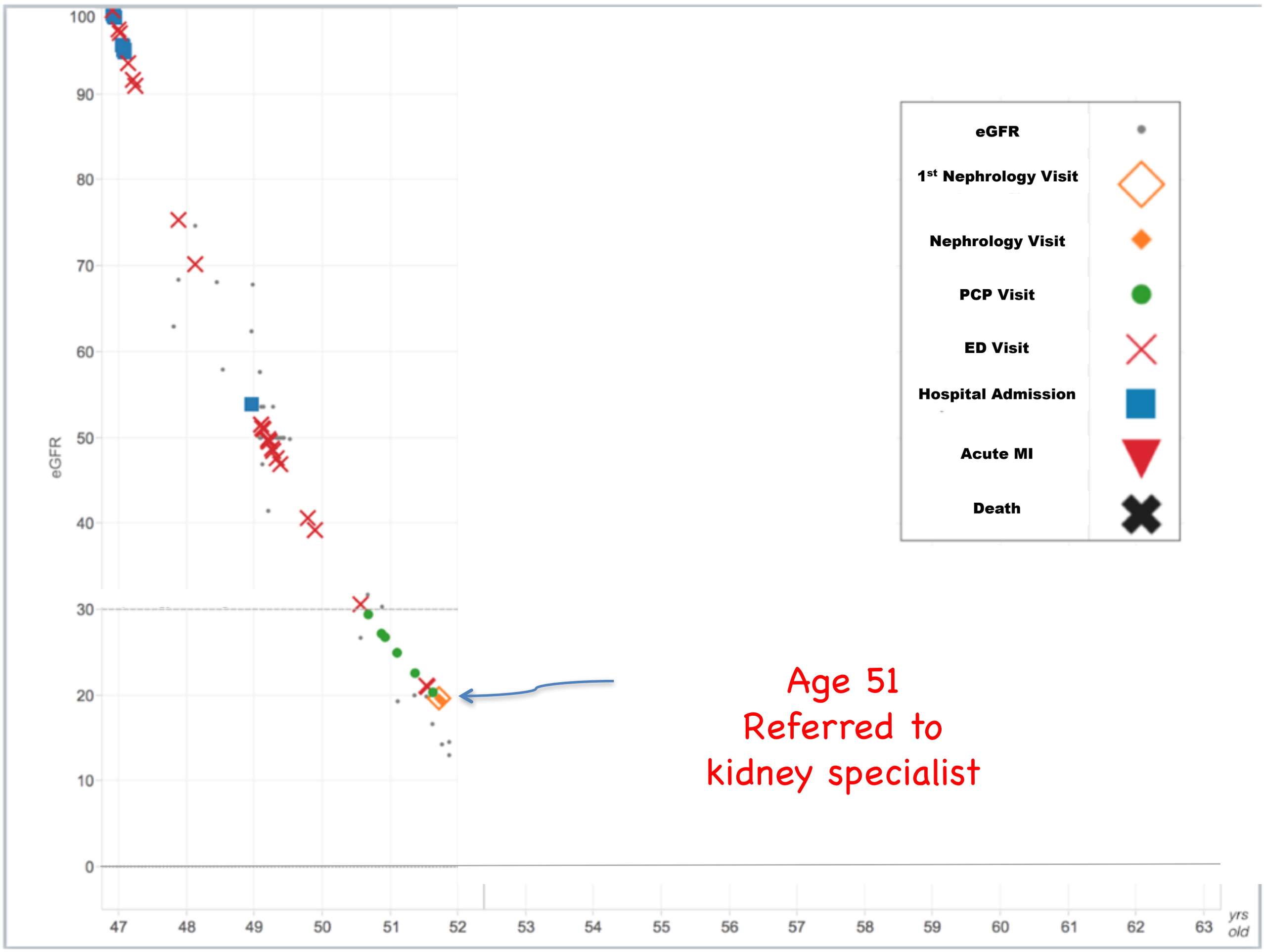
Untreated diabetes & high blood pressure.

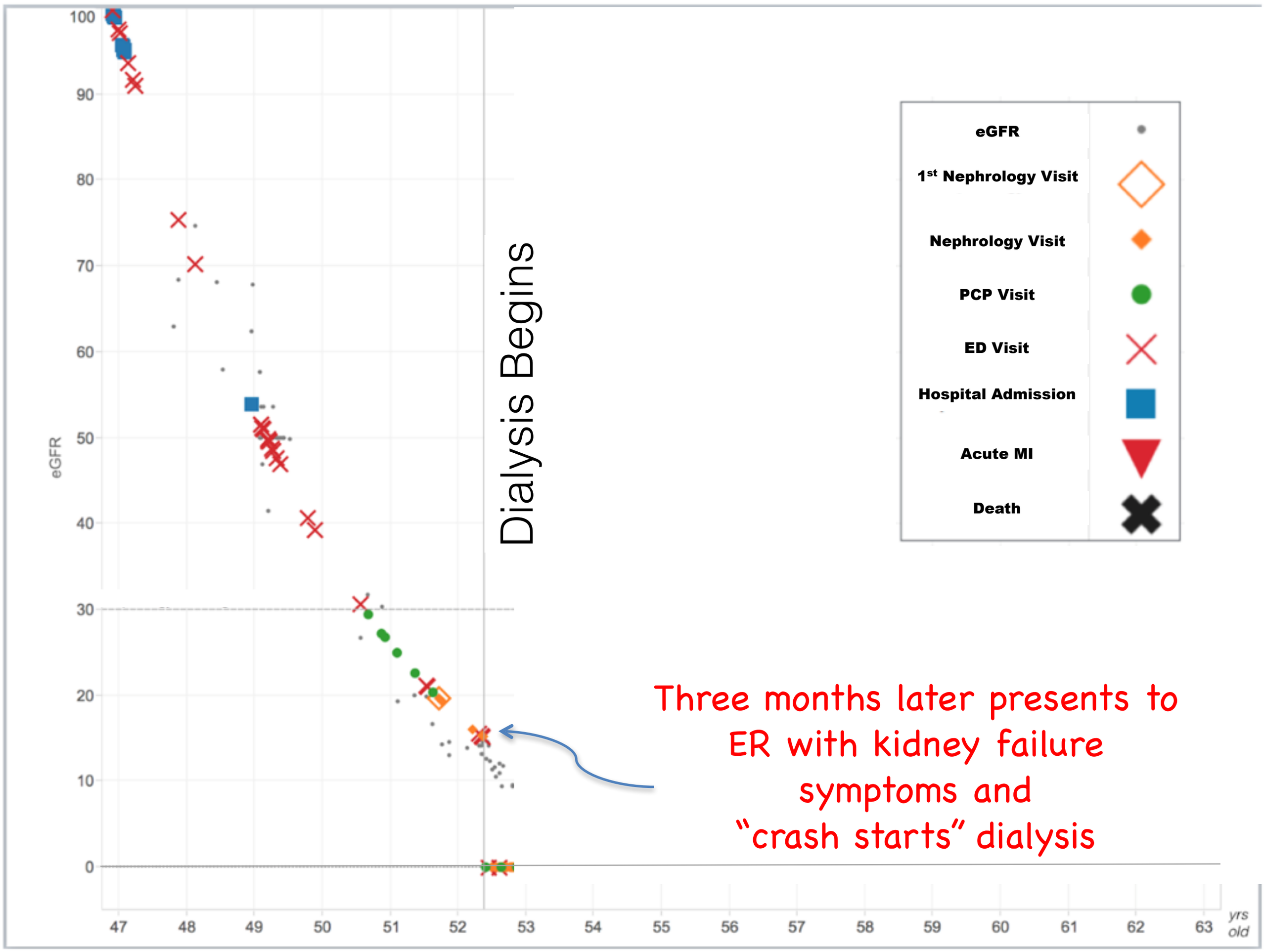
Normal kidney function, but with evidence of kidney damage.

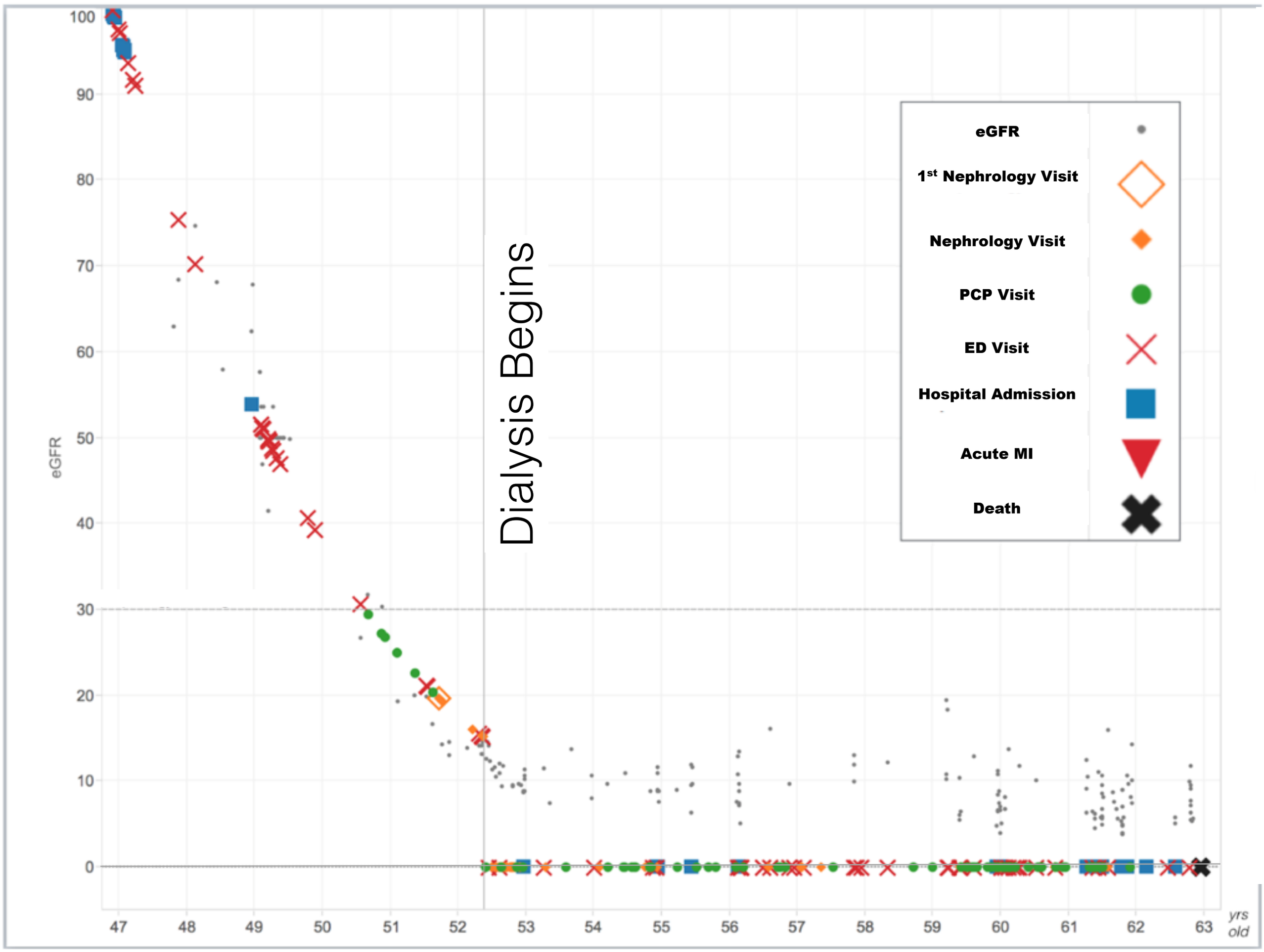
No regular medical care.




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
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

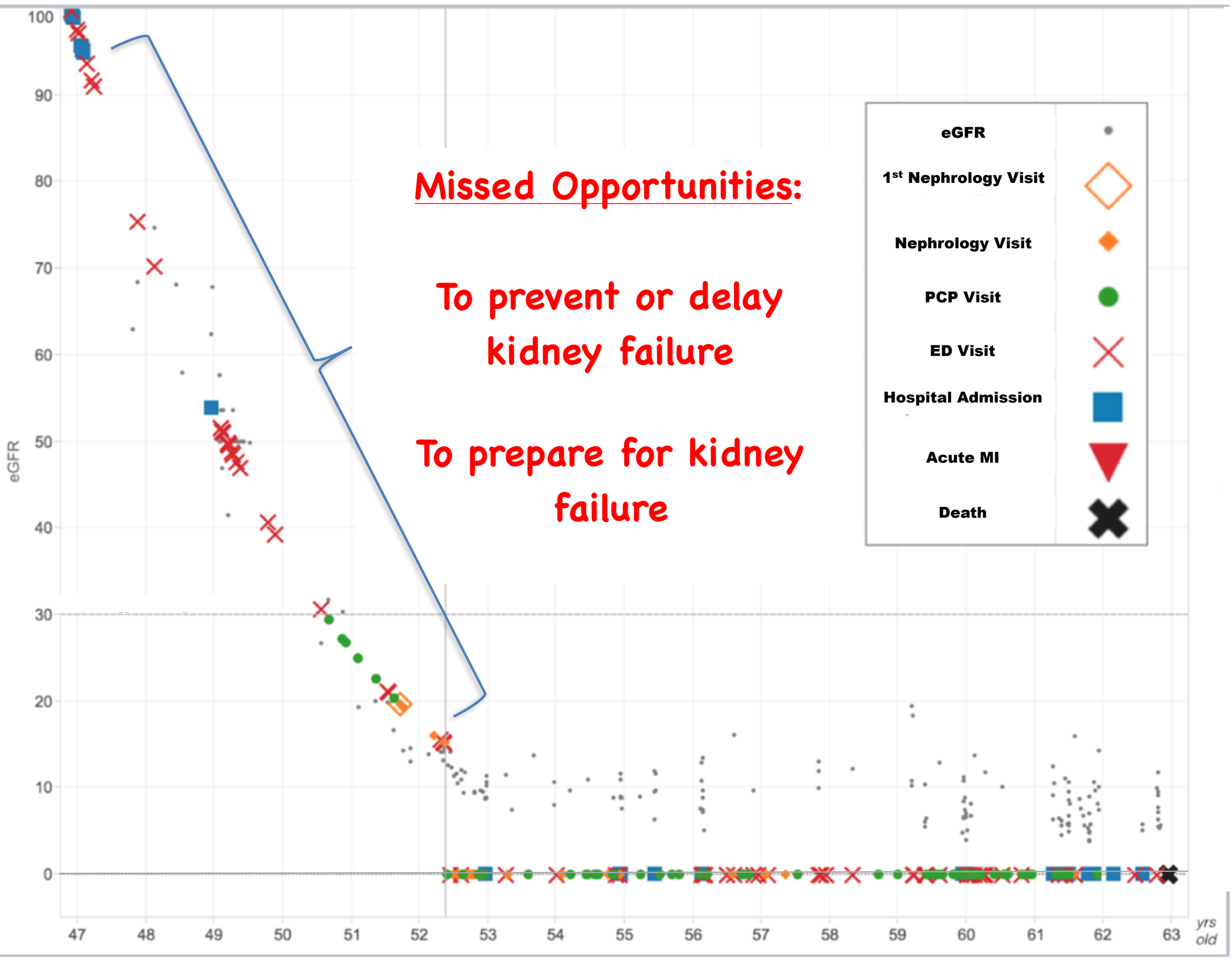

eGFR
100
90
80
70
60
50
40
30
20
10
0
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
yrs old
Dialysis Begins
eGFR
1st Nephrology Visit
Nephrology Visit
PCP Visit
ED Visit
Hospital Admission
Acute MI
Death
Three months later presents to ER with kidney failure symptoms and "crash starts" dialysis


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



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

American Journal of Nephrology

Am J Nephrol 2012;36:1–10
DOI: 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

- Can we use other clinical data (labs and/or vitals) to improve prediction of disease progression?
- Goal: flexible model for multivariate longitudinal data

Model for a single trajectory

Conditional likelihood factorizes across P labs: $p(\vec{y_i}|z_i, b_i, c_i; x_i) = \prod_{p=1}^{P} p(\vec{y_{ip}}|z_i, b_i, c_i; x_i)$

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

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

Population effect

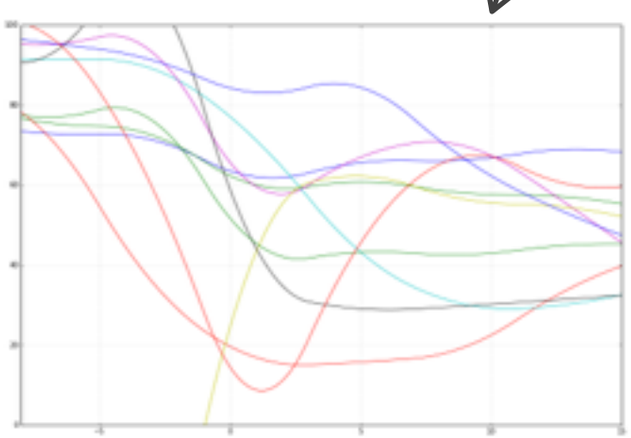
Latent subpopulation curve

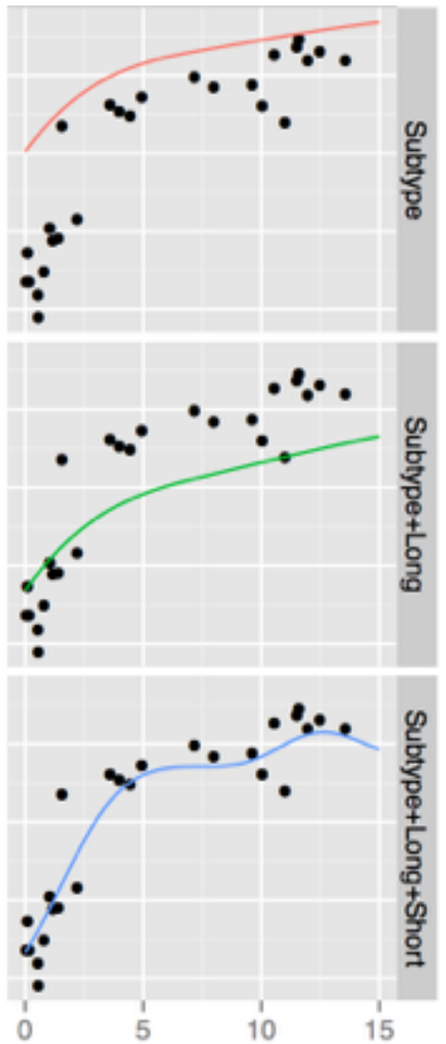
Individual long-term deviations

Individual transient deviations (GP)

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

Curves per subtype





Inducing Dependence

Dependence between mean functions
for the P labs in 2 ways:

$$\mu_{ip}(t) \sim \mathcal{GP}(\Lambda^{(p)} x_i + \Phi_z(t)^\top \beta^{(p)}_{z_{ip}} + \Phi_l(t)^\top b_{ip}, K_p)$$

Long-term deviations are correlated
via multivariate normal:

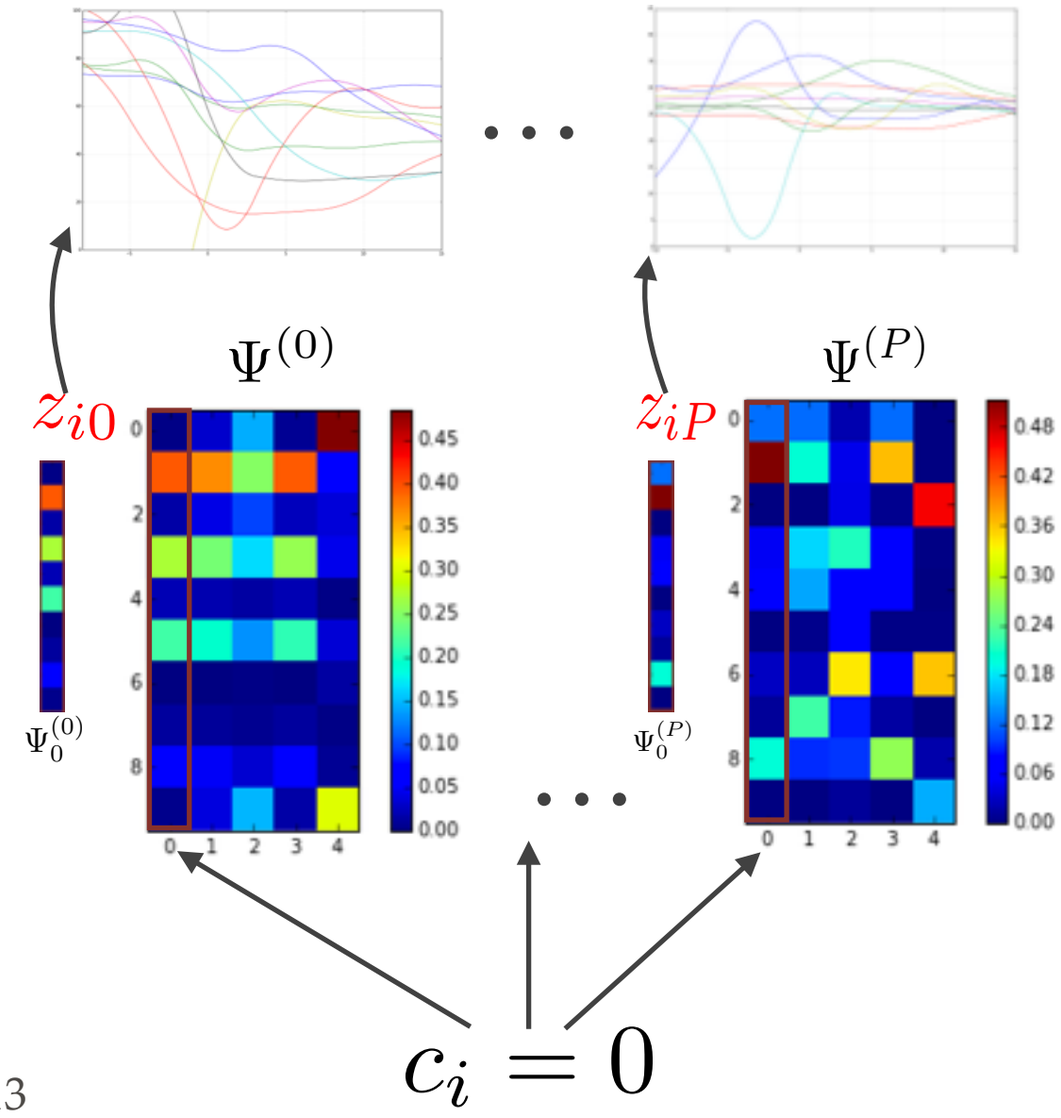
$$\vec{b_i} = (b_{i1}, \ldots, b_{iP})^\top \sim N(0, \Sigma_b)$$

Subtypes/clusters per lab are correlated
via mixture of multinomials:

$$z_{ip}|c_i = g \sim \text{Multinomial}(\Psi^{(p)}_g)$$

$$c_i \sim \text{Multinomial}(\pi_i)$$

$$\pi_{ig} = \frac{e^{w_g^\top x_i}}{\sum_{g'=1}^{G} e^{w_{g'}^\top x_i}}$$



Experimental Setup

- 6 variables of interest: eGFR, 5 other labs relevant to CKD
- Cohort of 44,000 patients at Duke with at least moderate stage CKD (Stage 3+) and 5+ measurements for eGFR
- For each test patient: use data before *t* to predict future labs
- Evaluation for each lab:
  - average MAE across test patients, in future time windows
- Baseline: [Schulam & Saria, 2015] trained independently

Quantitative Results

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	**8.86***	**10.43***	**12.05**	**13.69**	**8.84*****	**11.08****	**13.23***	**9.39*****	**12.29***
	Proposed	9.12	10.67	12.28	14.21	9.26	11.73	13.99	10.12	13.07
Serum Alb.	Schulam	0.59	0.79	1.09	1.53	0.60	0.88	1.28	0.63	0.96
	Proposed	**0.34*****	**0.39*****	**0.47*****	**0.63*****	**0.35*****	**0.45*****	**0.63*****	**0.40*****	**0.58*****
Serum Bicarb.	Schulam	1.92	2.06	2.13	2.31	1.93	2.06	2.21	1.89	**2.14**
	Proposed	1.87	1.97	2.04	2.31	1.89	1.99	2.31	1.87	2.24
Serum Calc.	Schulam	0.74	1.02	1.62	2.89	0.72	1.26	2.27	0.85	1.53
	Proposed	**0.37*****	**0.44*****	**0.58*****	**0.80*****	**0.39*****	**0.54*****	**0.80*****	**0.46*****	**0.73*****
Serum Phos.	Schulam	1.02	1.35	1.46	1.44	1.17	1.36	1.34	1.13	1.15
	Proposed	**0.57*****	**0.68*****	**0.88*****	**1.23**	**0.65*****	**0.88*****	1.25	**0.82*****	1.23
Urine ACR	Schulam	1.17	1.30	1.44	1.64	1.14	1.30	1.53	1.11	1.41
	Proposed	**0.92*****	**1.02*****	**1.17*****	**1.44**	**0.96*****	**1.13***	1.45	**1.02**	1.42

Conclusion

- Novel model for multivariate longitudinal clinical data
- Current clinical practice: “clinical gestalt”, no evidence-based method to inform decision-making process
- Future work:
  - More flexible dependence between labs
  - Jointly predict with events of interest (hospital admissions / high utilization, cardiac events, etc)
  - More clinically actionable metrics to evaluate models

Thank you!

Duke
UNIVERSITY

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.

Presenting tomorrow at clinical talks!

Duke Statistical Science

NDSEG NATIONAL DEFENSE SCIENCE & ENGINEERING GRADUATE FELLOWSHIP

Duke Institute for Health Innovation

Duke Translational Medicine Institute

Duke UNIVERSITY