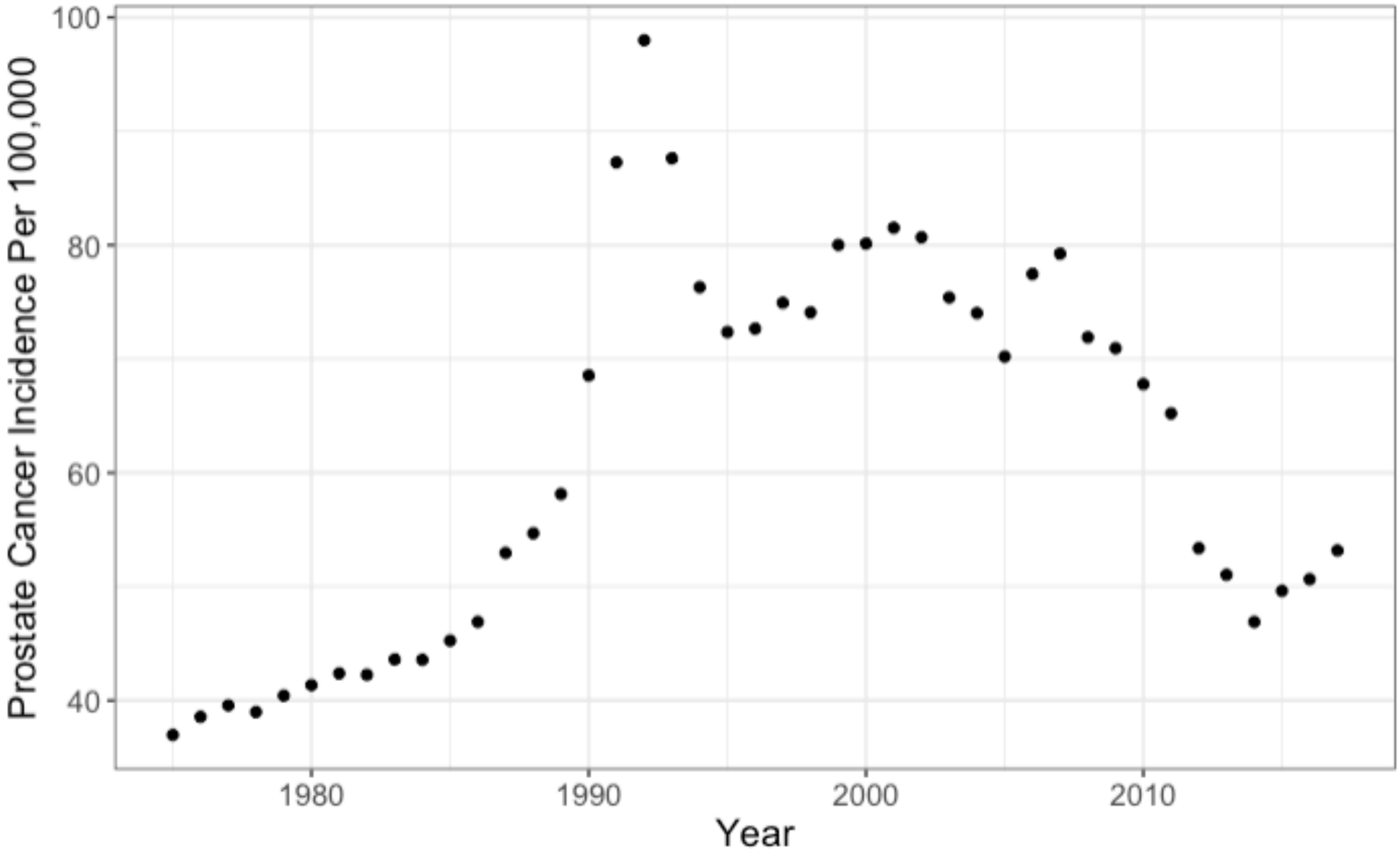
Modeling data using horseshoe processes

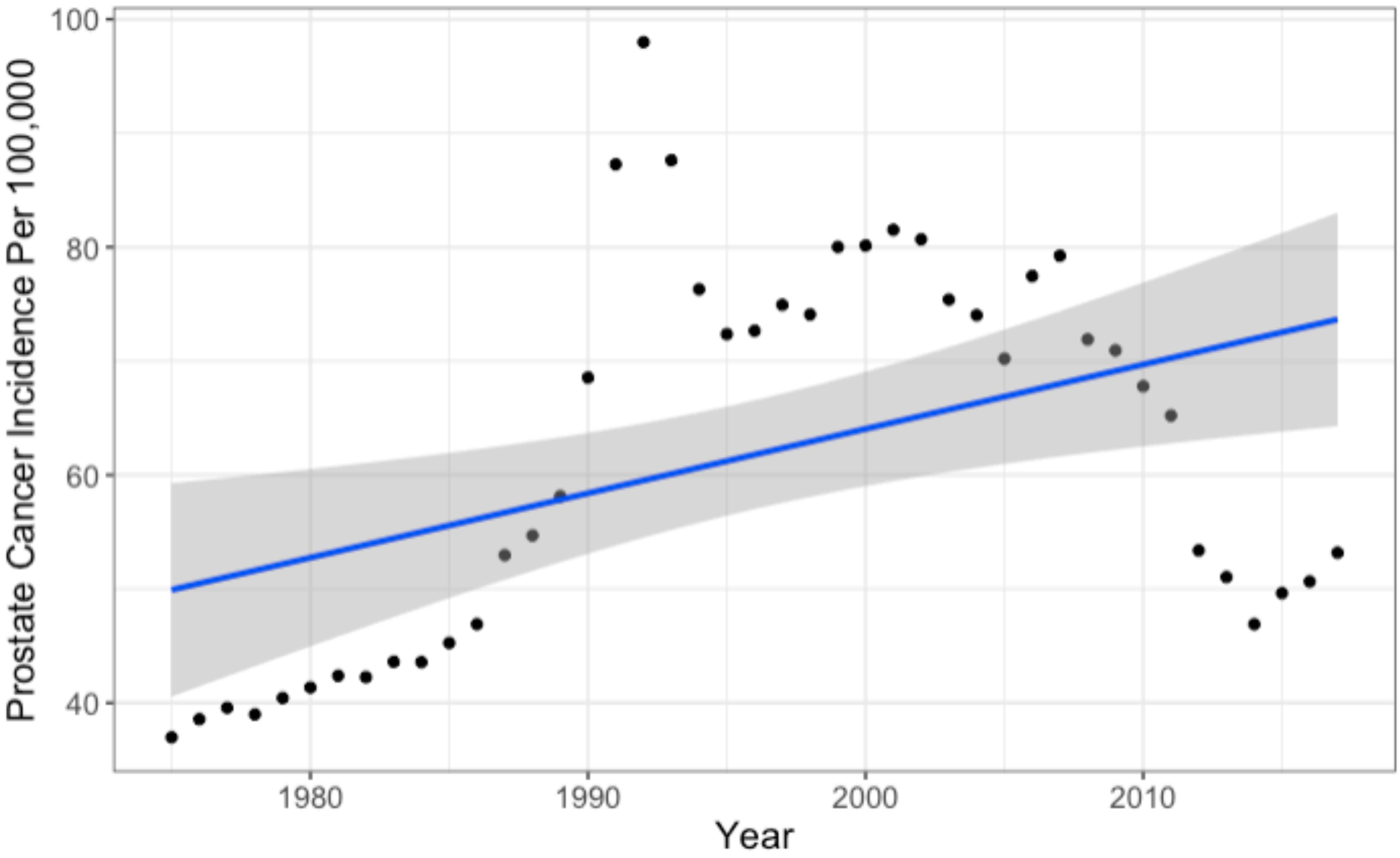
JSM 2021 Contributed Speed Talk: Advanced Bayesian Topics

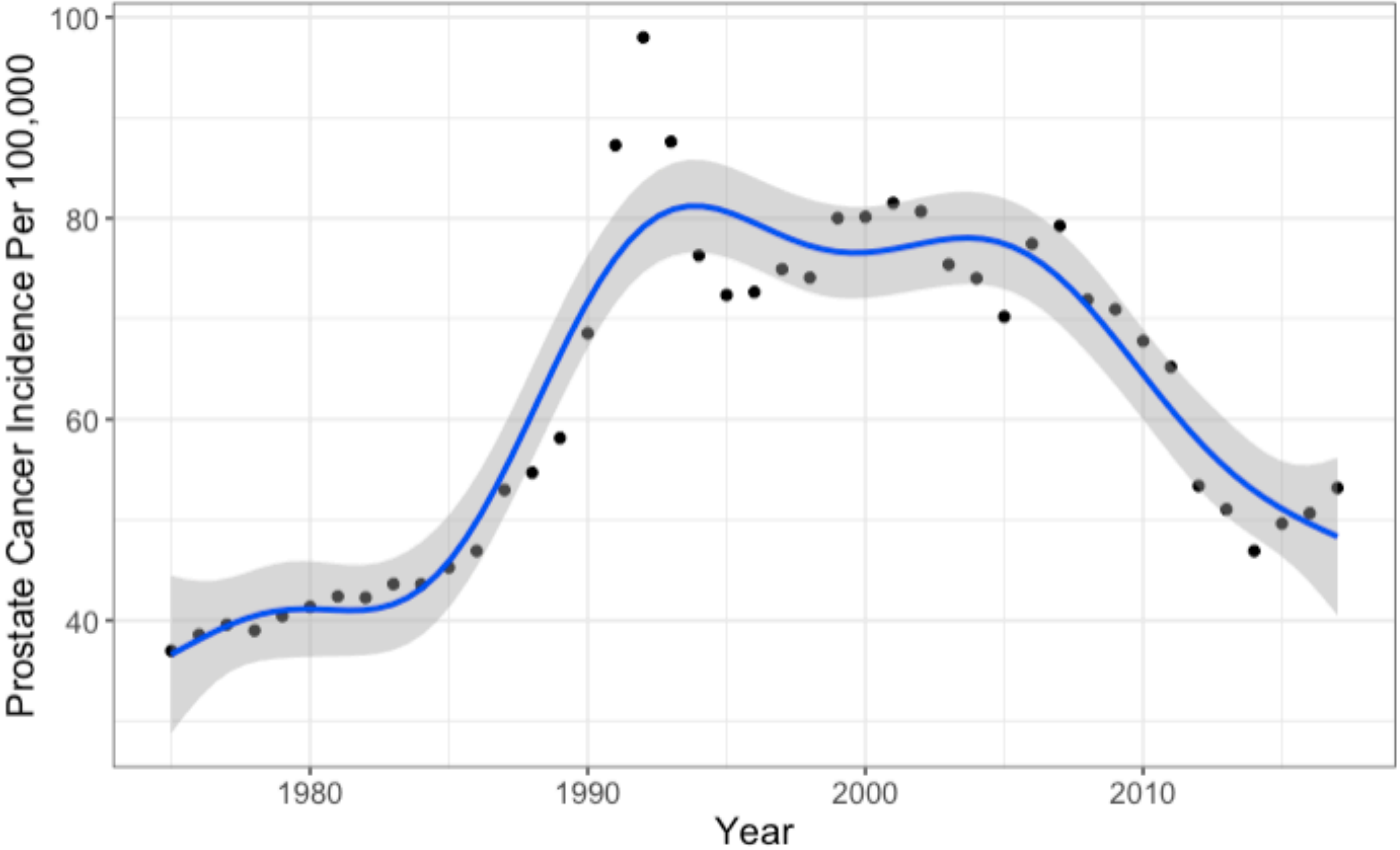
Elizabeth Chase, University of Michigan Biostatistics

Aug. 8, 2021

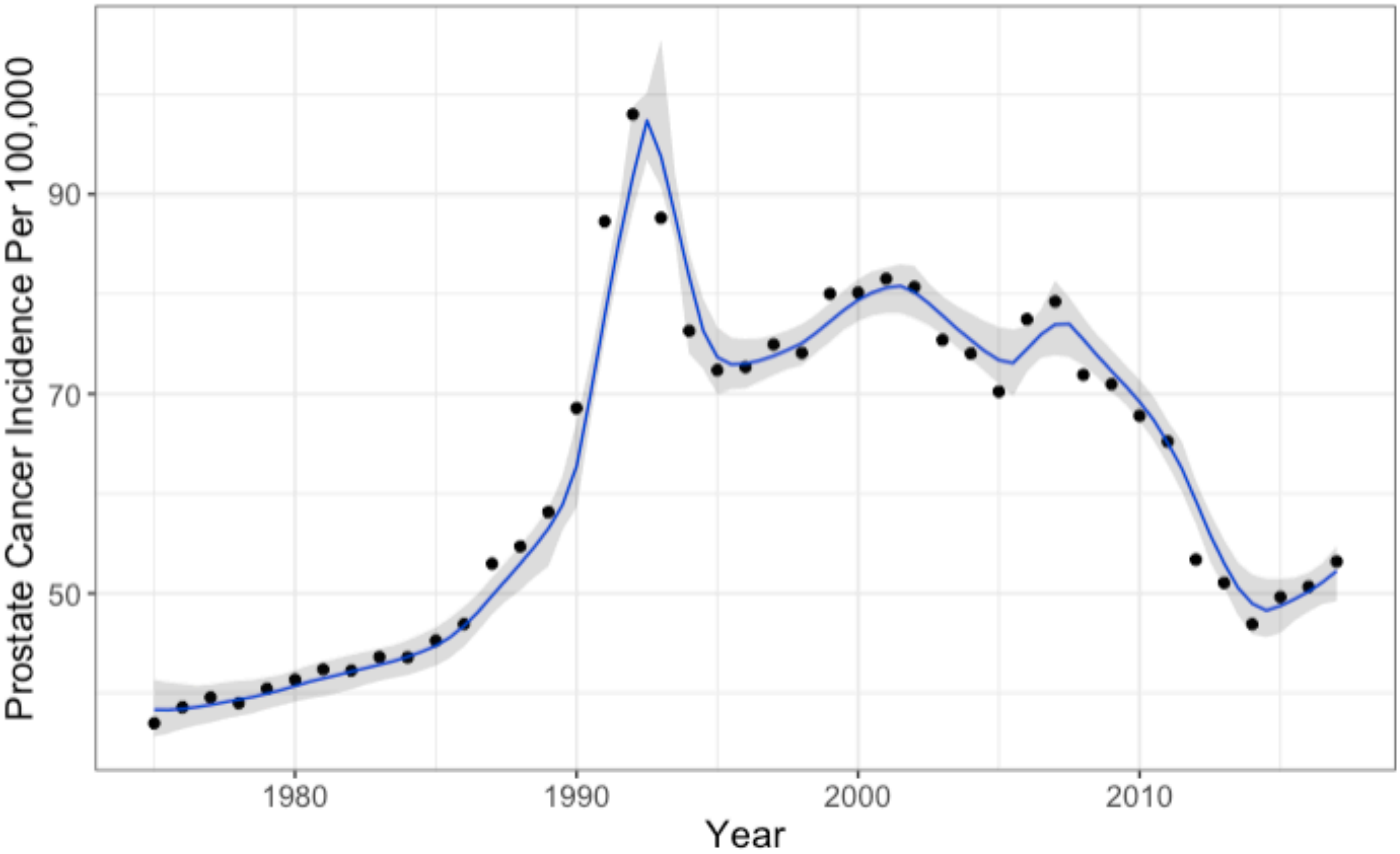


100
80
60
40
Prostate Cancer Incidence Per 100,000
1980
1990
2000
2010
Year

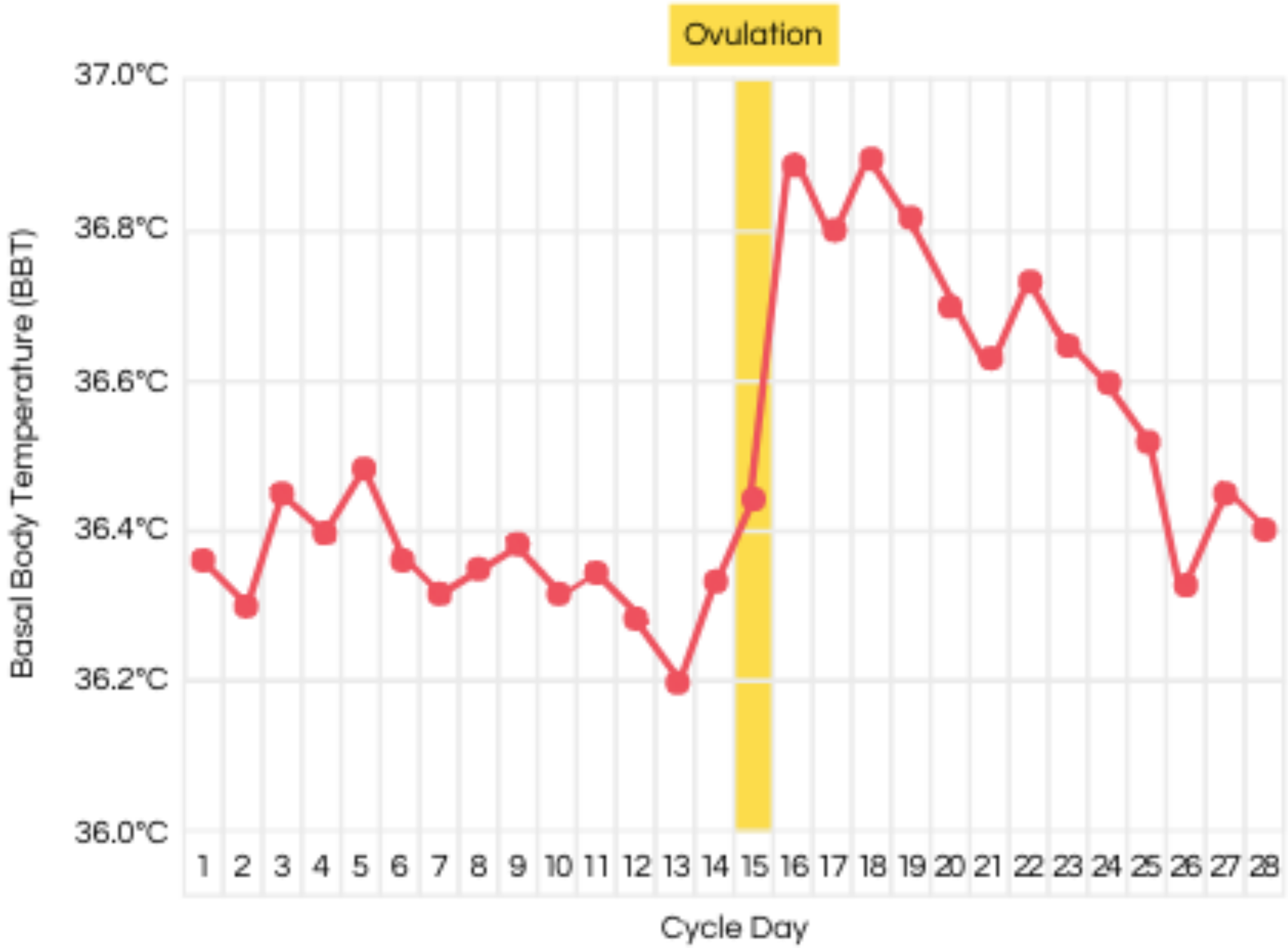

Prostate Cancer Incidence Per 100,000
100
80
60
40
1980
1990
2000
2010
Year

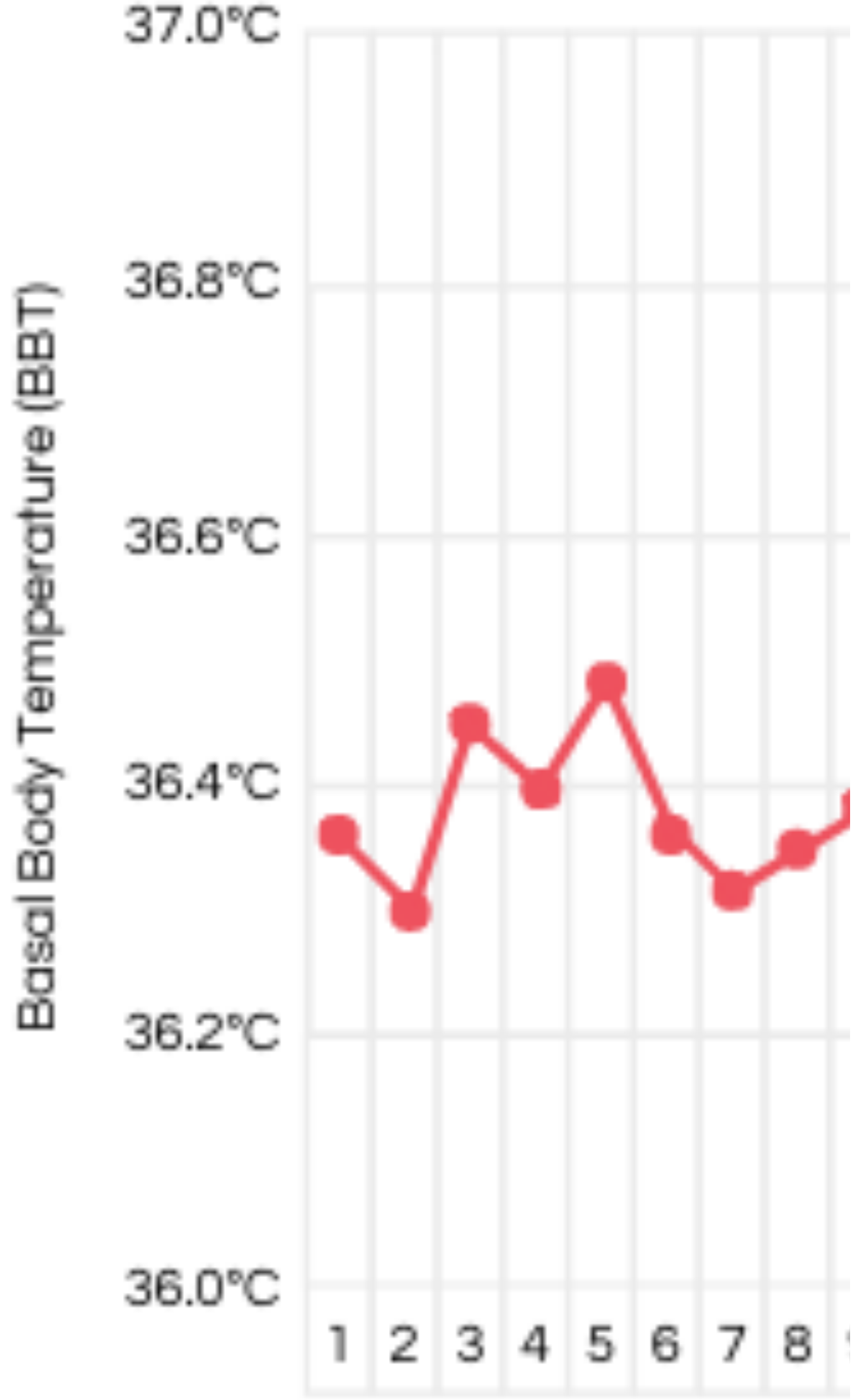


Prostate Cancer Incidence Per 100,000
100
80
60
40
1980
1990
2000
2010
Year

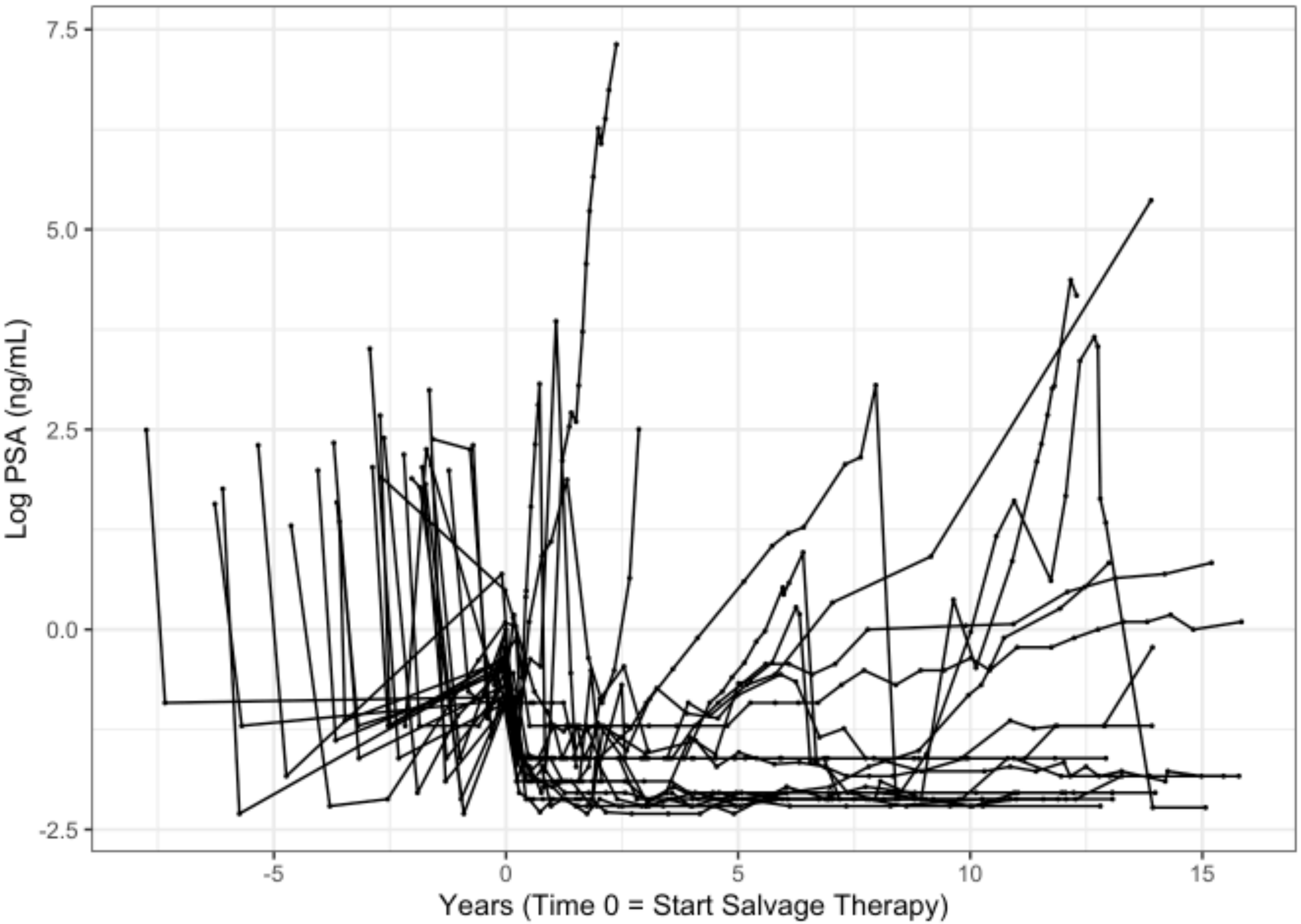


Prostate Cancer Incidence Per 100,000
90
70
50
1980
1990
2000
2010
Year


Ovulation
37.0°C
36.8°C
36.6°C
36.4°C
36.2°C
36.0°C
Basal Body Temperature (BBT)
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
Cycle Day


Basal Body Temperature (BBT)
37.0°C
36.8°C
36.6°C
36.4°C
36.2°C
36.0°C
1 2 3 4 5 6 7 8

Ovulation


Log PSA (ng/mL)
7.5
5.0
2.5
0.0
-2.5
-5
0
5
10
15
Years (Time 0 = Start Salvage Therapy)

Motivation

- Desired end-product: a flexible curve-fitting approach of a (possibly non-Gaussian) outcome on a continuous predictor that allows for *local changes in variance*

- Additional desirable features:

    - Equipped to work with small and large numbers of observations

    - Both equally and unequally spaced grids of the predictor

    - Possibly multiple measurements at the same value of the predictor

    - Provides uncertainty quantification

Elizabeth Chase, University of Michigan Biostatistics
ecchase@umich.edu

The Horseshoe Prior

- The horseshoe prior was developed as a Bayesian approach for variable shrinkage.
- Suppose we are considering the model for subjects i = 1, …, n and predictors j = 1, …, p:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

- If p is large, we might expect that most of our coefficients are zero, and just a couple effects are very large.

Elizabeth Chase, University of Michigan Biostatistics
ecchase@umich.edu

The Horseshoe Prior

- In 2010, Carvalho et al. proposed a continuous mixture prior, which they called the horseshoe:

$$\beta_j \sim N(0, \tau^2 \omega_j^2)$$

$$\omega_j \sim C^+(0,1),\ \tau \sim C^+(0,c)$$

- τ^2 is the global shrinkage parameter, and controls the overall level of shrinkage for all the effects.

- ω_j^2 is the local shrinkage parameter, and allows for a mixture of very large signals and zero signals.

A Horseshoe Process

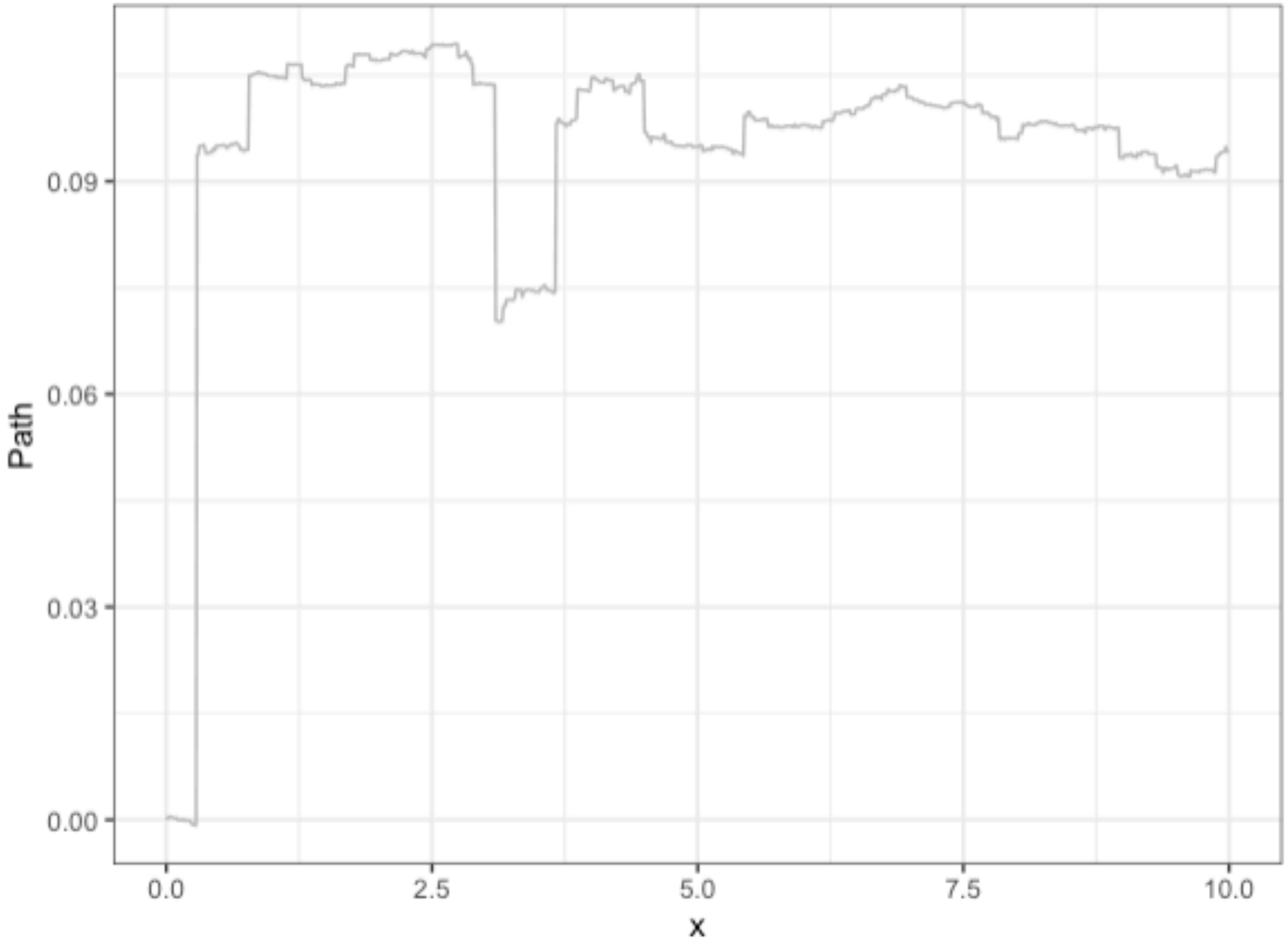
- We can extend the horseshoe distribution to the stochastic process setting, following Polson & Scott (2012).
- We define a horseshoe process $H(x)$ observed at discrete locations $x_0, x_1, \ldots, x_m$ as:

$$H(x_j) - H(x_{j-1}) \sim N(0, \tau^2 \omega_j^2 (x_j - x_{j-1}))$$

$$\omega_j \sim C^+(0,1),\ j = 1,\ldots,m$$

$$\tau \sim C^+(0,\ c),\ H(x_0) = 0$$

A Horseshoe Process



Elizabeth Chase, University of Michigan Biostatistics
ecchase@umich.edu

Second Derivative Shrinkage

Let y_{t_k} indicate some outcome (i.e. biomarker reading, disease incidence, etc.) at time $t_k, k = 0, \ldots, m$ with $t_0 = 0$.

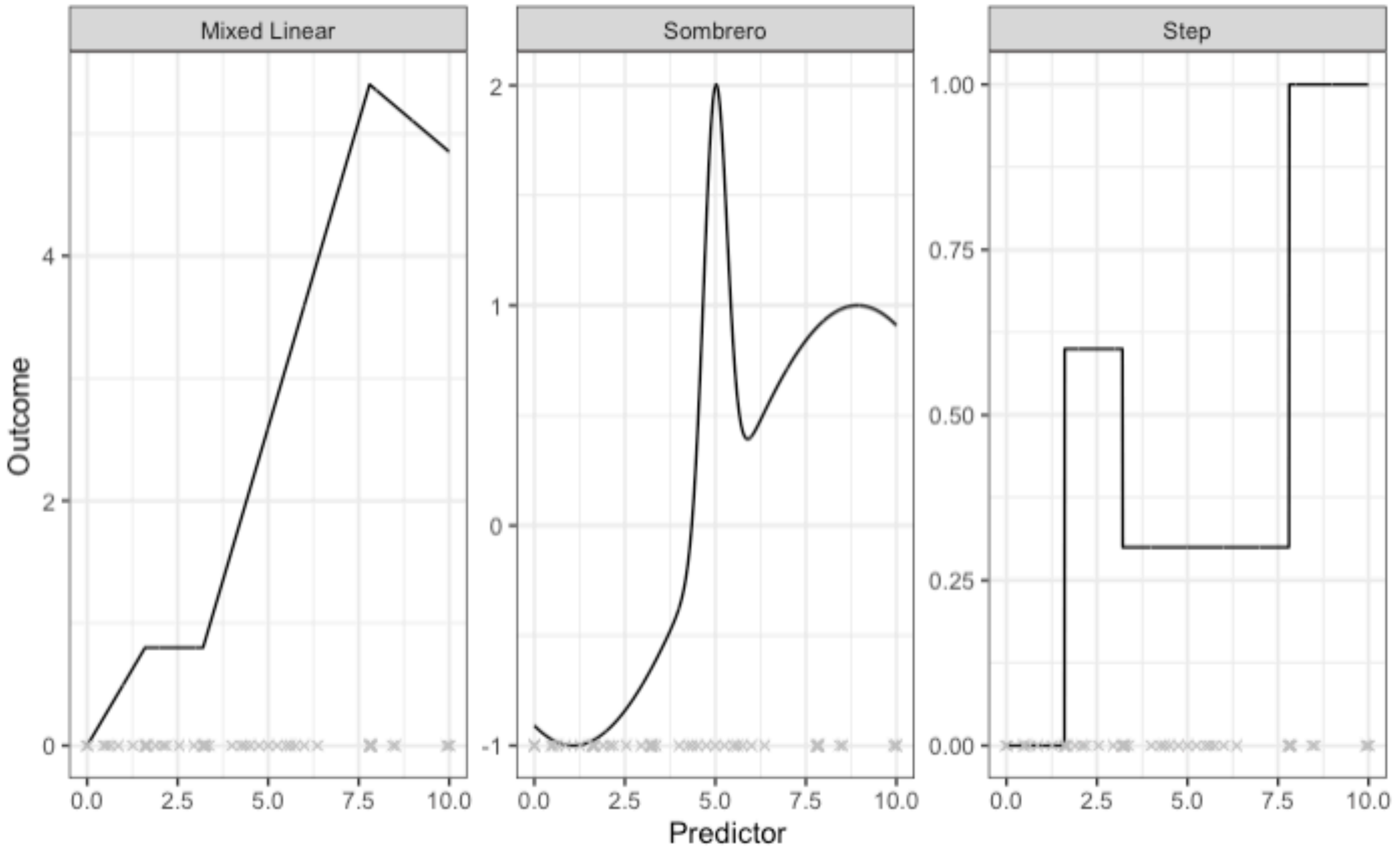
$$g(E(y_t)) = \alpha + \int_0^t H(u)du$$

- $\alpha \sim N(a, b^2)$
- $H(u) - H(u - s) \sim N(0, c^2\tau^2\omega(u)^2 s), \quad H(t_0) = 0$
- $\omega(u) \sim C^+(0,1), \quad u = t_1, \ldots, t_m, \quad \omega(t_0) = 0$
- $\tau \sim C^+(0,1)$

Computation

- The model is currently implemented using Hamiltonian Monte Carlo in Stan (via cmdstanr in R) using decentered parameterizations.
- This yields faster computational performance than other Bayesian methods that seek to address this problem.
- However, computation time is still a constraint, and issues with divergences still appear occasionally for select datasets.

Simulation Study



Elizabeth Chase, University of Michigan Biostatistics
ecchase@umich.edu

Simulation Study

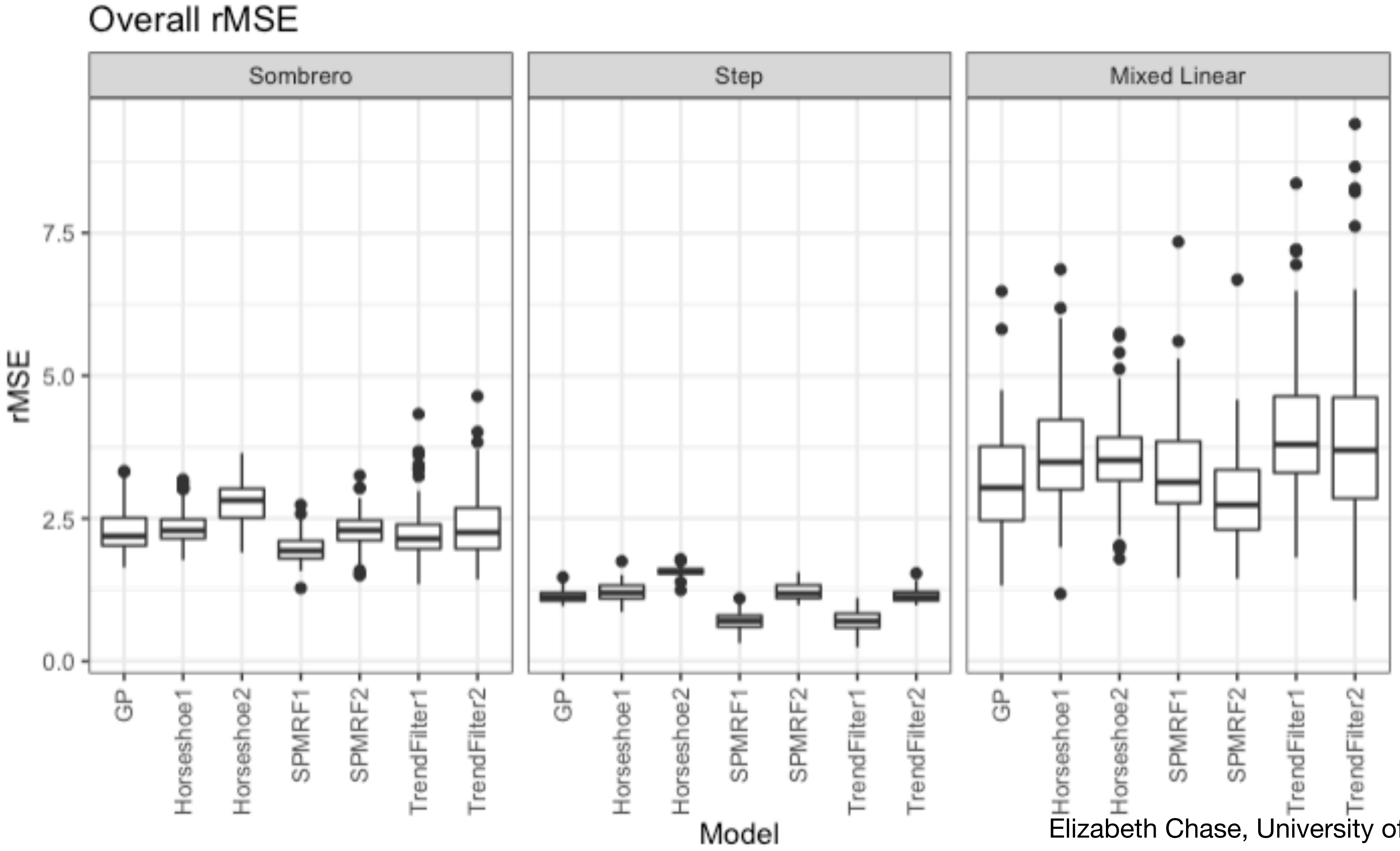
- Considered 3 data-generating scenarios, 2 sample sizes (n = 50, n = 100), and 3 types of outcome data (Gaussian, Binomial, Poisson)

- Ran 100 replicates of each scenario

- Comparison methods: Gaussian process regression, trend filtering model (Tibshirani 2014), shrinkage process Markov random fields (Faulkner & Minin 2018)

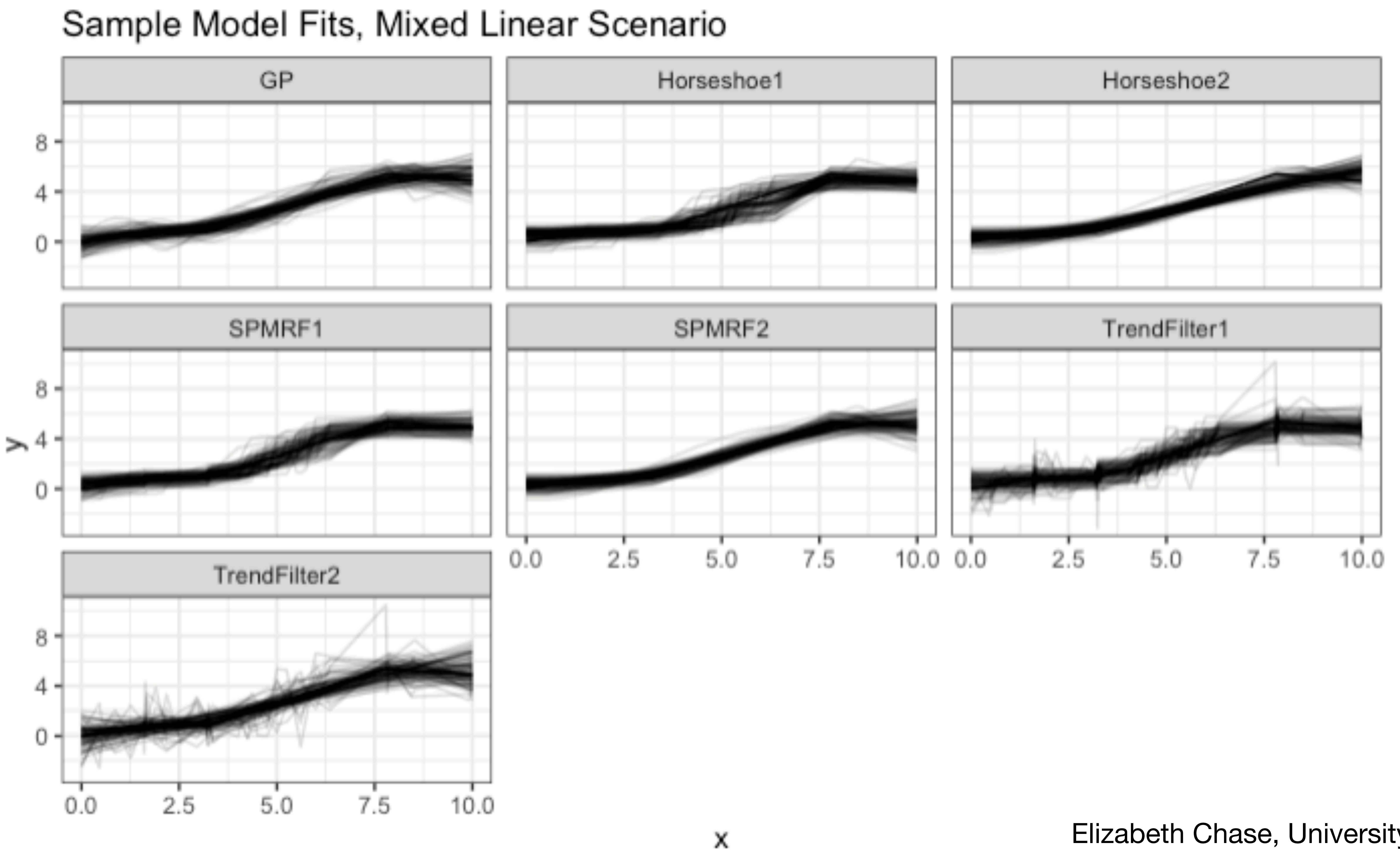
Elizabeth Chase, University of Michigan Biostatistics
ecchase@umich.edu

Simulation Study



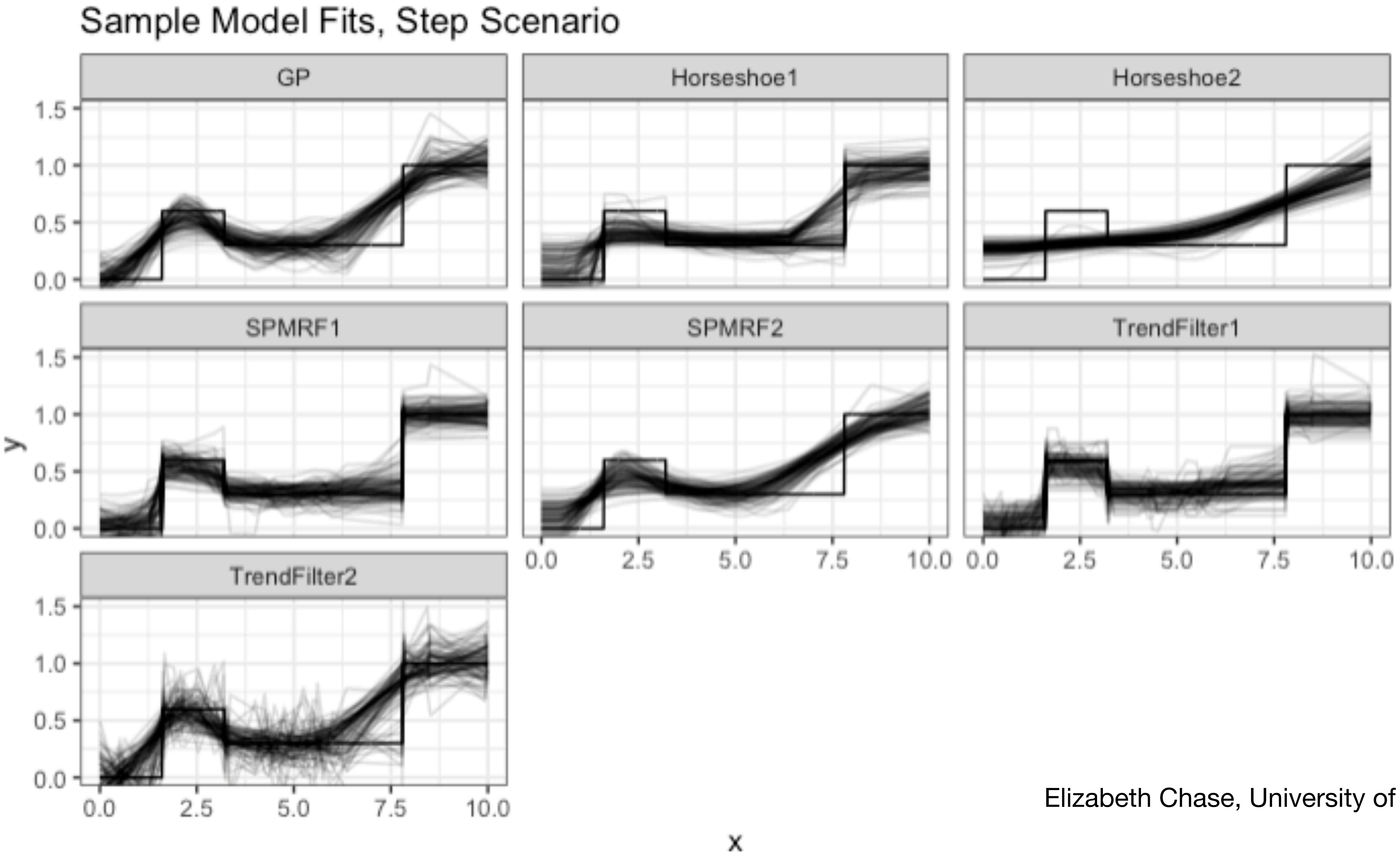
Elizabeth Chase, University of Michigan Biostatistics
ecchase@umich.edu

Simulation Study



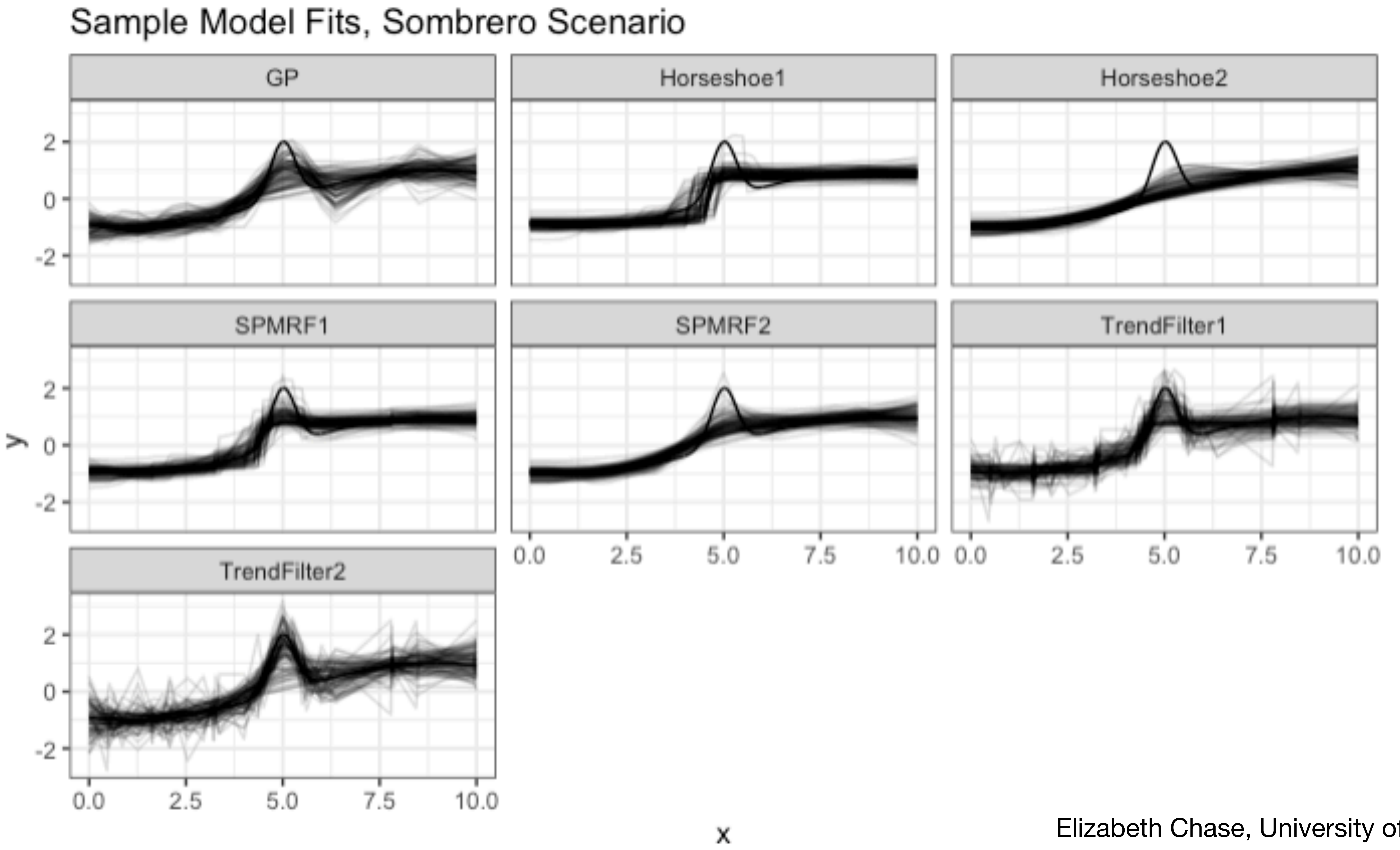
Elizabeth Chase, University of Michigan Biostatistics
ecchase@umich.edu

Simulation Study



Elizabeth Chase, University of Michigan Biostatistics
ecchase@umich.edu

Simulation Study



Future Work

- Continue workshopping method—how to set initial values, recommendations for setting hyperpriors
- Perhaps a variational Bayes implementation to further accelerate computation
- Consider use of regularized horseshoe or horseshoe + priors

References

- M Betancourt. “A Conceptual Introduction to Hamiltonian Monte Carlo.” *Arxiv* (July 2018).
- CM Carvalho, NG Poisson, JG Scott. “The horseshoe estimator for sparse signals.” *Biometrika* 97.2 (June 2010): 465-80.
- JR Faulkner and VN Minin. Locally adaptive smoothing with Markov random fields and shrinkage priors. *Bayesian Analysis* 13.1 (2018): 225–252.
- J Piironen and A Vehtari. “Sparsity information and regularization in the horseshoe and other shrinkage priors.” *Electronic Journal of Statistics* 11 (2017): 5018-51.
- NG Polson and JG Scott. “Local shrinkage rules, Lévy processes and regularized regression.” *JRSS-B* 74.2 (March 2012): 287-311.
- RJ Tibshirani. “Adaptive piecewise polynomial estimation via trend filtering.” *Annals of Statistics* 42.1 (Feb. 2014): 285-323.

Elizabeth Chase, University of Michigan Biostatistics
ecchase@umich.edu