Longitudinal Causal Inference with Latent Variable Models

Yizhen Xu

Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University
Department of Biostatistics, Johns Hopkins University

February 13th, 2023 @ University of Toronto

Project Types

Causal Inference

Latent Variable Modeling and Ensemble Methods

Projects - Causal Inference

Longitudinal Bayesian causal inference

- using multinomial probit BART to examine the impact of ART initiation on the HIV care cascade

  joint work with Joseph Hogan, Michael Daniels, et al.

  - no unmeasured confounding

- **using multivariate GLMM to investigate the efficacy of mycophenolate in treating scleroderma**

  joint work with Scott Zeger, Jisoo Kim, et al.

  - unmeasured time-invariant factors as patient heterogeneity

Multiply robust causal mediation analysis with continuous treatments

joint work with Amir Ghassami, Numair Sani, and Ilya Shpitser

Projects - Latent Variable Modeling and Ensemble Methods

Latent Variable Mixture Model for Assessing AD Biomarkers

joint work with Zheyu Wang

- time-varying unmeasured factors

Longitudinal Probabilistic Clustering of Biomarker Trajectories for Scleroderma

joint work with Jisoo Kim, Scott Zeger, et al.

Augmentation samplers for multinomial probit BART

joint work with Joseph Hogan, Michael Daniels, et al.

Submitted to Journal of Computational and Graphical Statistics

Student paper award, Applied Statistics Symposium, 2020

Classification using ensemble learning under weighted misclassification loss

joint work with Joseph Hogan, Tao Liu, et al.

Statistics in Medicine, 2019

Student paper award, ASA Section on Risk Analysis, 2017

Causal Framework for Treatment Evaluation using Multivariate Generalized Linear Mixed-Effects Models with Longitudinal Data

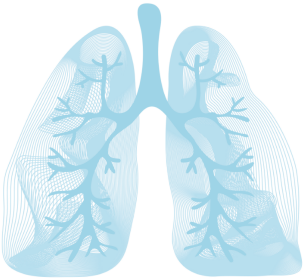
Question
Does mycophenolate (MMF) improve biomarker progression among different subgroups of patients with scleroderma?

Data
Longitudinal clinical data on 506 patients with scleroderma, 194 of whom were previously treated with MMF.

Contribution
Bayesian causal inference for a multivariate generalized linear mixed-effects model (MGLMM) comparing treatment paths

Scleroderma

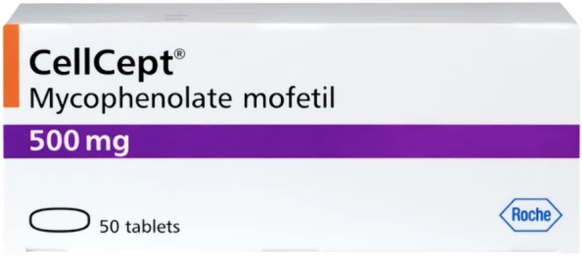
Scleroderma is a chronic autoimmune disease marked by hardening of the skin and internal organs.

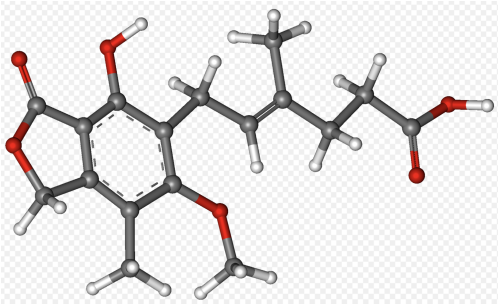


Sources: labblog.uofmhealth.org and sclerodermanews.com

Immunosuppressant medication

In Scleroderma Lung Study II, **mycophenolate** (MMF) resulted in improvements in the modified Rodnan skin score (mRSS) among diffuse patients over 2 yr.





Motivation

Aims

Decision support tool for treating patients in clinic

Compare disease progression under different treatment paths for **subgroups**

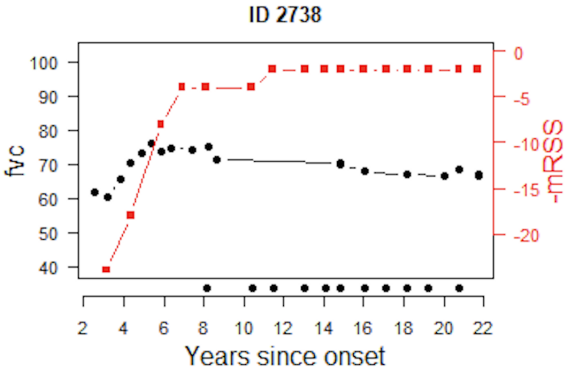
Complexities in Clinical Datasets

Treatment assignment involves important source of **unmeasured confounding**

Doctors decide by looking at **biomarker trends** across ten years for an individual

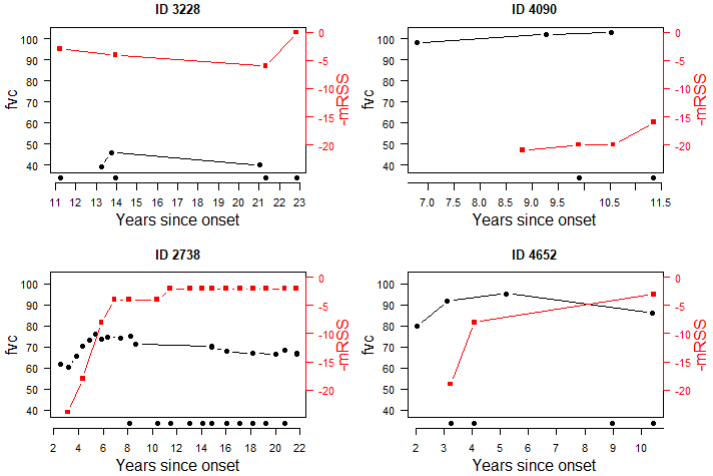
Irregular measurements and potentially informative missingness

Data Structure

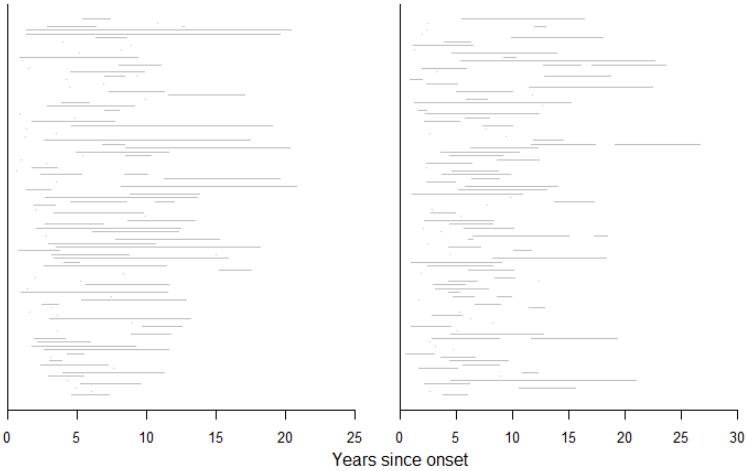


- Data source: The Johns Hopkins Precision Medicine Analytics Platform (PMAP) Registry
- Lung function measurements
  FVC (forced vital capacity), DLCO (Diffusing capacity for carbon monoxide)
- Skin thickness
  mRSS (modified Rodnan skin score)

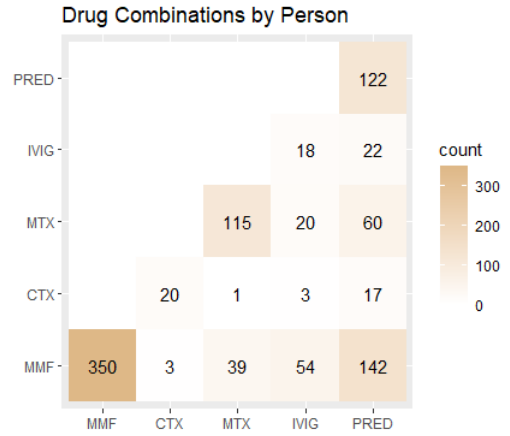
Data Structure

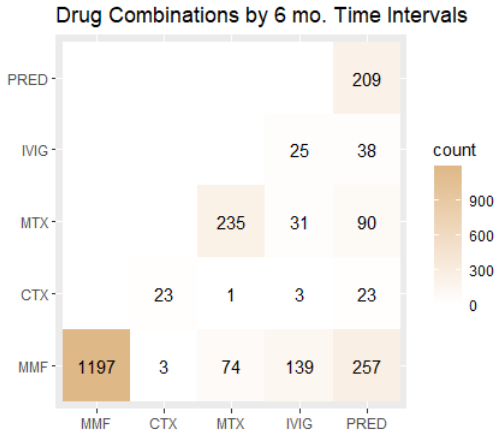


Data Structure



Data Structure - Framing a Treatment Question





Challenges

Unmeasured heterogeneity

- Clinician's treatment decisions are potentially related to unmeasured factors
- Unmeasured factors also influence biomarker progression
- Standard g-estimation methods lead to biased effect estimates when unmeasured confounders are present (Yang and Lok, 2018)

Proposal

Goal

- Account for patient heterogeneity in treatment assignment and biomarkers progression for longitudinal causal inference
- Estimate marginal population and subgroup average effect of dynamic treatment regimes

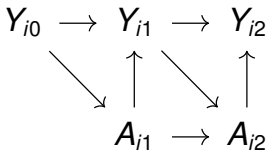
Solution

- Use random effects to characterize time-invariant unmeasured factors
- Adopt the potential outcomes framework and incorporate multivariate generalized linear mixed-effects models (MGLMM)

Formulating Treatment Effect

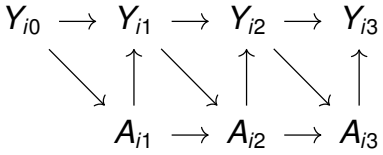
For patient i at time t,

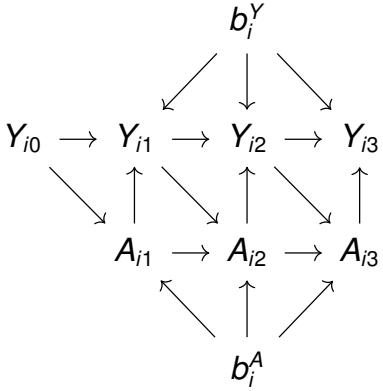
- Y_{it}: outcomes (biomarkers)
- A_{it}: binary treatment status



What is the causal effect of treatment paths, e.g. $\bar{a}_2 = (1, 1)$ versus $\bar{a}'_2 = (0, 0)$?

Time-invariant Unmeasured Heterogeneity





$$(b_i^Y, b_i^A) \sim MVN(0, G)$$

MGLMM

$$
\begin{aligned}
Y_{it} &= \eta_Y(V_i, \overline{A}_{i,t}, \overline{Y}_{i,t-1}; b_i^Y, \theta^Y) + e_{it}^Y \\
P(A_{it} = 1 | A_{i,t-1} = 0) &= \text{logit}^{-1}\{\eta_A(V_i, \overline{A}_{i,t-1}, \overline{Y}_{i,t-1}; b_i^A, \theta^A)\} \\
b_i &= (b_i^Y, b_i^A) \sim MVN(0, G)
\end{aligned}
$$

$$
G = \begin{pmatrix} s_Y^2 & \rho s_Y s_A \\ \rho s_Y s_A & s_A^2 \end{pmatrix}, \quad e_{it}^Y \sim N(0, \sigma^2), \quad (b_i^A, b_i^Y) \perp e_{it}^Y
$$

For patient i at time t,

- Y_{it}: outcomes
- A_{it}: binary treatment status
- V_i: baseline covariates

Connection between Random Effects and Unmeasured Confounders

- Model estimation relies on a posited value of s_A
- The random effects (b_i^Y, b_i^A) are context-specific representation of time-invariant unmeasured confounding.
  - Under model assumptions,
    $\rho = 0 \Rightarrow cov(b_i^A, b_i^Y) = 0 \Rightarrow$ no unmeasured confounders
  - Repeated measurement makes it possible to characterize individual heterogeneity that arises from unmeasured factors
  - Built-in sensitivity analysis, with sensitivity parameter being s_A in $b_i^A \sim N(0, s_A^2)$ for marginal effect

Assumptions

Regime q, potential outcome $Y(q)$. For $t = 0, \ldots, T$,

1. Consistency
$$\overline{Y}_t = \overline{Y}_t(q) \text{ if } \overline{A}_t = \overline{a}_t(q)$$

2. Positivity
$$P(A_{t+1} = a_{t+1}(q) | V, \overline{A}_t = \overline{a}_t(q), \overline{Y}_t, b_i^A) > 0$$
with probability 1 for $t \geq 0$

3. Sequential exchangeability given b_i^A:
$$\overline{Y}_{(t+1):(t+\tau)}(q) \perp A_{t+1} | V, \overline{A}_t = \overline{a}_t(q), \overline{Y}_t, b_i^A$$
for $\tau > 0$

Example

Assume conditional sequential exchangeabiltiy,

$$Y_{h+1}(q) \perp A_{h+1} | V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^A$$

marginal subgroup effect identified as

$$\begin{aligned} & P(Y_{h+1}(q) | V, \bar{A}_{0:h}, \bar{Y}_{0:h}) \\ = & \int P(Y_{h+1}(q) | V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^A) P(b_i^A | V, \bar{A}_{0:h}, \bar{Y}_{0:h}) \, db_i^A \\ = & \iint P(Y_{h+1}(q) | V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^Y) P(b_i^Y | V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^A) db_i^Y \\ & P(b_i^A | V, \bar{A}_{0:h}, \bar{Y}_{0:h}) \, db_i^A \end{aligned}$$

Special Case

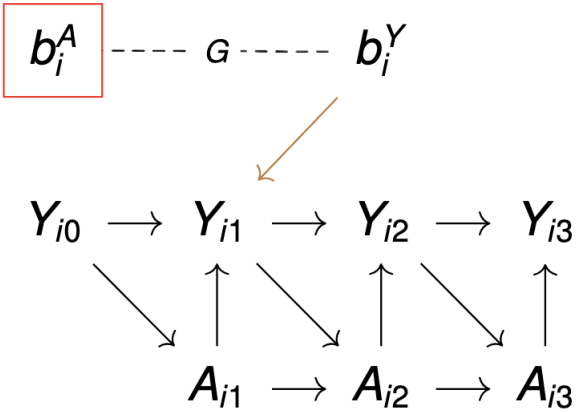
Assume no treatment assignment heterogeneity,

$$\text{var}(b_i^A) = 0.$$

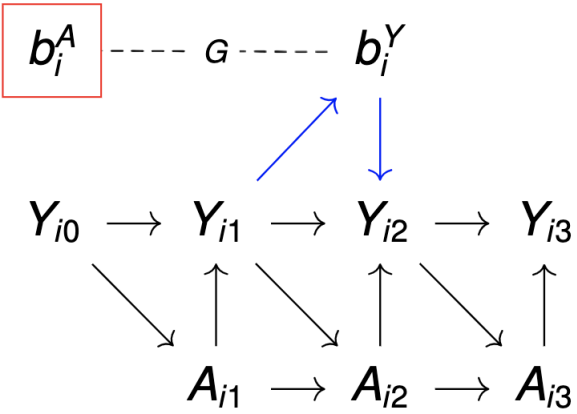
$\text{cov}(b_i^A, b_i^Y) = 0$, marginal subgroup effect identified as

$$P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}) = \int P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^Y) P(b_i^Y|V, \bar{A}_{0:h}, \bar{Y}_{0:h})\, db_i^Y$$

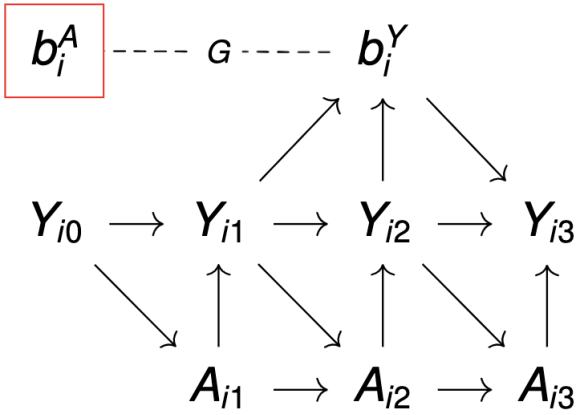
Information Flow Diagram in G-computation



Information Flow Diagram in G-computation



Information Flow Diagram in G-computation



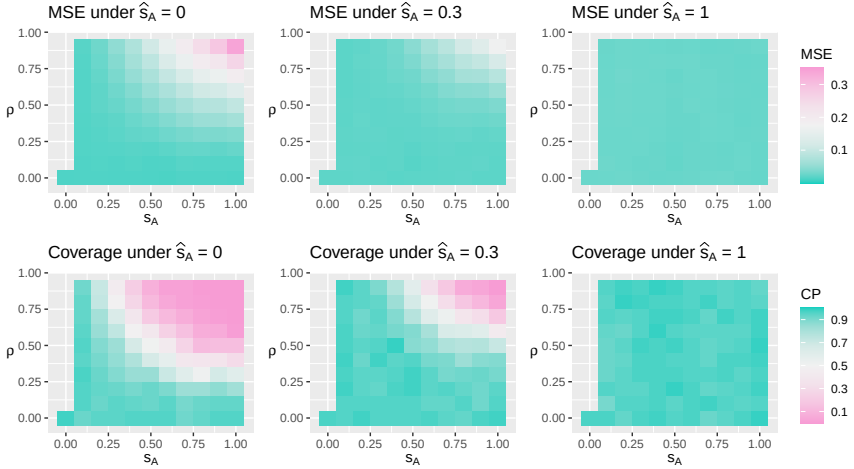
Simulation

$$Y_{it} = \beta_0^Y + V_i\beta_1^Y + t\beta_2^Y + \sum_{k=1}^{2} 1\{\sum_{s=1}^{t} A_{i,s} = k\}\beta_{3k}^Y + Y_{i,t-1}\beta_4^Y + b_{i0}^Y + e_{it}^Y$$

$$\text{logit}\{P(A_{it} = 1|A_{i,t-1} = 0)\} = \beta_0^A + V_i\beta_1^A + t\beta_2^A + Y_{i,t-1}\beta_4^A + b_{i0}^A$$

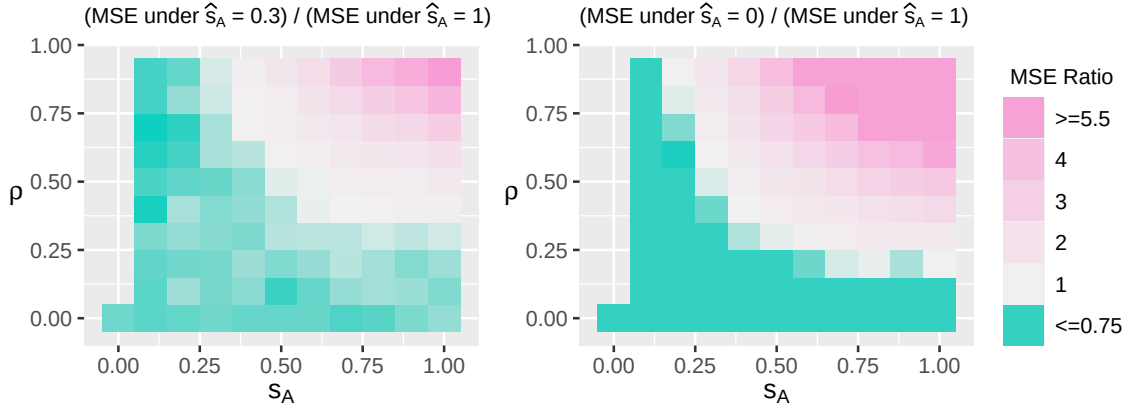
- $t \in \{1, 2\}$, $G = \begin{pmatrix} s_Y^2 & \rho s_Y s_A \\ \rho s_Y s_A & s_A^2 \end{pmatrix}$, $s_Y = 0.8$, $e_{it}^Y \sim N(0, 0.4^2)$
- V_i: baseline covariate
- simulate 100 datasets with sample size $n = 500$ for 101 combinations $(s_A, \rho) \in (0, 0) \cup \{s_A \in \{0.1, \ldots, 0.9, 1\}, \rho \in \{0, 0.1, \ldots, 0.9\}\}$.

Simulation



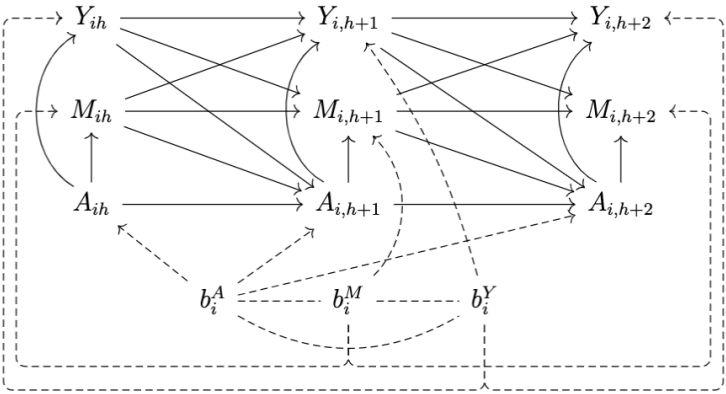
Under true treatment effect being 1.2 at the second time point, the figure displays mean squared error and posterior coverage for mixed population ATE under different assumed model parameter $\hat{s}_A$. Green indicates better estimation, e.g. lower MSE and higher coverage probability.

Simulation



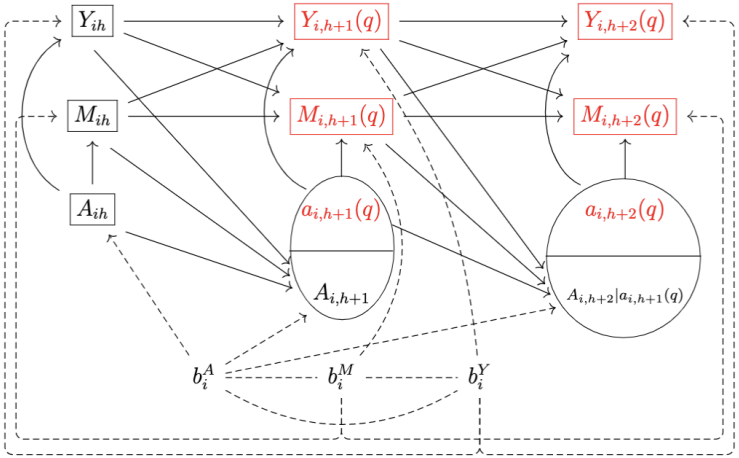
MSE ratio.

Analysis



- Y_{it} longitudinal outcomes
- M_{it} time-varying confounders
- A_{it}: treatments

Analysis



- red: counterfactuals
- box: variables related to the sequential update of (b_i^M, b_i^Y) in g-computation

Analysis

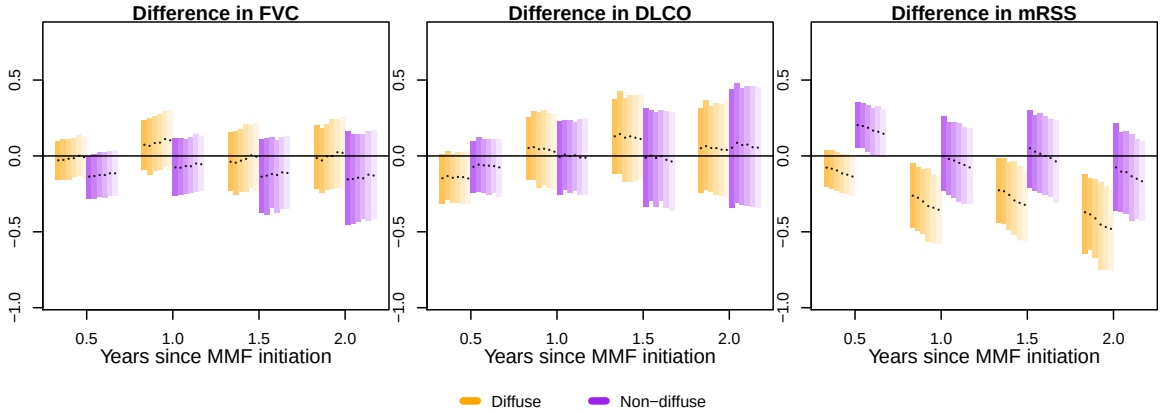


Figure: Application causal estimation by subgroup.

What's new about our method

- Random effects are involved in the g-computation for calculating subgroup causal effect
- We sequentially update subject-specific heterogeneity in biomarker dynamics as history information accumulates over time
- A way to represent time-invariant unmeasured factors in Bayesian causal inference
- Has the potential to be extended to guide the inclusion of latent variable models in Bayesian causal inference

What's new about our method

- Incorporates propensity score (PS) in Bayesian causal inference
  - A major debate: the role of PS in Bayesian causal inference (Li et al. 2022)
  - PS model is ignorable in Bayesian inference of population ATE and mixed ATE
  - Existing ways: specify outcomes distribution based on PS, shared parameters/priors between PS and outcome models, posterior-based IPW or DR estimators

Future Directions

Statistical

Relax the normality of random effects

Make the predictive part less parametric, e.g. BART

Accommodate more general treatment patterns, e.g. dosage and taper-offs

Applied

Drug combinations

Other organ function biomarkers

Joint analyses with other autoimmune diseases

Other projects

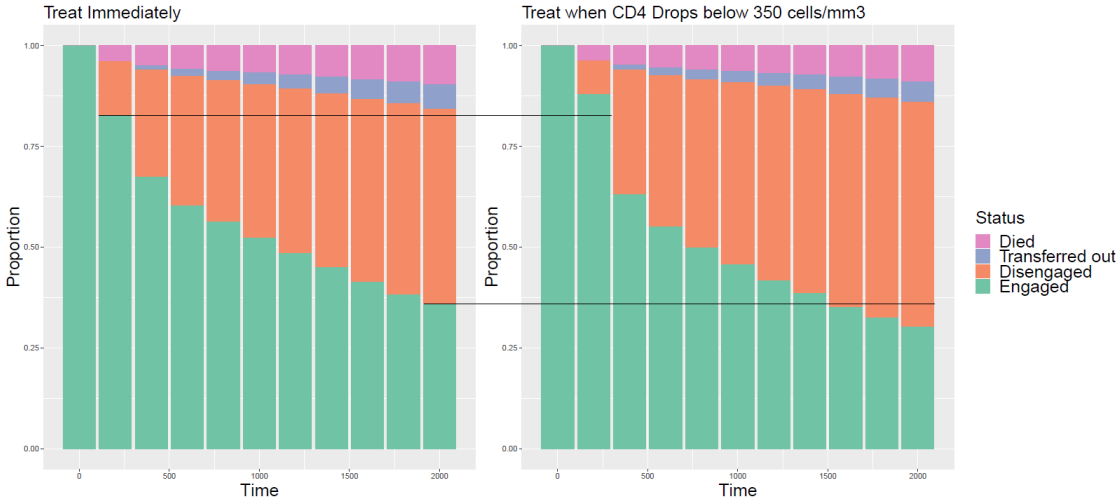
Project 1
Bayesian Framework for Predictive and Causal Modeling with Application to HIV Care Cascade

Question
Does early initiation of ART promote engagement and retention in HIV care?

Data
Longitudinal EHRs from 90,000 people in an HIV care program in Kenya

Contribution
Bayesian causal inference for a multistate model comparing dynamic treatment regimes

Counterfactual Simulation



Project 1
Bayesian Framework for Predictive and Causal Modeling with Application to HIV Care Cascade

- No unmeasured confounding
- Multinomial probit BART in Bayesian g-computation

  - assumption of correct model specification
  - inter-related outcome categories
  - extremely sparse outcome categories

Project 1

Bayesian Framework for Predictive and Causal Modeling with Application to the HIV Care Cascade

Abstract

We develop a Bayesian method for semiparametric predictive and causal inference on longitudinal multi-state outcomes. We use a Bayesian causal formulation of the g computation algorithm (GCA) and incorporate Bayesian additive regression trees (BART) as the generative components to reduce the need for parametric model specifications and enable machine learning based prediction of events using time-evolving models. Our method provides a way to conduct predictive and causal inference based on posterior predictive distributions of the counterfactual outcomes over time. The work is motivated by the electronic health records (EHRs) from the Academic Model Providing Access to Healthcare (AMPATH) in Kenya. We use the data to investigate dynamic treatment regimes by comparing their causal effects on the progression of patients' engagement in care through the HIV care cascade, which is framed as a time-evolving multi-state outcome with dependent alternatives. Under settings involving dynamical systems that can be described by state transitions over time, the proposed framework can be applied broadly to understand complex interventions that may depend on the progression of outcome and confounders when massive data are available.

Keywords: Bayesian G computation, BART, competing risks

Project 2

Augmentation Samplers for Multinomial Probit Bayesian Additive Regression Trees

Abstract

The multinomial probit (MNP) (Imai and van Dyk, 2005) framework is based on a multivariate Gaussian latent structure, allowing for natural extensions to multilevel modeling. Unlike multinomial logistic models, MNP does not assume independent alternatives. Kindo et al. (2016) proposed multinomial probit BART (MPBART) to accommodate Bayesian additive regression trees (BART) formulation in MNP. The posterior sampling algorithms for MNP and MPBART are collapsed Gibbs samplers. Because the collapsing augmentation strategy yields a geometric rate of convergence no greater than that of a standard Gibbs sampling step, it is recommended whenever computationally feasible (Imai and van Dyk, 2005, Liu, 1994). While this strategy necessitates simple sampling steps and a reasonably fast converging Markov chain, the complexity of stochastic search for posterior trees may undermine its benefit. We address this problem by sampling posterior trees conditional on the constrained parameter space and compare our proposals to that of Kindo et al. (2016), who sample posterior trees based on an augmented parameter space. We also compare to the approach by Sparapani et al. (2021) that specified the multinomial model in terms of conditional probabilities. In terms of MCMC convergence and posterior predictive accuracy, our proposals are comparable to the conditional probability approach and outperform the augmented tree sampling approach. We also show that the theoretical mixing rates of our proposals are guaranteed to be no greater than the augmented tree sampling approach.

Keywords: keyword: Additive Regression Trees, Bayesian Data Augmentation, Categorical Outcomes, Latent Models

https://github.com/yizhenxu/GcompBART

Project 3

Causal Framework for Treatment Evaluation using Multivariate Generalized Linear Mixed-Effects Models with Longitudinal Data

Yizhen Xu, Jisoo Kim, Ami Shah, Scott Zeger

December 9, 2022

Abstract

Dynamic prediction of causal effects under different treatment regimes conditional on individual's characteristics and longitudinal history is an essential problem in precision medicine. This is challenging in practice because outcomes and treatment assignment mechanisms are unknown in observational studies, an individual's treatment efficacy is a counterfactual, and the existence of selection bias is often unavoidable.

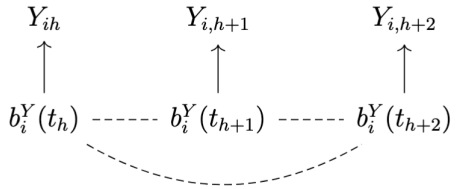
We propose a Bayesian framework for identifying the counterfactual benefits of treatment regimes using Bayesian g-computation [26;41] with multivariate generalized linear mixed effect models. Unmeasured time-invariant factors are identified as subject-specific random effects in the joint distribution of outcomes, time-varying confounders, and treatment assignments. Existing methods mostly focus on balancing the confounder distributions of observations between different treatments. We propose a sequential ignorability assumption conditional on the treatment assignment heterogeneity. This is analogous to balancing the distribution of observable variables as well as the latent tendency toward each treatment due to unmeasured time-invariant factors.

Longitudinal causal inference; latent variable modeling; random effects models; g-computation

Project 4
Probabilistic Clustering using Shared Latent Variable Model for Assessing Alzheimer's Disease Biomarkers

Identify subgroups with higher risk profile before irreversible damage

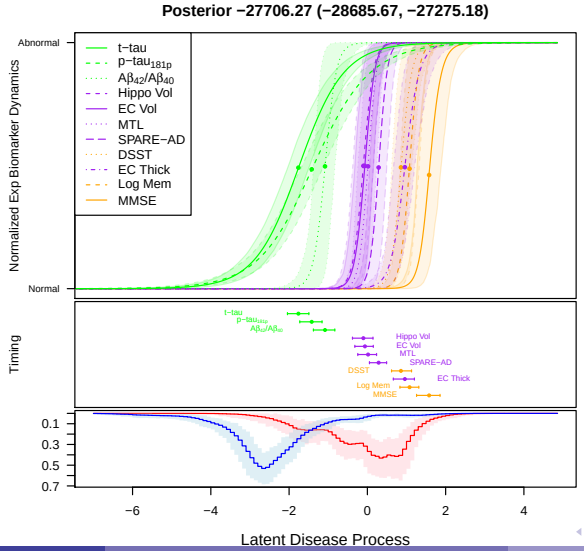
Sequential ordering of biomarkers along the latent pathophysiological pathway



Solution

Mixture model with continuous cluster and subject-specific latent utility

Project 4
Probabilistic Clustering using Shared Latent Variable Model for Assessing Alzheimer's Disease Biomarkers



Project 4

Probabilistic Clustering using Shared Latent Variable Model for Assessing Alzheimer's Disease Biomarkers

Yizhen Xu* Zheyu Wang

December 6, 2022

Abstract

The pathophysiological process of neurodegenerative diseases can initiate a decade or more before noticeable symptoms, causing irreversible brain structure changes by the time of clinical diagnosis. It is therefore crucial to study biomarker dynamics during the preclinical stages, when patients' brain functions are still largely intact for more efficient intervention and treatment development. Two of the main challenges in biomarker-disease modeling are: first, the disease progression is not directly observable, especially during the preclincial stage of neurodegenerative diseases where most interest resides; second, biomarker and disease dynamics have notable individual heterogeneity. Increasing research identified subgroups with not well-understood biomarker patterns, which may be explained by co-morbidities in the elderly population[4] or brain resilience[2].

To study the sequential ordering of biomarkers along the pathophysiological pathway, we focus on the Alzheimer's disease (AD) and propose a latent variable mixture model that quantifies disease progression by a dynamic latent metric. The latent metric accounts for individualized time-varying heterogeneity and is cluster-specific for identifying subgroups with systematically different risk profiles. Model estimation uses Hamiltonian Monte Carlo and is validated by simulation studies. The number of clusters is determined by BIC. When applied to the BIOCARD data, the estimated sequential order of biomarkers is consistent with the hypothetical model of biomarker dynamics in Jack Jr et al.[13]. Application results are further evaluated by investigating the conversion between AD clinical diagnoses within posterior clusters.

Project 5

Stat Med. 2019 May 20; 38(11): 2002–2012.
PMID: 30609090
Published online 2019 Jan 4. doi: 10.1002/sim.8082

Classification using Ensemble Learning under Weighted Misclassification Loss

Yizhen Xu,[1,*] Tao Liu,[1] Michael J. Daniels,[2] Rami Kantor,[3] Ann Mwangi,[4,5] and Joseph W Hogan[1,4]

Author information · Copyright and License information Disclaimer

Abstract

Go to:

Binary classification rules based on covariates typically depend on simple loss functions such as zero-one misclassification. Some cases may require more complex loss functions. For example, individual-level monitoring of HIV-infected individuals on antiretroviral therapy (ART) requires periodic assessment of treatment failure, defined as having a viral load (VL) value above a certain threshold. In some resource limited settings, VL tests may be limited by cost or technology, and diagnoses are based on other clinical markers. Depending on scenario, higher premium may be placed on avoiding false-positives which brings greater cost and reduced treatment options. Here, the optimal rule is determined by minimizing a weighted misclassification loss/risk.

We propose a method for finding and cross-validating optimal binary classification rules under weighted misclassification loss. We focus on rules comprising a prediction score and an associated threshold, where the score is derived using an ensemble learner. Simulations and examples show that our method, which derives the score and threshold jointly, more accurately estimates overall risk and has better operating characteristics compared with methods that derive the score first and the cutoff conditionally on the score especially for finite samples.

Project 6

Multiply Robust Causal Mediation Analysis with Continuous Treatments

AmirEmad Ghassami*, Numair Sani*, Yizhen Xu*, Ilya Shpitser

Abstract

In many applications, researchers are interested in the direct and indirect causal effects of an intervention on an outcome of interest. Mediation analysis offers a rigorous framework for the identification and estimation of such causal quantities. In the case of binary treatment, efficient estimators for the direct and indirect effects are derived by Tchetgen Tchetgen and Shpitser (2012). These estimators are based on influence functions and possess desirable multiple robustness properties. However, they are not readily applicable when treatments are continuous, which is the case in several settings, such as drug dosage in medical applications. In this work, we extend the influence function-based estimator of Tchetgen Tchetgen and Shpitser (2012) to deal with continuous treatments by utilizing a kernel smoothing approach. We first demonstrate that our proposed estimator preserves the multiple robustness property of the estimator in Tchetgen Tchetgen and Shpitser (2012). Then we show that under certain mild regularity conditions, our estimator is asymptotically normal. Our estimation scheme allows for high-dimensional nuisance parameters that can be estimated at slower rates than the target parameter. Additionally, we utilize cross-fitting, which allows for weaker smoothness requirements for the nuisance functions.

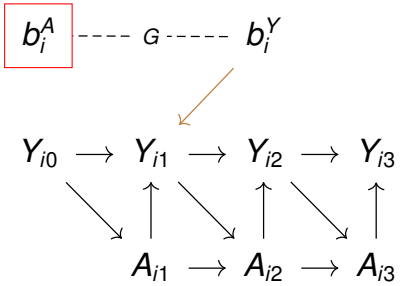
Acknowledgment

We thank Xiaoxuan Cai, Kelsey Alexovitz, Ji Soo Kim, Emily Scott, Bonnie Smith, Emily Scott, Ami Shah, Laura Hummers, Aki Nishimura, Becky Genberg, Rami Kantor, Ann Mwangi, Yuxin Zhu, Wei Jin, Victor Omodi for their great help and suggestions.

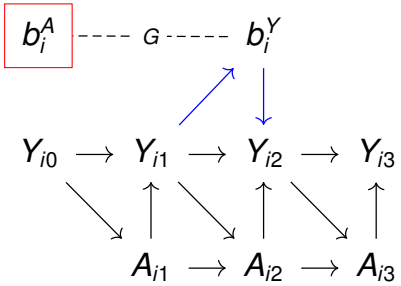
yxu143@jhu.edu
https://yizhenxu.github.io/

Thank you!

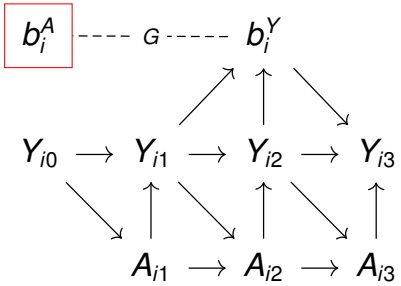
Information Flow Diagram in G-computation



Information Flow Diagram in G-computation



Information Flow Diagram in G-computation



Proposal

$$
\begin{aligned}
g(\bar{a}_2 | V_i, b_i^A) &= \mathbb{E}(Y_{i2} | V_i, \overline{A}_{i2} = \bar{a}_2, b_i^A) \\
&= \iint \{\mathbb{E}(Y_{i2} | V_i, \overline{Y}_{i1}, \overline{A}_{i2} = \bar{a}_2, b_i^A) \\
&\qquad \times P(Y_{i1} | V_i, Y_{i0}, A_{i1} = a_1, b_i^A)\} \; P(Y_{i0}) \, dY_{i0} \, dY_{i1}
\end{aligned}
$$

$$
\begin{aligned}
g(\bar{a}_2 | V_i) &= \mathbb{E}(Y_{i2} | V_i, \overline{A}_{i2} = \bar{a}_2) \\
&= \int g(\bar{a}_2 | V_i, b_i^A) \; P(b_i^A | V_i) \, db_i^A
\end{aligned}
$$

Proposal

$$f(Y_{i1}|V_i, Y_{i0}, A_{i1} = a_1, b_i^A) = \int P(Y_{i1}|V_i, Y_{i0}, A_{i1} = a_1, b_i^A, b_i^Y = u_0) \underline{P(b_i^Y = u_0|b_i^A)}\, du_0$$

$$\mathbb{E}(Y_{i2}|V_i, \overline{Y}_{i1}, \overline{A}_{i2} = \overline{a}_2, b_i^A) = \int \mathbb{E}(Y_{i2}|V_i, \overline{Y}_{i1}, \overline{A}_{i2} = \overline{a}_2, b_i^A, b_i^Y = u_1) \underline{P(b_i^Y = u_1|V_i, \overline{Y}_{i1}, A_{i1} = a_1, b_i^A)}\, du_1$$

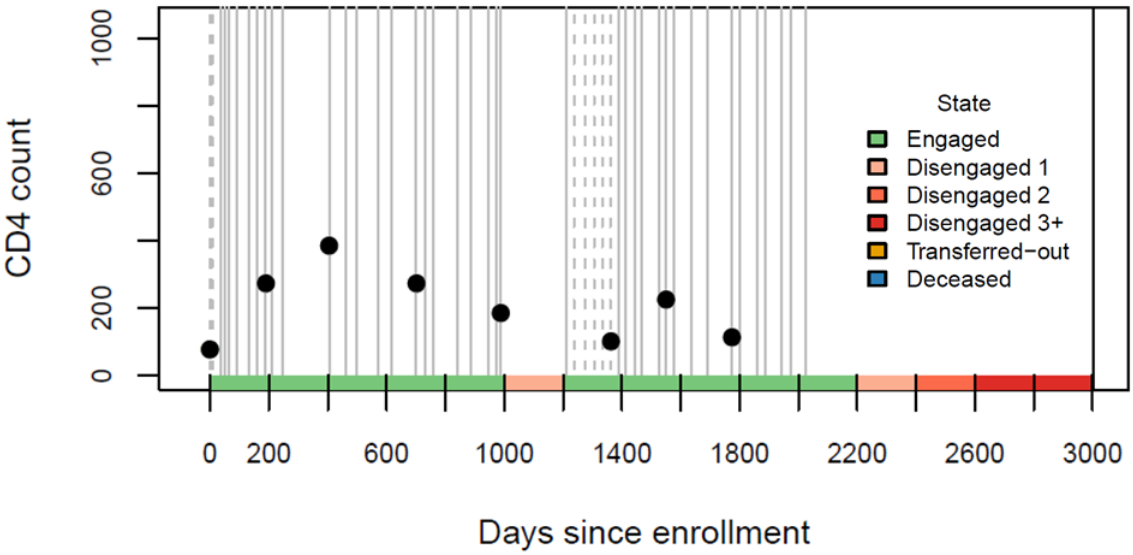
What's new about our method

- Existing work on longitudinal causal inference with mixed-effects models usually expresses a causal estimand as a function of the treatment sequence, covariates, and fixed effect coefficients
- Random effects are included in our calculation of the causal estimand to better address patient heterogeneity
- We propose to sequentially update subject-specific heterogeneity in biomarker dynamics as history information accumulates over time

What's new about our method

- A way to represent time-invariant unmeasured factors in Bayesian causal inference
- Degree of unmeasured confounding is encoded by covariances in the MGLMM
- Has the potential to be extended to guide the inclusion of latent variable models in Bayesian causal inference

Translating Patient-level Data into States



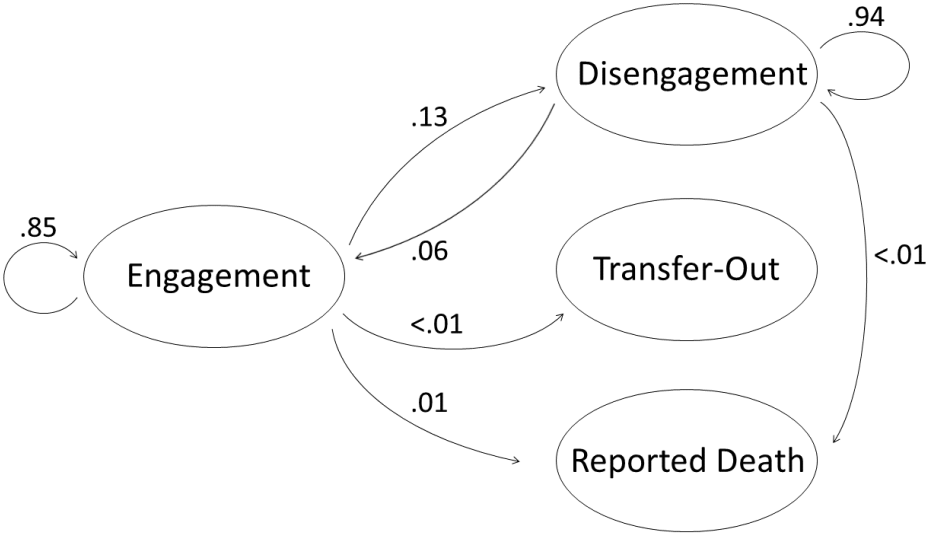
Engagement in Care

- **Retention** of patients in care in the months after treatment initiation may depend in part on the manner of treatment initiation
- Good retention and engagement in HIV care (Horstmann et al. 2010)
  - Promote adherence to medications – deter development of drug resistance, improve health outcomes, reduce health care costs
  - Encourage positive behavioral change – lowers HIV transmission rate
- START Trial (2015): **immediate initiation** of ART lowers the risk of developing serious illness or death

Motivation

- Primary question: Does immediate treatment result in better retention?
- Causal problem: What would have happened to patients retention and survival if the target population followed a certain regime over time?
- Regime: Static or dynamic
- Motivating application:
  - EHRs from AMPATH
  - Adults enrolled in HIV care at AMPATH (6/1/2008 - 8/23/2016)
  - 1,687,415 records from 96,045 patients before data cleaning
  - **Number of records ranges from 1 to 115. Timing of HIV care visits is highly irregular**

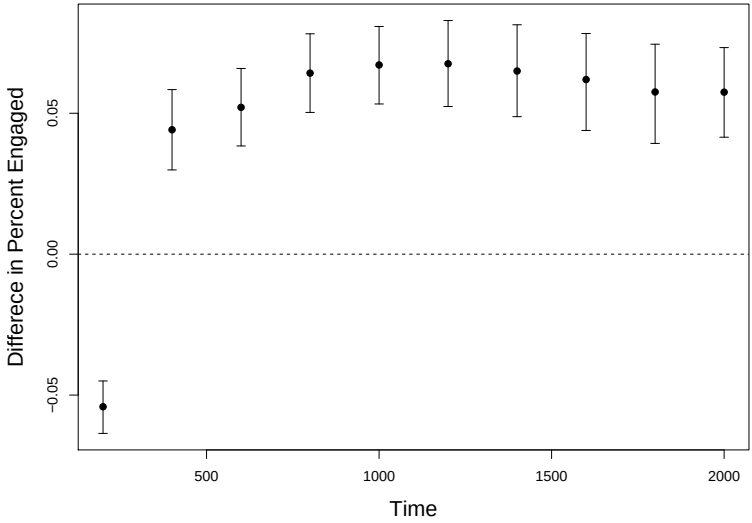
Transition Rates



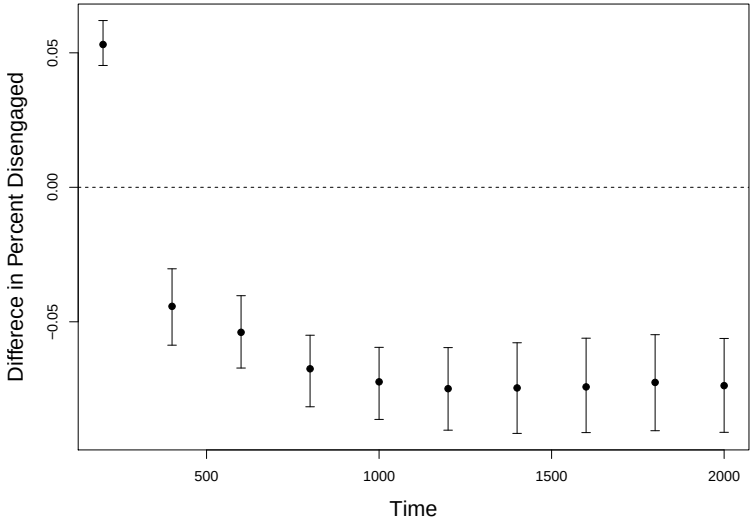
Causal Evaluation – The Big Picture

- Causal inference using Bayesian g-computation algorithm (GCA)
  - Require correct specification of predictive models for consistent estimations
  - Use Bayesian additive regression trees (BART)
- Challenge: fitting multinomial probit BART (MPBART) for outcome models
  - The four categories are inter-related
  - Transfer-out and reported death are extremely sparse

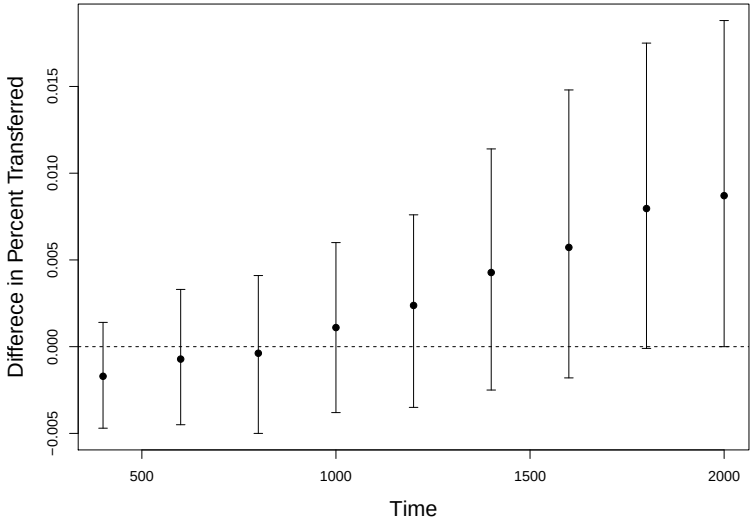
Inference about Causal Effectiveness



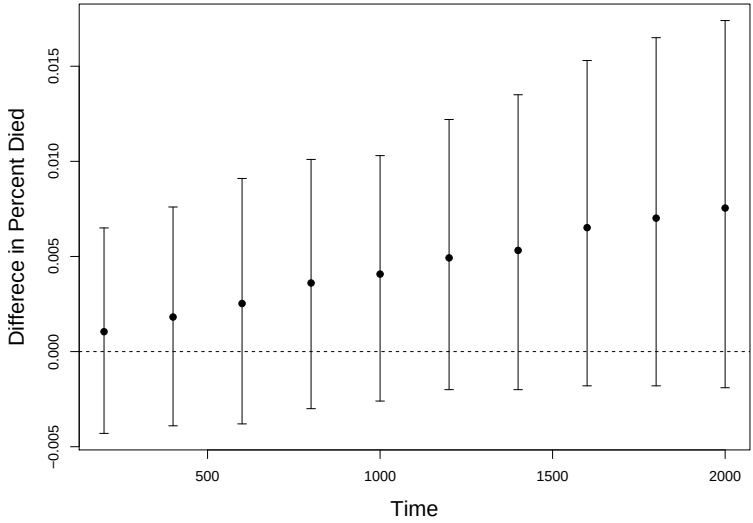
Inference about Causal Effectiveness



Inference about Causal Effectiveness



Inference about Causal Effectiveness



GCA using Linear Models

Covariate effects for transition from 'engaged' at $j = 3$ (day 600)

State at t_{j-1}	Engaged		
State at t_j	Disengaged	Transfer	Death
age	–0.02*	–0.01*	0.01*
male	0.18*	–0.05	0.10
Enrollment Year	0.011	–0.04*	–0.06*
TravelTime	–0.01	0.01	–0.04
WHO stage	0.05*	0.06	0.09*
Married	–0.15*	–0.08	–0.16*
Height	–0.002	0.00	0.00
log Weight	–0.26*	–0.13	–0.29*
undetectable VL	–0.62	–0.05	–6.21
Previous ARV	–0.38*	0.21*	–0.12
CD4 Update	–2.20*	–1.49*	–0.52*
latest log CD4+1	–0.20*	–0.13*	–0.31*

Augmentation Samplers for MPBART

Motivation

Accurate predictive modeling of patient engagement in HIV care, while accounting for death and transfer out of care as competing endpoints.

Challenge

- Collapsing augmentation strategy needs simple sampling steps and a reasonably fast converging Markov chain
- Complexity of stochastic search for posterior trees may undermine the procedure

Contribution

An example of a complicated Gibbs sampler that is sensitive to parameter expansion and needs special considerations in its data augmentation schemes.