Flexible Bayesian Models for Prediction and Causal Estimation with Return Time Outcomes

Applications in HIV Care Retention

Arman Oganisian[1]

Allison DeLong[1], Ann Mwangi[2], Edwin Sang[2], Hamish Fraser[3], Joseph Hogan[1]

[1]Department of Biostatistics, Brown University
[2]Moi University
[3]Center for Biomedical Informatics, Brown University

BROWN
School of Public Health

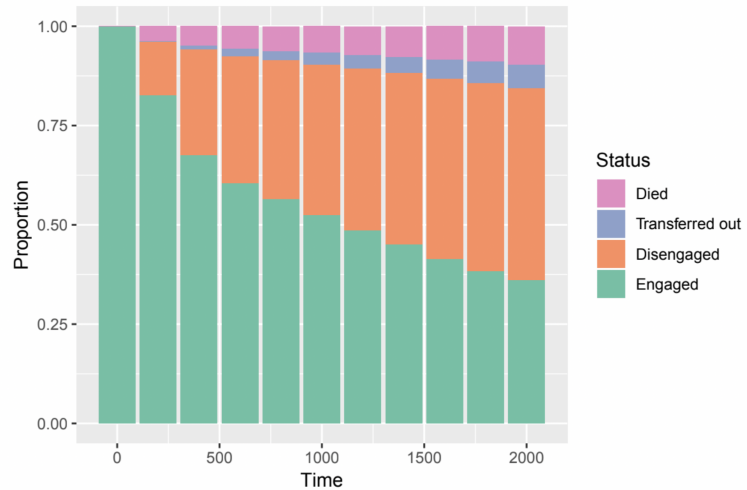
Study Motivation - Closing the Retention Gap

- Retention in care is a crucial component of Joint United Nations Programme on HIV/AIDS (UNAIDS) 90-90-90 plan goals.
- Ideally: patients repeatedly attend follow-up clinical appointments on time.
- Reality: clinic visits may be difficult to make for certain patients.
- Key decision: scheduling of follow-up appointments.
- Scheduling too frequently or infrequently can lead to retention loss.

AMPATH Care Program in Western Kenya

The Academic Model Providing Access to Healthcare (AMPATH) care program treats roughly 150,000 patients with HIV at over 60 urban and rural clinics in western Kenya.

- At AMPATH, single-visit retention rates at many clinics fluctuate below 90%.
- Preliminary analyses show that long-term retention drops off considerably after initial enrollment.



AMRS Data Structure and Notation

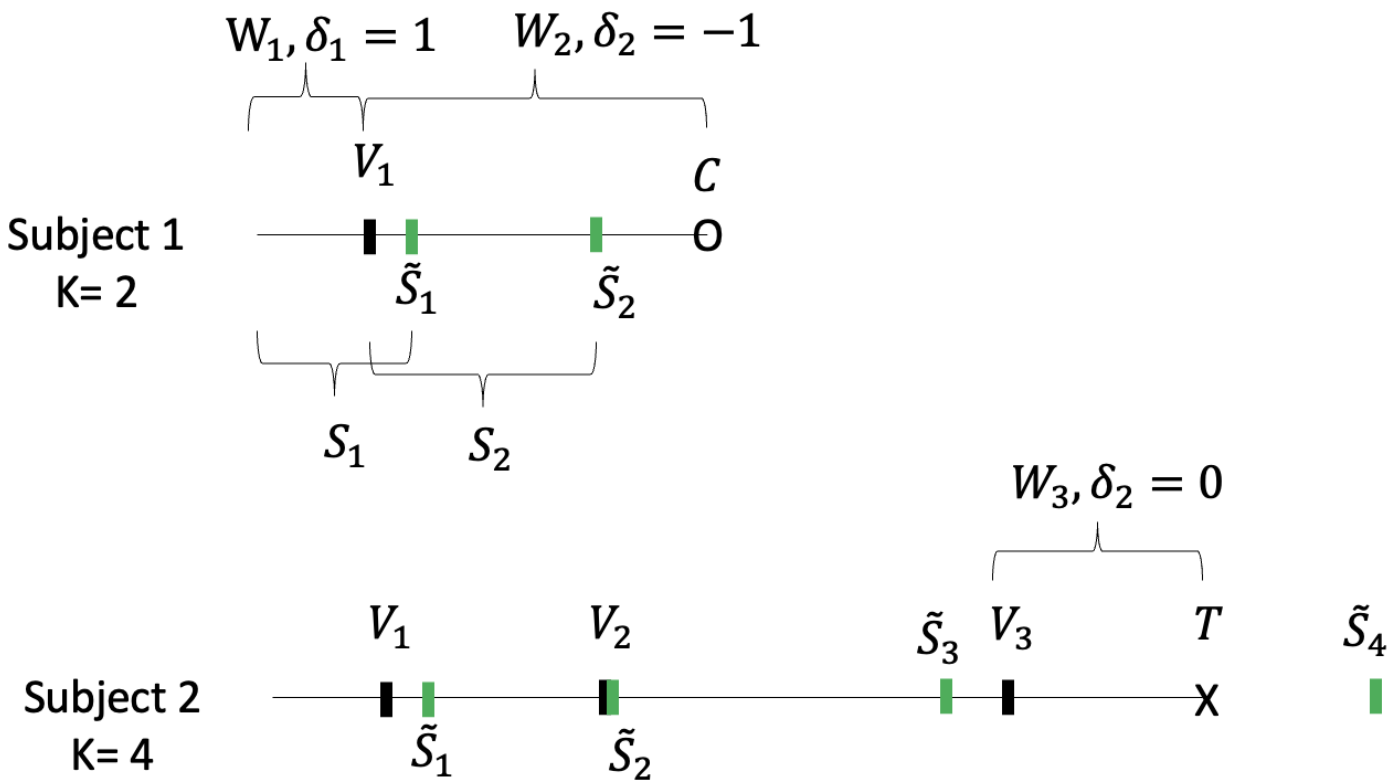
AMPATH Medical Record System (AMRS) - longitudinal data on scheduling times, event times, and patient features.

- For each patient, we observe data on $k = 1, 2, \ldots, K$ events.
- Visit $k = 1, 2, \ldots, K-1$ scheduled for time $\tilde{S}_k$ and occurred at time V_k.
- At visit $k = 1, 2, \ldots, K-1$ we record a set of features L_k.
- Define relative waiting times and scheduled return times: $S_k = \tilde{S}_k - V_{k-1}$ and $W_k = \min(V_k, T, C) - V_{k-1}$.
- Event indicator $\delta_k = \in \{-1, 0, 1\}$ for censoring, death, return visit - respectively.
- Δ-retention indicator:

$$Y_k(\Delta) = I(W_k - S_k < \Delta, \delta_k = 1)$$

- Available history $H_k = (\bar{S}_{k-1}, \bar{V}_{k-1}, \bar{L}_k, \bar{\delta}_{k-1} = 1)$.

AMRS Data Structure and Notation



Δ-Retention Definitions

- Kenyan Ministry of Health (MOH) HIV treatment guidelines describes a patient as lost to followup (LTFU) "a client who has not turned up or come back to the clinic for either a clinical visit or refills for more than 90 days from the last scheduled visit".
- Kenyan MOH guidelines also define a "defaulter" as "a client who has not turned up for either a clinical visit or refills 7 days after their scheduled appointment date."

Potential Outcomes and Estimands

Define potential waiting time and event type of event k under hypothetical scheduling decision $s_k \in \mathcal{S}_k$

$$W_k^{s_k} = \min(V_k^{s_k}, T_k^{s_k}) \qquad \delta_k^{s_k} = I(V_k^{s_k} < T_k^{s_k})$$

Potential retention status: $Y_k^{s_k}(\Delta) = I(W_k^{s_k} - s_k < \Delta, \delta_k^{s_k} = 1)$

Relevant Estimands. For some $s_k, s_k' \in \mathcal{S}_k$.

- Conditional effects:

$$\Psi_k(\Delta; h_k) = P(Y_k^{s_k}(\Delta) = 1 \mid H_k = h_k) - P(Y_k^{s_k'}(\Delta) = 1 \mid H_k = h_k)$$

- Marginal effects:

$$\Psi_k(\Delta) = \int_{\mathcal{H}_k} P(Y_k^{s_k}(\Delta) = 1 \mid h_k) - P(Y_k^{s_k'}(\Delta) = 1 \mid h_k) dF_k(h_k)$$

- Optimization:

$$s_k^*(\Delta) = \underset{s_k \in \mathcal{S}_k}{\operatorname{argmax}} P(Y_k^{s_k}(\Delta) = 1 \mid H_k = h_k)$$

- Prediction: for some new $H_k = h_k$,

$$P(Y_k(\Delta) = 1 \mid S_k = s_k, H_k = h_k)$$

Estimation is challenging

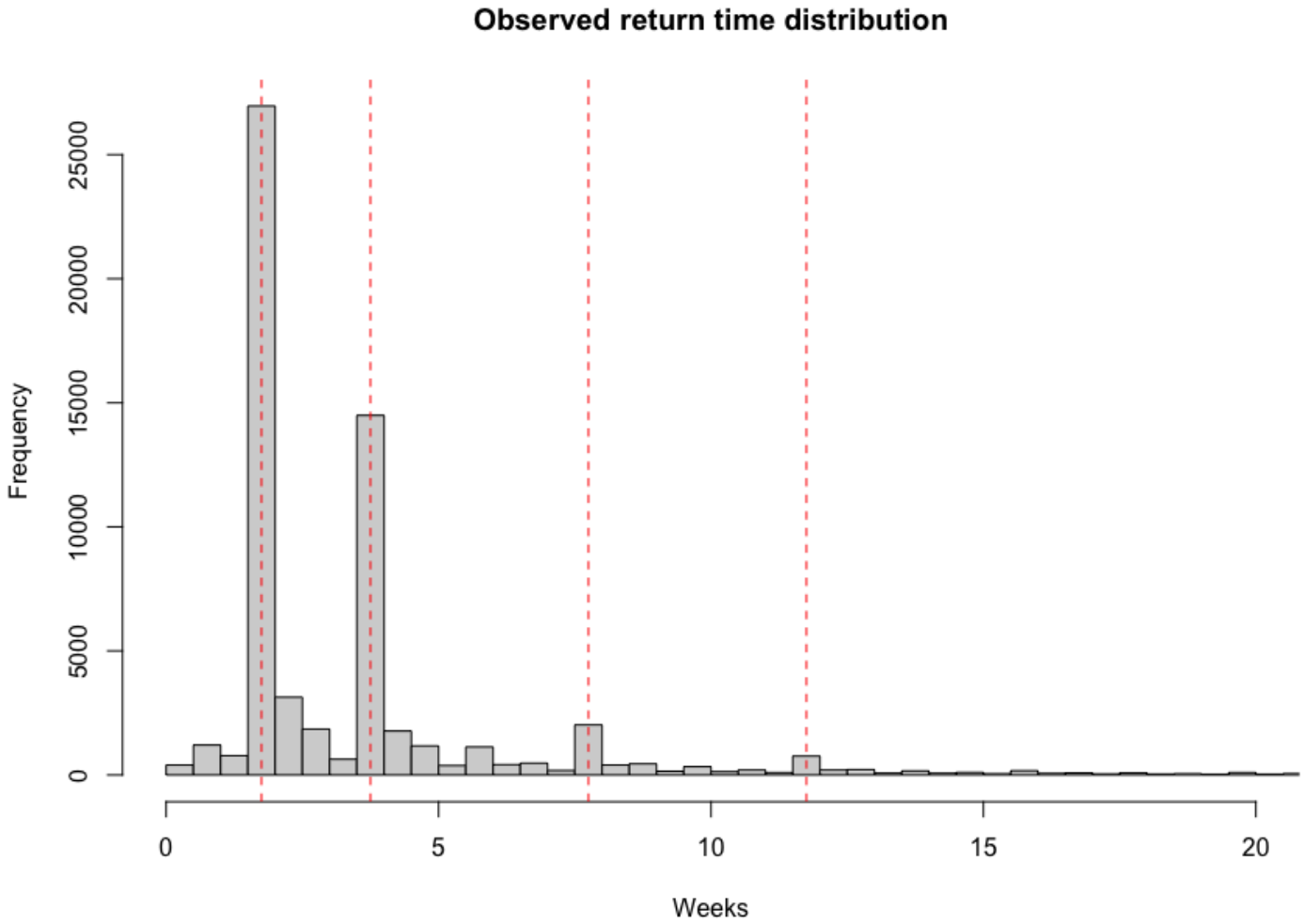
$$f^*(w_k^{s_k}, \delta_k^{s_k} = j \mid H_k = h_k)$$

for $j \in \{0, 1\}$. Estimands are functionals of the joint distribution of potential outcomes

Challenges:

- Counfounding: Followup frequency is like driven by factors also related to retention (e.g. travel time from clinic, CD4 count, etc.)
- Complex return time distributions: clumping around the scheduling visit time, censoring, visit process terminated by death.

Estimation is challenging



Identification Assumptions

Recall: $H_k = (\bar{S}_k, \bar{V}_{k-1}, \bar{\delta}_{k-1} = 0)$

1. Conditionally ignorable scheduling

$$W_k^{s_k}, \delta_k^{s_k} \perp S_k \mid \bar{S}_{k-1}, \bar{V}_{k-1}, \bar{L}_k, \bar{\delta}_{k-1} = 1$$

2. Positivity for each $s_k \in \mathcal{S}_k$

$$P(S_k = s_k \mid \bar{S}_{k-1}, \bar{V}_{k-1}, , \bar{L}_k, \bar{\delta}_{k-1} = 1) > 0$$

3. Non-informative censoring

$$\lim_{dw \to 0} \frac{P(w_k \leq W_k < w_k + dw, \delta_k = -1 \mid W_k > w_k, H_k = h_k, V_k^{s_k}, T_k^{s_k})}{dw}$$
$$= \lim_{dw \to 0} \frac{P(w_k \leq W_k < w_k + dw, \delta_k = -1 \mid W_k > w_k, H = h_k)}{dw}$$

Other assumptions: SUTVA, censoring positivity.

Observed data models

Under previous assumptions, joint distribution of potential outcomes can be expressed in terms of observed data cause-specific hazards,

$$f^*(w_k^{s_k}, \delta_k^{s_k} = j \mid H_k = h_k) = f(w_k, \delta_k = j \mid S_k = s_k, H_k = h_k)$$
$$= \lambda_j(w_k \mid s_k, h_k) \exp\Big(- \int_0^{w_k} \sum_{j \in \{0,1\}} \lambda_j(u \mid s_k, h_k) du \Big)$$

Requires modeling cause-specific hazards $\lambda_j(w_k \mid s_k, h_k)$.

Identification Assumptions - Missing Features

In medical records, features may be inconsistently recorded across subjects so that some elements of $L_k = (L_{1k}, L_{2k}, \ldots, L_{pk})$ are missing.

- Define monitoring indicators: $M_k = (M_{1k}, M_{2k}, \ldots, M_{1k})$.
- Observed covariates from visit k: $L_k^{O,M_k} = L_k \cdot M_k$
- Observed history: $(\bar{L}_k^{O,\bar{M}_k}, \bar{M}_k)$.

Assumptions must be modified, e.g. modified ignorability

$$W_k^{s_k}, \delta_k^{s_k} \perp S_k \mid \bar{S}_{k-1}, \bar{V}_{k-1}, \bar{L}_k^{O,\bar{m}_k}, \bar{M}_k = \bar{m}_k, \bar{\delta}_{k-1} = 1$$

For full flexibility, should stratify waiting time models by S_k and monitoring pattern $\bar{M}_k$.

Bayesian Semi-parametric Hazard Models

Stratified proportional hazard models

$$\lambda_j(w_k \mid s_k, h_k) = \lambda_{js_k0}(w_k)\exp(g_{j,s_k}(h_k))$$

- E.g., $g_{j,s_k}(h_k;\beta_{j,s_k}) = \bar{L}'_k\beta_{1,j,s_k} + \beta_{2,j,s_k}(\sum_{h=1}^{3} I(V_{k-h} - S_{k-h} < \Delta))$ where $\beta_{j,s_k} = (\beta_{1,j,s_k}, \beta_{2,j,s_k})$.
- Piecewise baseline hazard specification

$$\lambda_{js_k0}(w_k;\theta_{j,s_k}) = \sum_{u=1}^{U} I(w_k \in I_u^{js_k})\theta_{u,j,s_k}$$

Where $\{I_u^{js_k}\}_{u=1}^{U}$ is a partition of $[0, \max W_k]$ into U equally-sized intervals and $\theta_{j,s_k} = \{\theta_{u,j,s_k}\}_u^U$.

Autoregressive Smoothing Prior

$\{X_u\}_{u=1}^{U} \sim LNAR1(\rho, \eta, \sigma)$ with initial condition $\log X_1 = \eta + \sigma\epsilon_1$ and transition relation for $u = 2, 3, \ldots, K$

$$\log X_u = \eta(1-\rho) + \rho \log X_{u-1} + \sigma\epsilon_u$$

where $0 \leq \rho < 1$, $-\infty < \eta < \infty$, and $\sigma > 0$ are the hyperparameters of the process and ϵ_u are i.i.d. $N(0,1)$ random variables.

- Prior mean process: $E[\log X_u] = \eta$
- Prior covariance process: $Corr(\log X_u, \log X_{u-v}) = \rho^v$

Autoregressive Smoothing Prior

$\{X_u\}_{u=1}^U \sim LNAR1(\rho, \eta, \sigma)$ with initial condition $\log X_1 = \eta + \sigma\epsilon_1$ and transition relation for $u = 2, 3, \ldots, K$

$$\log X_u = \eta(1-\rho) + \rho \log X_{u-1} + \sigma\epsilon_u$$

where $0 \leq \rho < 1$, $-\infty < \eta < \infty$, and $\sigma > 0$ are the hyperparameters of the process and ϵ_u are i.i.d. $N(0,1)$ random variables.

- Prior mean process: $E[\log X_u] = \eta$
- Prior covariance process: $Corr(\log X_u, \log X_{u-v}) = \rho^v$

We place separate priors on the baseline hazard rates:

$$\{\theta_{u,j,s_k}\}_{u=1}^U \sim LNAR1(\rho_{j,s_k}, \eta_{j,s_k}, \sigma_{j,s_k})$$

Posterior Causal Estimation via Monte Carlo

We use standard Markov Chain Monte Carlo (MCMC) methods to obtain posterior draws of baseline hazard rates $\{\theta_{u,j,s_k}\}$ and covariate effects β_{j,s_k} for each $s_k \in \mathcal{S}_k$ an $j \in \{0, 1\}$.

Posterior Causal Estimation via Monte Carlo

We use standard Markov Chain Monte Carlo (MCMC) methods to obtain posterior draws of baseline hazard rates $\{\theta_{u,j,s_k}\}$ and covariate effects β_{j,s_k} for each $s_k \in \mathcal{S}_k$ an $j \in \{0, 1\}$.

For each subject i simulate $b = 1, 2, \ldots, B$ events,

$$V_k^{(b)} \sim \lambda_1(w_k \mid s_k, h_{ik}; \theta_{1,s_k}, \beta_{1,s_k})$$

$$T_k^{(b)} \sim \lambda_0(w_k \mid s_k, h_{ik}; \theta_{0,s_k}, \beta_{0,s_k})$$

Set $W_k^{(b)} = \min(V_k^{(b)}, T_k^{(b)})$ and $\delta_k^{(b)} = I(V_k^{(b)} < T_k^{(b)})$

$$P(Y_k^{s_k}(\Delta) = 1 \mid H_k = h_{ik}) \approx \frac{1}{B} \sum_{b=1}^{B} I(W_k^{(b)} - s_k < \Delta, \delta_k^{(b)} = 1)$$

Repeat for each posterior draw to obtain draws of $P(Y_k^{s_k}(\Delta) = 1 \mid H_k = h_{ik})$.

Alternative Approaches

Alternatively, we can fit stratified models for $Y_k(\Delta)$ directly:

$$P(Y_k(\Delta) = 1 \mid S_k = s_k, H_k = h_k) = g^{-1}(f_{s_k}(h_k))$$

- E.g. BART $f_{s_k}(h_k) = \sum_{r=1}^{R} f_{r,s_k}(h_k; \mathcal{T}_{r,s_k})$.
- E.g. Logistic regression $f_{s_k}(h_k) = h_k' \beta_{s_k}$.

Upsides:

- Simple to implement.
- Uncertainty estimation is easy with Bayesian models.

Downsides:

- Must re-train model for new Δ values.
- Information loss.

Predictive Performance Under Censoring

	Low Censoring			High Censoring		
	BayesHaz	BART	Logistic	BayesHaz	BART	Logistic
No Missing Features	.70	.65	.65	.67	.56	.56
Missing Features	.64	.56	.59	.61	.53	.55

Table: Area under the curve (AUCs) of predicted engagement probabilities in held out test set under various covariate missingness/censoring scenarios.

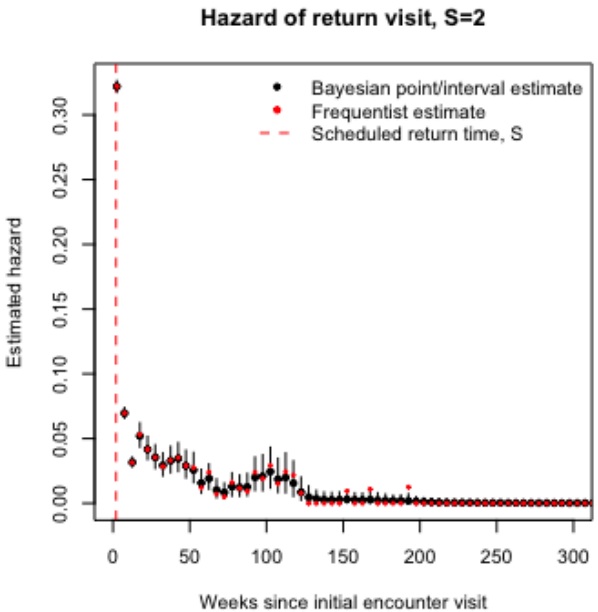
Analyses of First Return in AMPATH

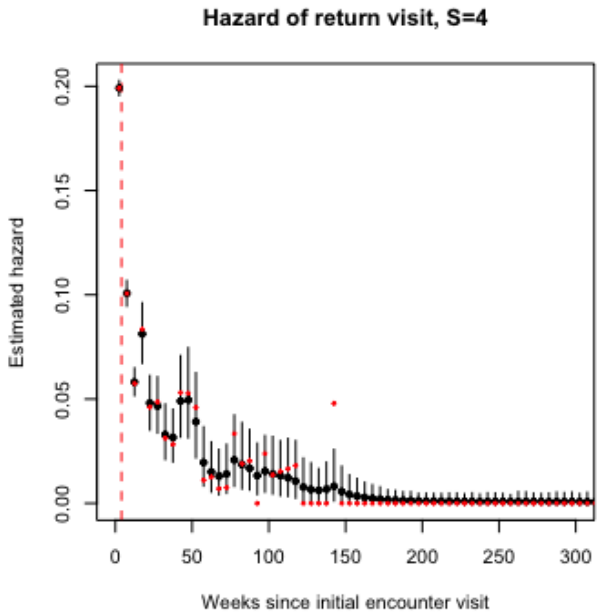
$k = 1$. Data cut with $N \approx 74,000$ patients at initial enrollment visit scheduled to return at either 2, 4, 8, or 12 weeks later.

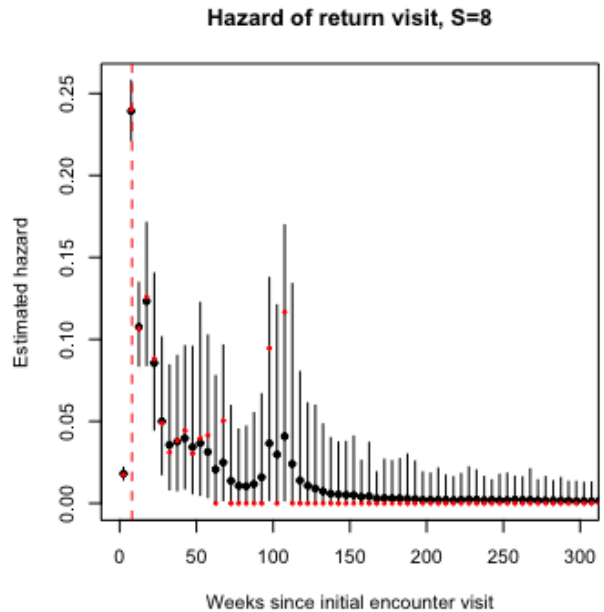
- 85% return for subsequent visit.
- Covariate information on: sex, age, ARV assignment, CD4, travel time.
- CD4 (33%) and travel time (45%) have high rates of missingness.

Bayesian Hazard Estimation

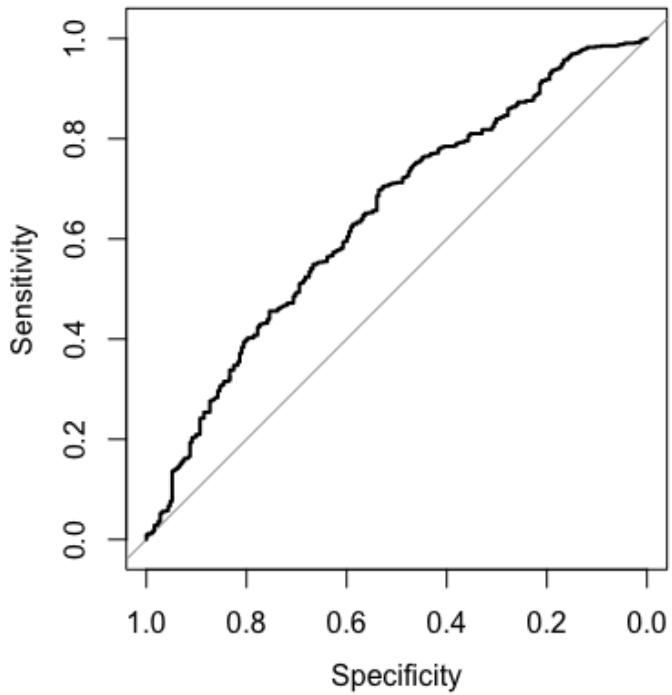
Posterior Inference for $\lambda_{js_k0}(w_k)$ using AMPATH data with $j = 1$ and $k = 1$ and $s_k \in \{2, 4, 8\}$.

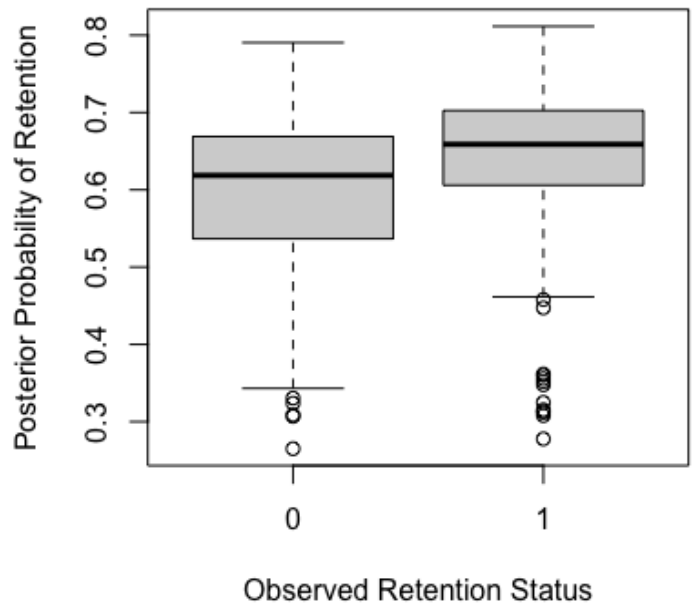






Predictive Assessments





- Out-of-sample AUCs fairly similar with with BART (fully nonparametric) - $\approx 66\%$.

Summary

- Generative modeling - single set of models for several estimands at once.
- Priors provide useful, tailored regularization.
- Bayesian methods provide full inference for all functionals.
- Can be more efficient use of granular data relative to off-the-shelf alternatives.