Bayesian Models for Counterfactual Prediction and Optimization with Incomplete Information: Applications in HIV Care Retention

Arman Oganisian[1],
Allison DeLong[1], Ben Mosong[3],
Ann Mwangi[2,3], Edwin Sang[3], Joseph Hogan[1]

[1] Department of Biostatistics, Brown University, USA
[2]Department of Mathematics, Physics and Computing, Moi University, Kenya
[3]Academic Model Providing Access to Healthcare (AMPATH), Kenya

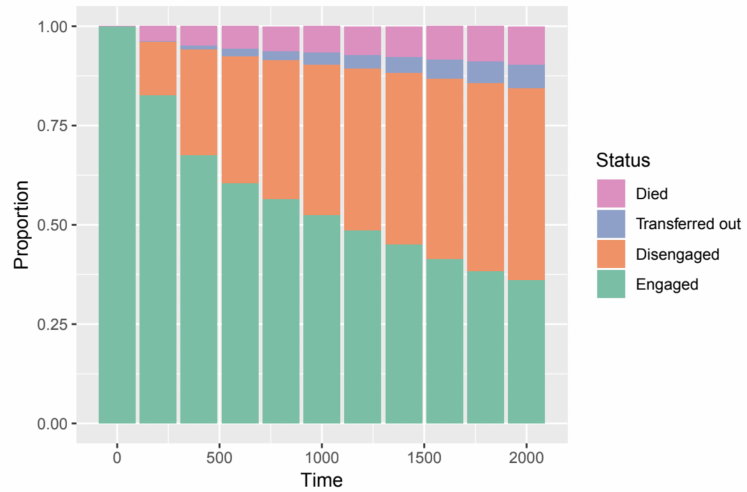
Study Motivation - Closing the HIV Retention Gap

- 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.

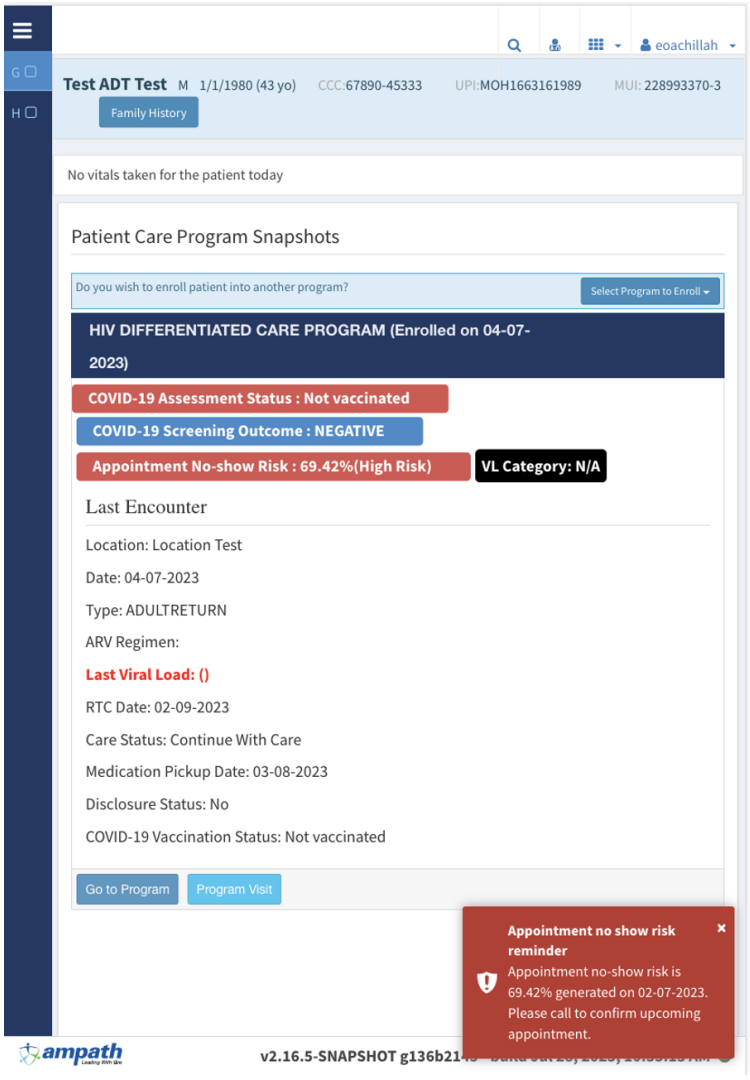
AMPATH Care Program in Western Kenya

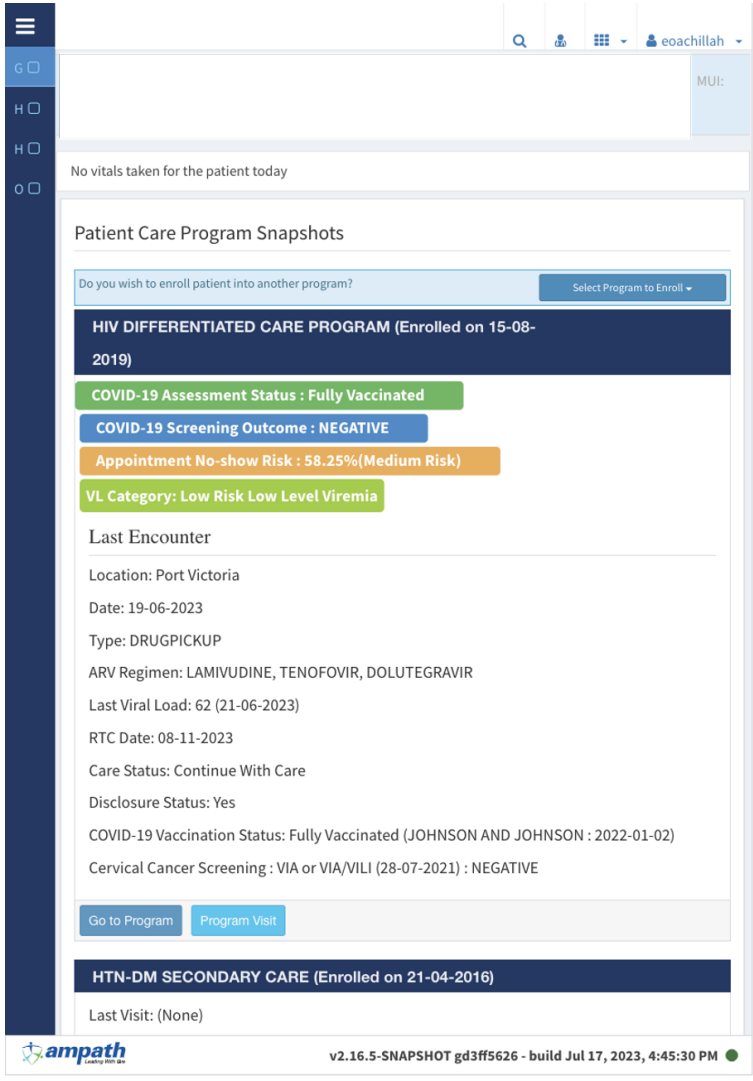
The Academic Model Providing Access to Healthcare (AMPATH) care program treats roughly 150,000 patients with HIV at over 60 urban and rural clinics in western Kenya.

- 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 visit.

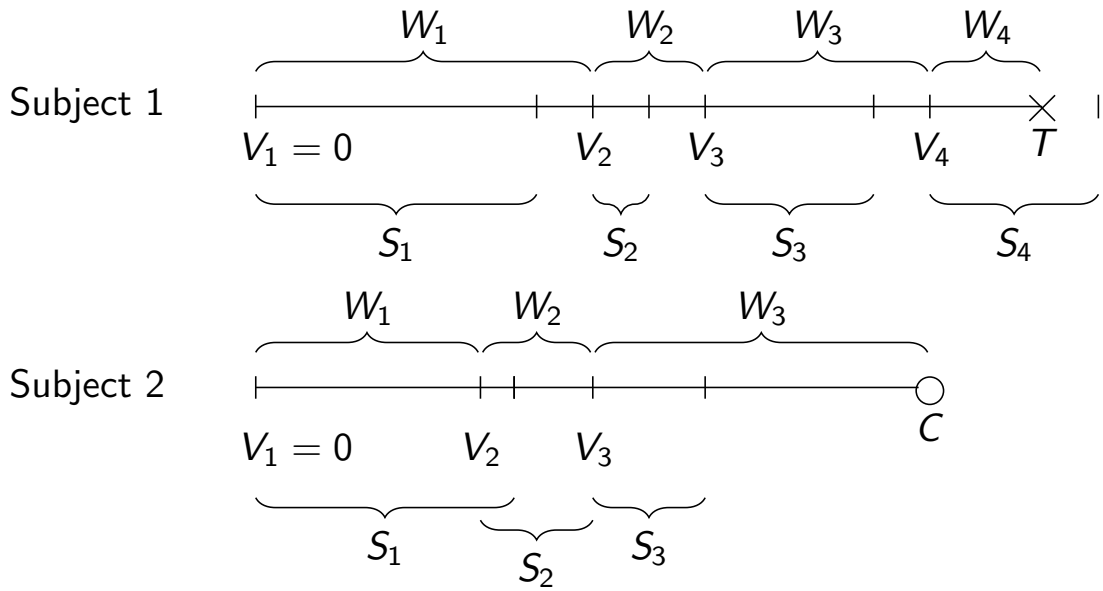


Point-of-Care Decision Support





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 after enrollment visit, $k = 1$.
- Events can be: visit, death, or censoring event.
- $\tilde{L}_k$: at visit k, a set of P features that may (or may not) be available.
- S_k: scheduled return time.
- W_k: Observed weeks until next event.
- $\delta_k \in \{-1, 0, 1\}$: Event indicator for censoring, death, return visit - respectively.

Incomplete Covariate Information

Covariate information such as viral load and CD4 count are assessed sporadically at each visit.

- $\tilde{L}_k = (\tilde{L}_{1k}, \tilde{L}_{2k}, \ldots, \tilde{L}_{Pk})$: $P-$dimensional covariate vector at visit k.
- $M_k = (M_{1k}, M_{2k}, \ldots, M_{Pk})$: indicators at visit k, where $M_{pk} = 1$ indicates that $\tilde{L}_{pk}$ was monitored and $M_{pk} = 0$ indicates that it was not monitored.
- $L_k = \tilde{L}_k \odot M_k$: Observed set of covariates at visit k.
- $L_k^U = \tilde{L}_k \odot (1 - M_k)$ denote the unobserved values at visit k.

Example: $P = 2$ covariates: four possible missingness patterns:

- Pattern $M_k = (1, 1)$: $L_k = (\tilde{L}_{1k}, \tilde{L}_{2k})$;
- Pattern $M_k = (0, 0)$: $L_k = (0, 0)$;
- Pattern $M_k = (1, 0)$: $L_k = (\tilde{L}_{1k}, 0)$;
- Pattern $M_k = (0, 1)$: $L_k = (0, \tilde{L}_{2k})$;

Our Contributions

Complicated data-generating process with many moving parts:

- Stochastic visit process.
- Terminal and censoring events.
- Sporadically observed covariate process.

What do we really mean by "retention"? And how do we compare effects of different scheduling decisions with observational data?
In our work, we

- Formalize retention in the presence of censoring and death.
- Define estimands in terms of counterfactual retention under different hypothetical scheduling decisions.
- Use flexible Bayesian approaches to model return-time distributions.

Potential Δ-Retention and Causal Estimands

Potential Δ-Retention: For a given $\Delta \geq 0$, indicator for if a patient would have returned with Δ weeks of a hypothetical scheduled return time $S_k = s$ following visit k as

$$Y_k^s(\Delta) = I(\underbrace{W_k^s - s}_{\text{Delay Time}} \leq \Delta, \delta_k^s = 1)$$

Where $W_k^s = \min(W_{Tk}^s, W_{Vk}^s)$ and $\delta^s = I(W_{Vk}^s < W_{Tk}^s)$.
Several Δ may be of interest: E.g. Kenya Ministry of Health defines

- Defaulter: a patient who misses their visit by 7 days.
- Loss to follow-up (LTFU): a patient who misses their visit by 90 days.

Potential Δ-Retention and Causal Estimands

Counterfactual Prediction: proportion of the subpopulation at-risk of an event at visit k ($\bar{\delta}_{k-1} = 1$) with available history $H_k = (\bar{L}_k, \bar{M}_k, \bar{S}_{k-1}, \bar{W}_{k-1}, \bar{\delta}_{k-1} = 1)$ that would have been retained under scheduling decision s.

$$\Psi_k^s(H_k; \Delta) = P(Y_k^s(\Delta) = 1 \mid \bar{L}_k, \bar{M}_k, \bar{S}_{k-1}, \bar{W}_{k-1}, \bar{\delta}_{k-1} = 1) \quad \text{for } s \in \mathcal{S}$$

Counterfactual Optimization: Finding an optimal scheduling decision

$$s^*(h_k) = \underset{s \in \mathcal{S}}{\text{argmax}}\ \Psi_k^s(h_k; \Delta)$$

Identification Task

$Y_k^s(\Delta) = I(W_k^s - s \leq \Delta, \delta_k^s = 1)$ is a function of W_k^s and δ_k^s,

So we require identification of sub-density function of potential waiting time

$$f_{kj}^*(W_k^s = w_k, \delta_k^s = j \mid H_k = h_k)$$

Instantaneous probability of transitioning to an event of type j at w_k weeks after visit k had we made scheduling decision s.

Identification Assumptions

We require conditionally exchangeable scheduling

$$W^s_{Vk}, W^s_{Tk} \perp S_k \mid \bar{S}_{k-1}, \bar{W}_{k-1}, \bar{L}_k, \bar{M}_k = \bar{m}_k, \bar{\delta}_{k-1} = 1$$

- Within each missingness pattern, $\bar{M}_k = \bar{m}_k$, covariates observed in that pattern, $\bar{L}_k$, are sufficient to control for confounding.
- Reasonable because missingness is due to monitoring: non-monitored values $\bar{L}^U_k$, were unknown to clinician and so could not have influenced S_k.
- Motivates estimation approach: estimate sub-density models stratified by missingness pattern, each model adjusting for covariates observed in that pattern.

Identification Assumptions

Noninformative censoring: cause-specific hazard of censoring conditionally independent of potential waiting times.

$$\lim_{dw_k \to 0} dw_k^{-1} P(w_k < W_k < w + dw_k, \delta_k = -1 \mid W_k > w, h_k, w_{Vk}^s, w_{Tk}^s) =$$

$$\lim_{dw_k \to 0} dw_k^{-1} P(w_k < W_k < w_k + dw_k, \delta_k = -1 \mid W_k > w, h_k)$$

Related positivity assumptions for scheduling and censoring also required.

Nonparametric Identification

Joint distribution of potential outcomes can be expressed in terms of observed data cause-specific hazards,

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

Requires modeling cause-specific hazards $\lambda_j(w_k \mid s, h_k)$ for $j = \{0, 1\}$.

Bayesian Semiparametric Transition Models

For some covariate vector x

$$\lambda_j(w_k \mid x_k) = \lambda_{jk0}(w_k)\exp(g_{jk}(x_k))$$

- E.g., $g_{jk}(h_k; \beta_{jk})$ can include main effects, interaction effects, spline functions of covariates, etc - governed by coefficient vector β_{jk}.
- Piecewise baseline hazard specification $\lambda_{jk0}(w_k)$ with an autoregressive prior for smoothing.
- Hazard models fully stratified by each scheduling and missingness pattern combination.

Posterior Counterfactual Prediction and Optimization

We use Markov Chain Monte Carlo (MCMC) methods to obtain posterior draws of $\{\lambda_{jk0}, g_{jk}\}$ for each j, k. Then for each $s \in \mathcal{S}_k$ an $j \in \{0, 1\}$.

Posterior Counterfactual Prediction and Optimization

We use Markov Chain Monte Carlo (MCMC) methods to obtain posterior draws of $\{\lambda_{jk0}, g_{jk}\}$ for each j, k. Then for each $s \in \mathcal{S}_k$ an $j \in \{0, 1\}$.

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

$$W_{Vk}^{(b)} \sim \lambda_1(w_k \mid s, h_{ik})$$

$$W_{Tk}^{(b)} \sim \lambda_0(w_k \mid s, h_{ik})$$

Set $W_k^{(b)} = \min(W_{Vk}^{(b)}, W_{Tk}^{(b)})$ and $\delta_k^{(b)} = I(W_{Vk}^{(b)} < W_{Tk}^{(b)})$:

Posterior Counterfactual Prediction and Optimization

We use Markov Chain Monte Carlo (MCMC) methods to obtain posterior draws of $\{\lambda_{jk0}, g_{jk}\}$ for each j, k. Then for each $s \in \mathcal{S}_k$ an $j \in \{0, 1\}$.

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

$$W_{Vk}^{(b)} \sim \lambda_1(w_k \mid s, h_{ik})$$

$$W_{Tk}^{(b)} \sim \lambda_0(w_k \mid s, h_{ik})$$

Set $W_k^{(b)} = \min(W_{Vk}^{(b)}, W_{Tk}^{(b)})$ and $\delta_k^{(b)} = I(W_{Vk}^{(b)} < W_{Tk}^{(b)})$:

Posterior prediction:

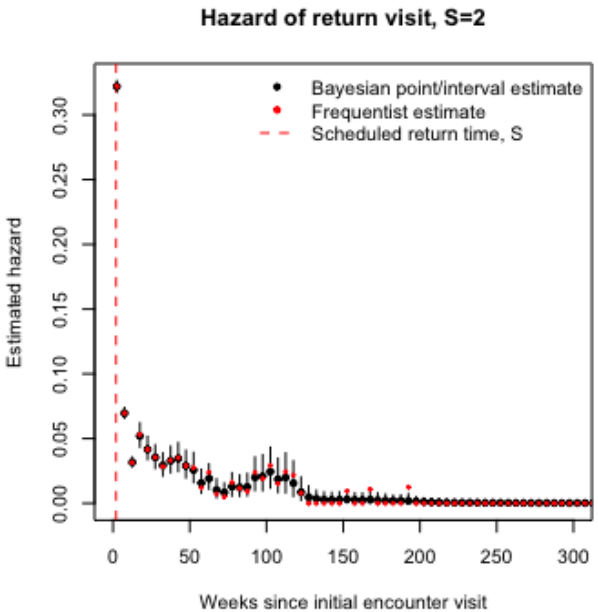
$$P(Y_k^s(\Delta) = 1 \mid H_k = h_{ik}) \approx \frac{1}{B} \sum_{b=1}^{B} I(W_k^{(b)} - s < \Delta, \delta_k^{(b)} = 1)$$

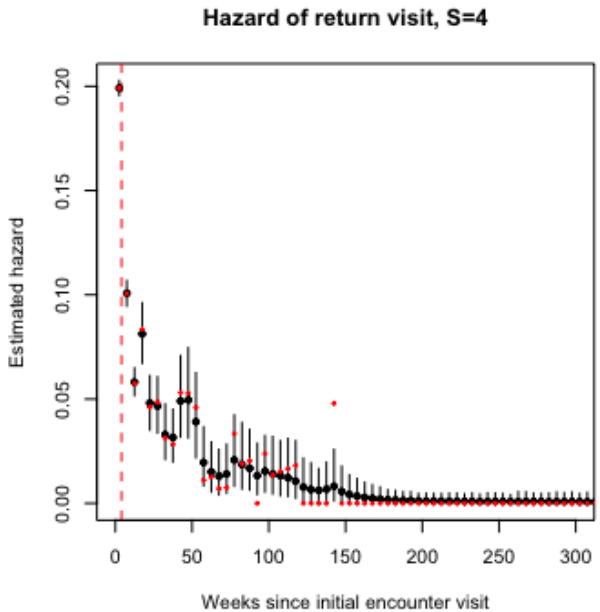
Posterior Optimization: For each draw, $s^*(h_{ik}) = \underset{s \in \mathcal{S}}{\text{argmax}}\ P(Y_k^s(\Delta) = 1 \mid h_{ik})$.

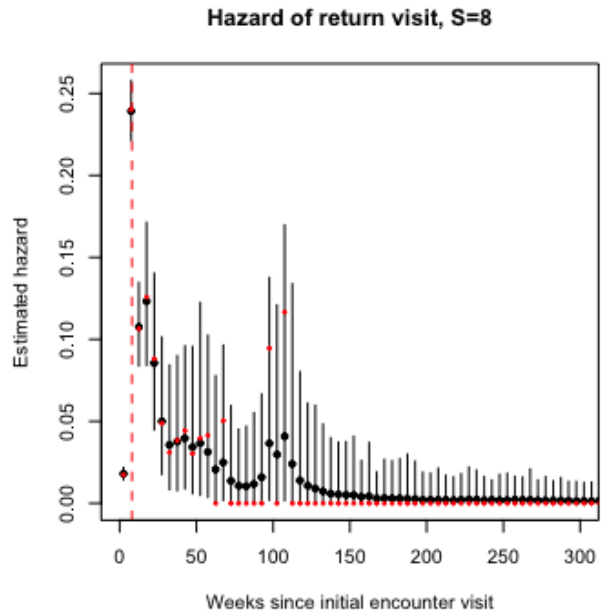
Return-Time Analysis in AMPATH: Summary Statistics

	Clinic			
	Kitale (n=3398)	Busia (n=2891)	UGDH (n=2438)	KCRH Module A (n=2284)
Missing:				
viral load only	1389 (.41)	607 (.21)	303 (.12)	895 (.39)
CD4 only	93 (.03)	114 (.04)	50 (.02)	73 (.03)
both	1907 (.56)	2167 (.75)	2077 (.85)	1313 (.57)
neither	9 (<.01)	3 (<.01)	8 (<.01)	3 (<.01)
Obs. log viral load	7.7 (4.8-10.1)	8.0 (5.7-10.6)	7.5 (4.7-9.9)	7.1 (5.0-7.7)
Obs. log CD4	5.2 (4.6-6.1)	5.6 (5.1-6.3)	5.4 (4.9-6.2)	5.2 (4.8-6.0)
Age at enroll.	36.1 (28.0-42.4)	36.0 (28.0-42.7)	35.0 (26.8-41.7)	40.2 (31.8-47.5)
Male	1261 (0.37)	1064 (0.37)	895 (0.37)	890 (0.39)
On ARV	2710 (0.80)	2688 (0.93)	2169 (0.89)	2265 (0.99)
Sched. return in				
Two weeks	2848 (0.84)	2099 (0.73)	2004 (0.82)	1899 (0.83)
Four weeks	485 (0.14)	671 (0.23)	354 (0.15)	348 (0.15)
Eight weeks	65 (0.02)	121 (0.04)	80 (0.03)	37 (0.02)

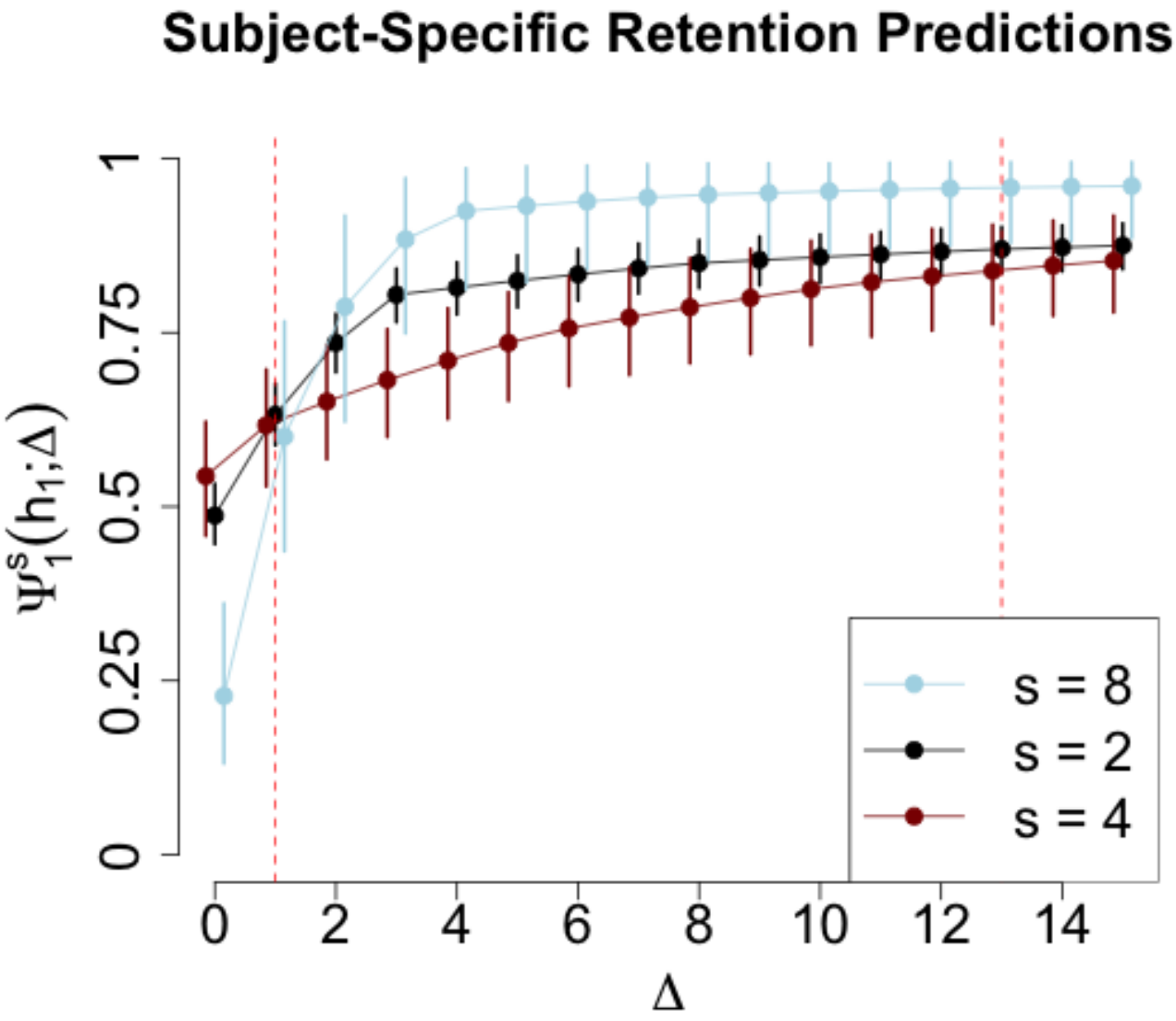
Return-Time Hazard Estimates and Clumping



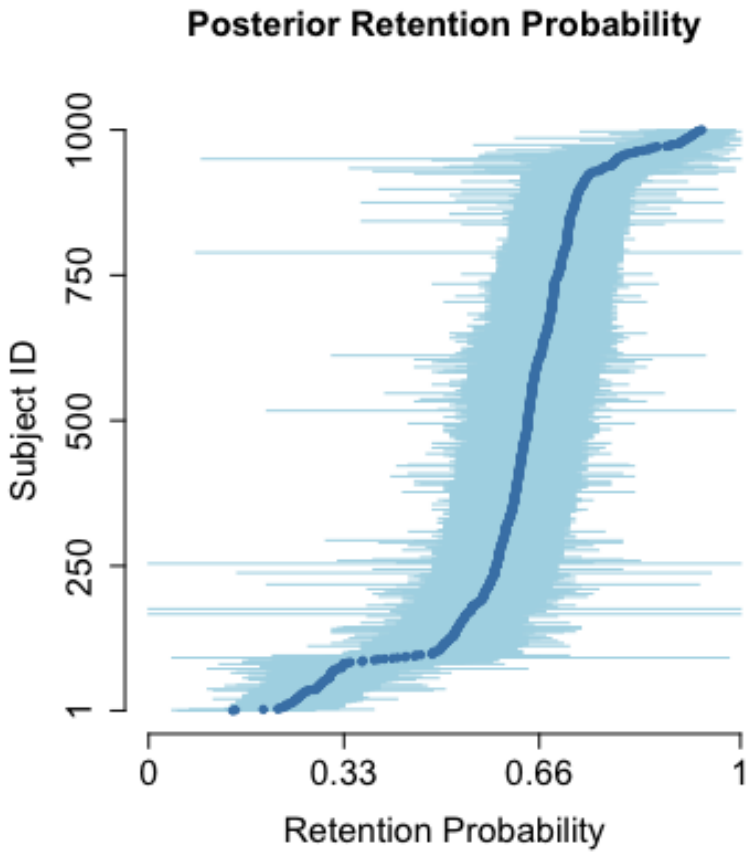


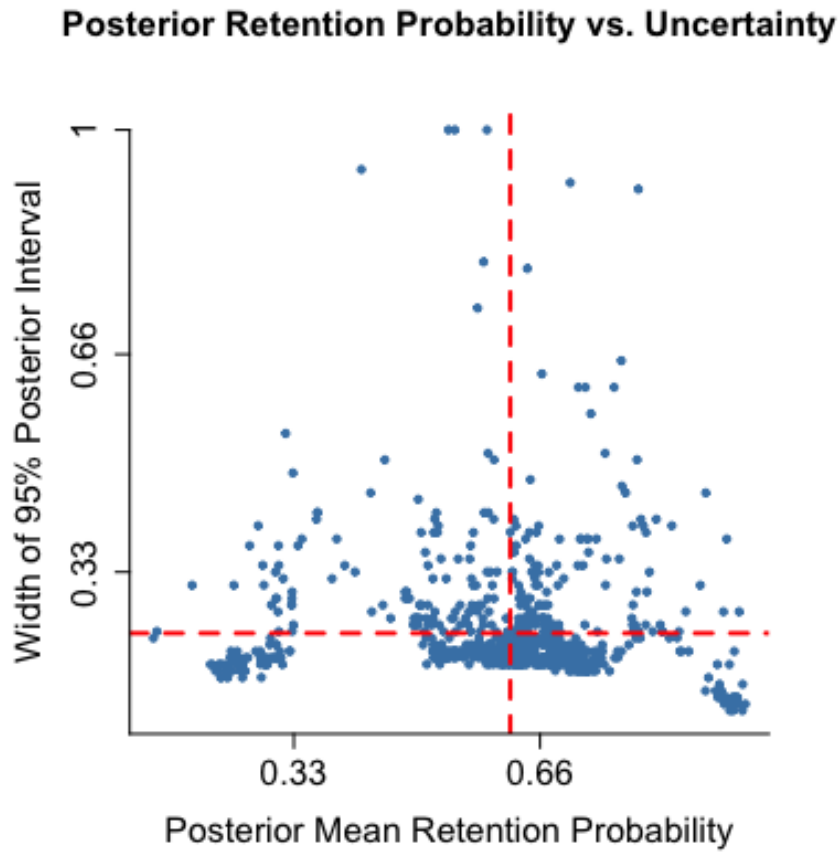


Posterior Retention Predictions



Posterior Retention Predictions





Conclusions

- Decision recommendation/optimization is properly framed as a causal task.
- When predicting potential retention, it's important to account for competing and censoring events.
- Generative Bayesian modeling of underlying return-time outcome superior to dichotomizing in some cases.
- What's the causal effect of the causal inference?

Related Work

- Ji H, Oganisian A. (2023). causalBETA: An R Package for Bayesian Semiparametric Casual Inference with Event-Time Outcomes. arXiv preprint arXiv:2310.12358.
- Oganisian A, Getz KD, Alonzo TA, Aplenc R, Roy JA. (2024). Bayesian semiparametric model for sequential treatment decisions with informative timing, Biostatistics, `https://doi.org/10.1093/biostatistics/kxad035`
- Oganisian A, Roy JA. (2021) A practical introduction to Bayesian estimation of causal effects: Parametric and nonparametric approaches. Statistics in Medicine. `https://doi.org/10.1002/sim.8761`

Contact: email: arman_oganisian@brown.edu; twitter: @StableMarkets
Acknowledgements: NIH R01 AI167694; PCORI Award ME-2023C1-31348