# Predicting Renal Failure in Patients with C3 Glomerulopathy

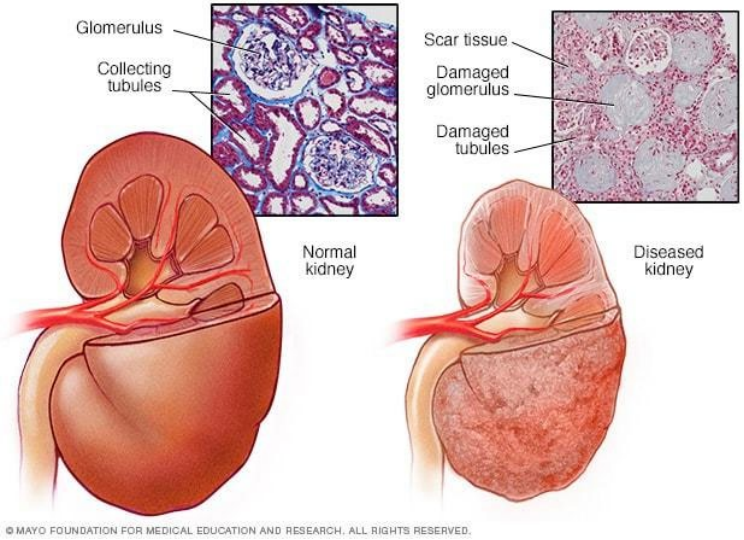
**ISIB 2022**

Bethany Astor, Delaney Underwood, and Alan Wang

Patrick Breheny, PhD and Logan Harris, MS

# Kidney Disease

- Progressive disease in which the kidneys lose their ability to properly filter waste and excess fluids from the blood
- Glomerulus-filtering unit of the kidney
  - Glomerular filtration rate (GFR) is a measure of kidney function
- Current treatment option
  - Kidney Transplant



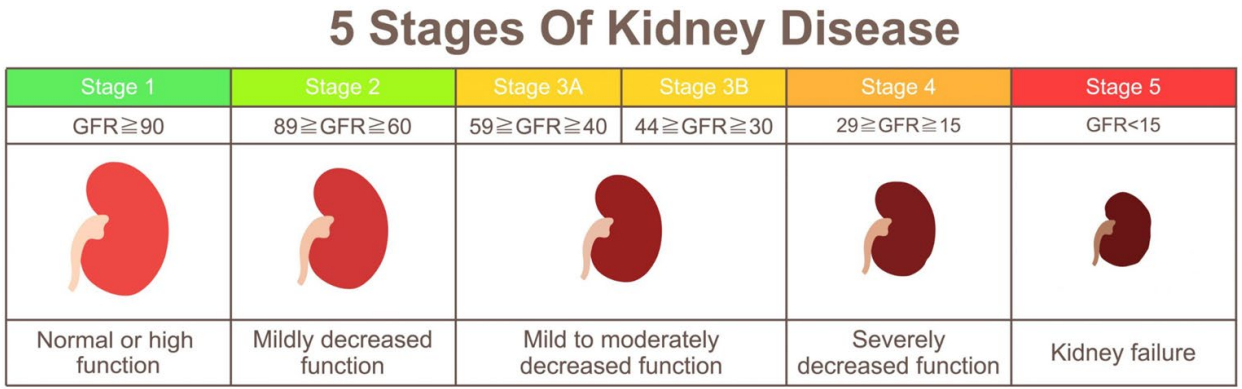

https://www.mayoclinic.org/diseases-conditions/chronic-kidney-disease/symptoms-causes/syc-20354521.

# C3 Glomerulopathy (C3G)

- A form of kidney disease in which an abundance of the protein C3 is present in the glomerular capillaries
- 1 in 1,000,000 people have C3G
- Sub-types of C3G
  - Dense Deposit Disease (DDD)
  - C3 Glomerulonephritis (C3GN)

# Purpose

- About half of patients with C3G have end-stage renal disease (ESRD) within 10 years
- Improve predictions of future renal failure using various biomarkers
- GFR vs. eGFR
- Capping eGFR measurements




http://www.renalcareconsult.com/renal-disease-education/stages-of-kidney-disease/

# The Data - UI C3G Natural History Study

## Observations

- Initially 1935 observations from 208 individuals
- Predictions made from 467 observations from 109 individuals

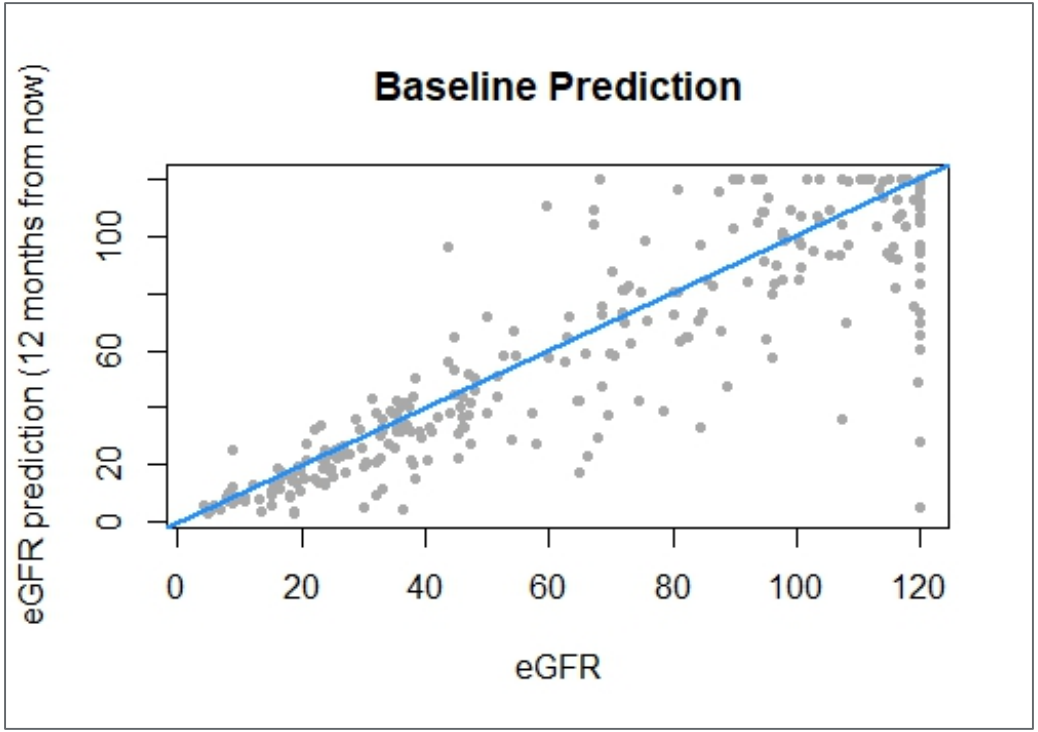
## Measurements

- eGFR
- C3
- UPCR
- UACR
- Soluble C5b-9 level
- Chronic Kidney Disease stage
- Sex
- Etiology
- Years with disease
- Age at Biopsy
- Race/Ethnicity
- Transplant Status

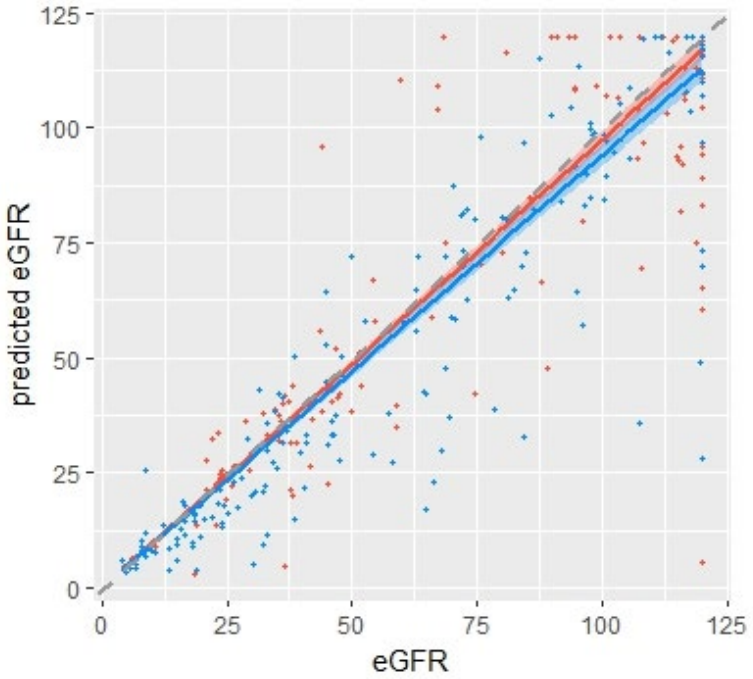
# Three Models

1. Baseline
2. Years with disease
3. Incorporating biomarkers

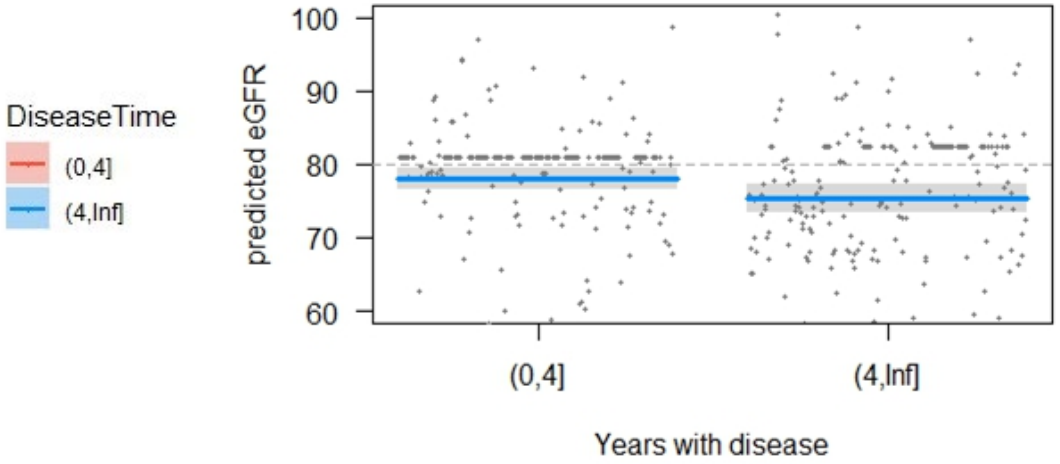
# 1. Baseline Prediction



# 2. Years with Disease




<4 years: 3% decline
>4 years: 6% decline

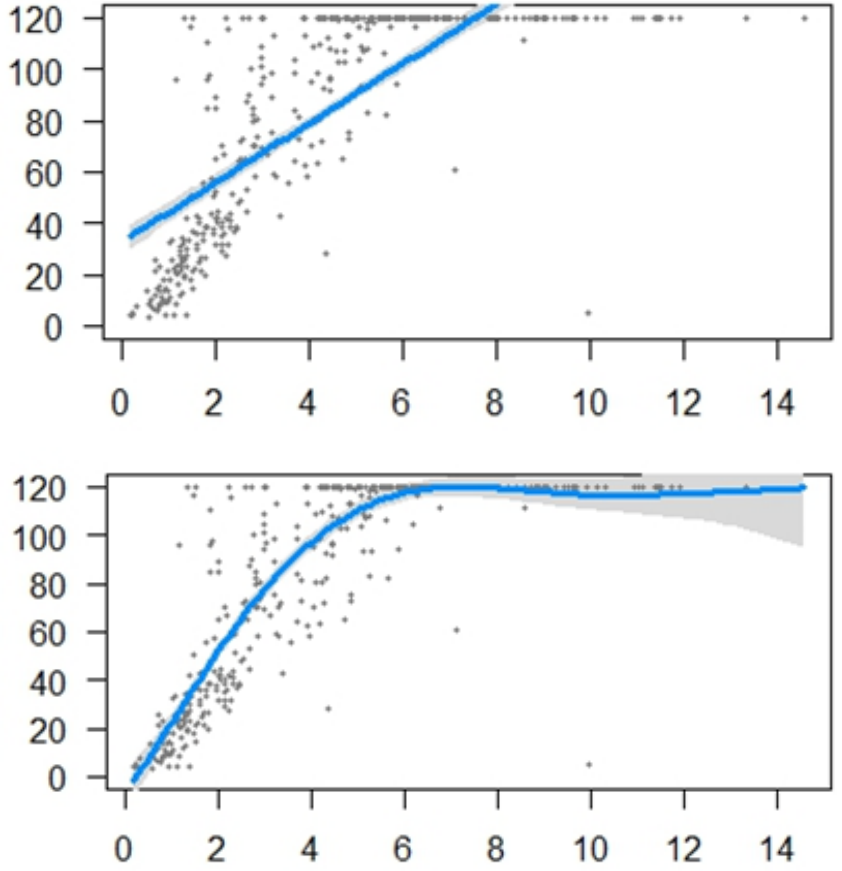


# 3. Biomarkers

- eGFR
- Years with disease
- Urine Protein-Creatinine Ratio (UPCR)
- Soluble C5b-9 level

# Methods

- Linear Models
- Generalised Additive Models (GAMs)
  - Goal: extend linear-regression models to be more flexible while avoiding overfitting the data



# Generalised Additive Models (GAMs)

- *GAM:* a sum of smooth functions

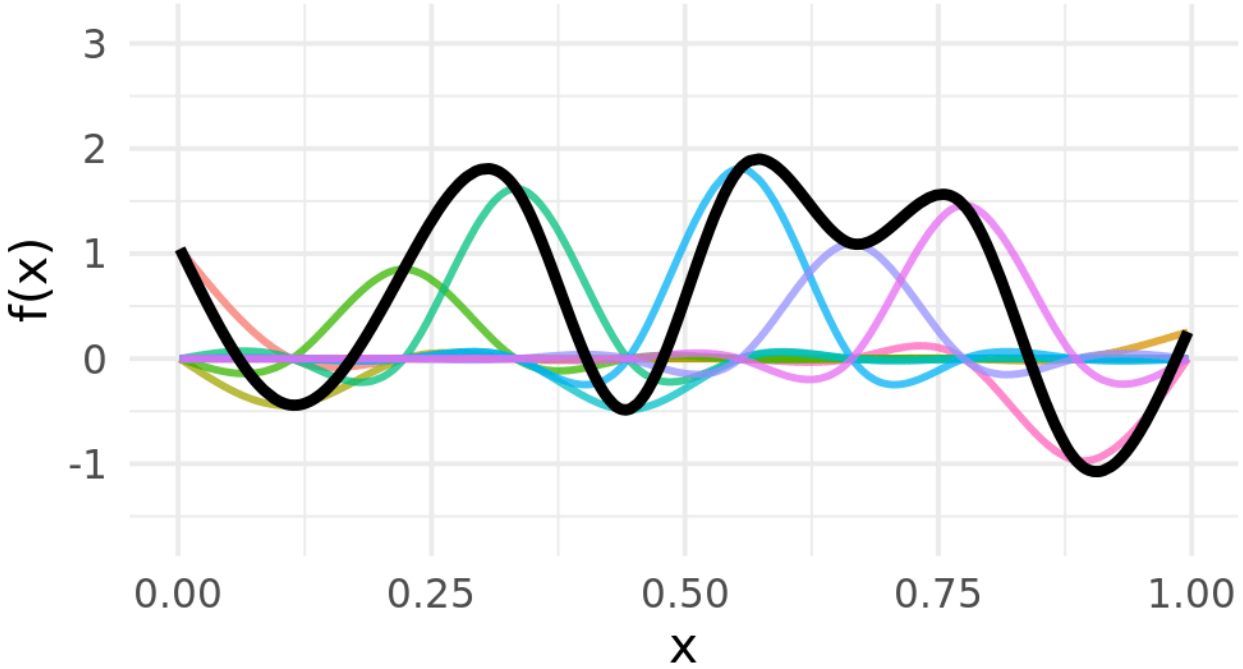
$$y_i = \beta_0 + \sum_j s_j(x_{ji}) + \epsilon_i$$

- *Splines*: functions made of simpler functions, $b_k$, each with weight $\beta_k$

$$s(x) = \sum_{k=1}^{K} \beta_k b_k(x)$$

Equations from: Simpson, Gavin, 'Generalised Additive Models (GAMs)' for the 'Statistical Methods' webinar series hosted by the Ecological Forecasting Initiative and the ESA Statistical Ecology Section. Given on January 3, 2022. bit.ly/gam-efi-22.

# GAMs



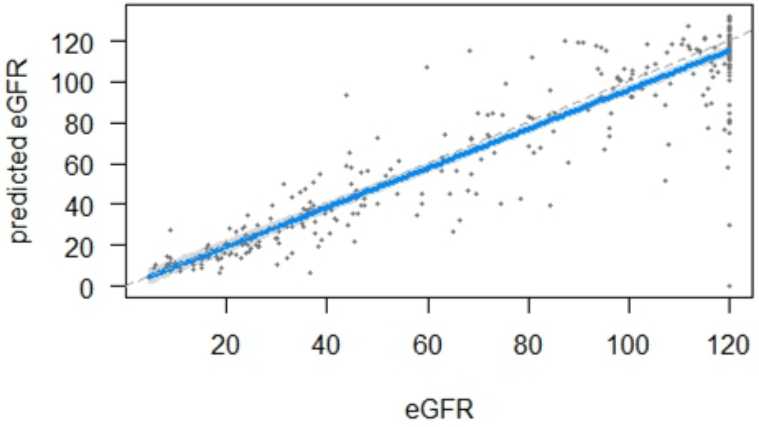

Animation from: Simpson, Gavin, 'Generalised Additive Models (GAMs)' for the 'Statistical Methods' webinar series hosted by the Ecological Forecasting Initiative and the ESA Statistical Ecology Section. Given on January 3, 2022. bit.ly/gam-efi-22.

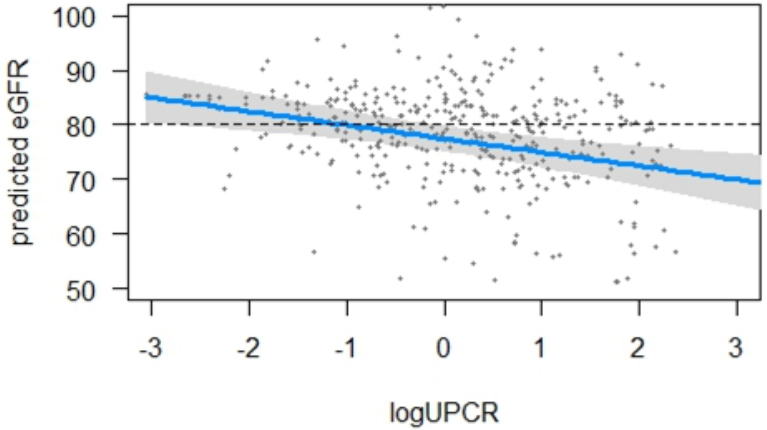
# Final Model

```
fit <- gam(eGFR2_cap ~ eGFR2_12m6w_cap +
             logUPCR_12m6w +
             s(soluble_level, sp=0.07) +
             s(YrsDisease),
           data=c3g)
```

eGFR
coefficient = 0.95971

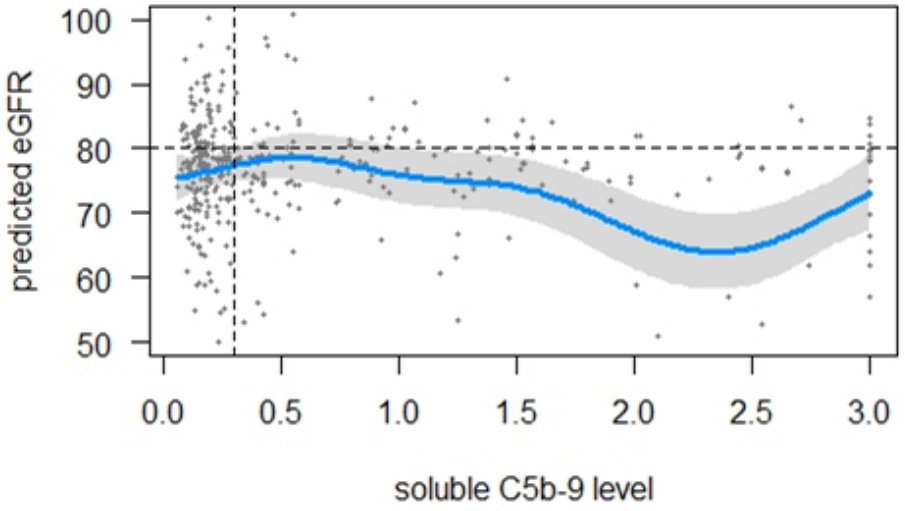
log(UPCR)
coefficient = -2.486



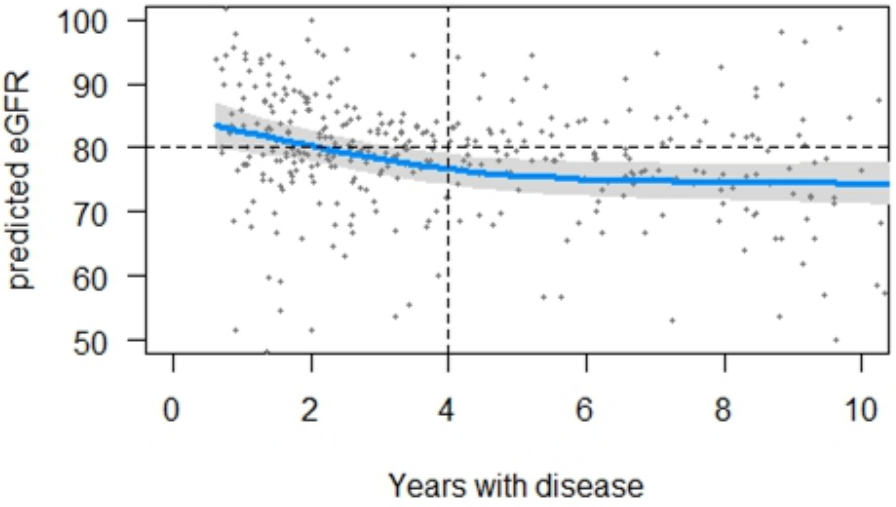



# Final Model

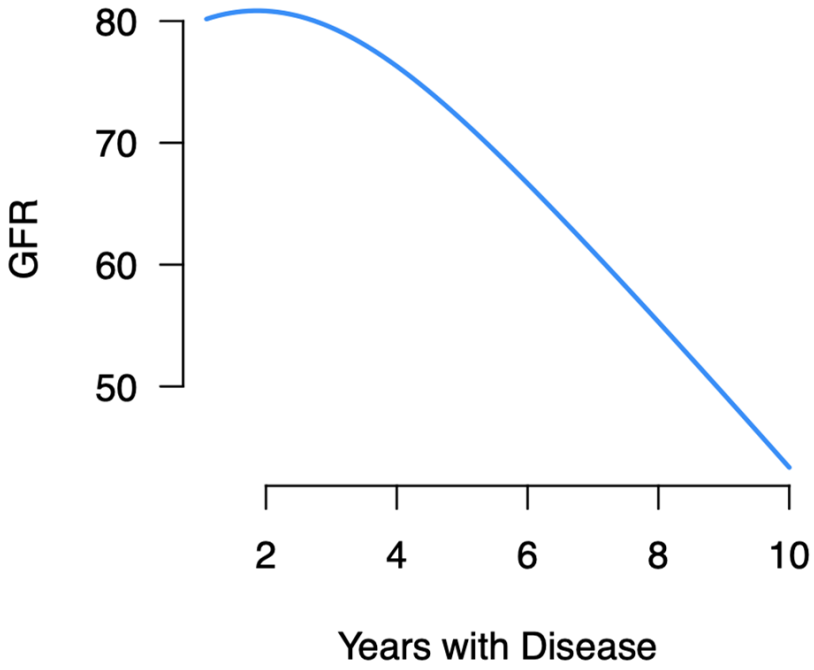
soluble level




years with disease



# Final Model - Patient Trajectory



# Model Accuracy

- Adjusted R-squared:
  - Adjusted R-squared for baseline prediction = 0.879
  - Adjusted R-squared for years prediction = 0.879
  - Adjusted R-squared for our model = 0.89
- Cross Validation:
  - Generalised Cross Validation
    - GCV for baseline prediction = 230.08
    - GCV for years prediction = 230
    - GCV for our model = 212.01
  - Training and testing splits

# Limitations

- Some observations come from the same individual multiple times
- Not taking into account dependence
- Have missing data

# Acknowledgements


- Patrick Breheny, PhD
- Logan Harris, MS
- Richard J H Smith, MD
- Carla M Nester, MD, MSA, FASN
- ISIB Program sponsored by the National Heart Lung and Blood Institute (NHLBI), grant # HL-147231