Bayesian Nonparametrics, Missing Data, and Causal Inference

M. Daniels

University of Florida

Bayes Biostat 2018, Cambridge (UK)

Problems for causal inference and missing data

- Always untestable (from the observed data) assumptions needed for drawing inference in the presence of missing data and for causal inference
- Bias from mis-specified (parametric) models when use Bayesian inference
- we will discuss each next

How to specify uncheckable assumptions I

- assumptions need to be reasonable and carefully chosen for the specific application
- but since uncheckable from observed data: how address deviations/uncertainty?

How to specify uncheckable assumptions II

- strategy: parametrize assumptions via sensitivity parameters
  - sensitivity parameters should be easy to 'understand' in at least one (of two) ways
    1. so can elicit reasonable values based on expert opinion
    2. so can calibrate based on 'corresponding features' of the observed data
       - for 2., often can do via identifying restrictions - connect unidentified distributions, moments, etc. to 'corresponding' observed data 'equivalents'
  - informative prior specification for sensitivity parameters
- not uncommon to need assumptions for missing data AND causality in real applications

Models for observed data

- use Bayesian nonparametrics (BNP) to avoid bias from misspecified parametric assumptions, but not at loss of efficiency
- what BNP models to use?
  - can be factor of (causal) estimand of interest (specific estimand (e.g. mean) or not)
  - computational considerations
  - but if model entire distribution of potential outcomes 'carefully', allow any causal estimand
- fit of model to observed data should not be impacted by uncheckable assumptions (and sensitivity parameters)

Case studies

- use two case studies to illustrate approach
  - Case study I: EDP for comparative effectiveness in EHRs (with missing confounders)
  - Case study II: DDP-GP for causal inference for semi-competing risks

Case Study I: Notation

- A: treatment
- L: potential (pre-treatment) confounders
- Y: observed outcome
- Y^a: potential outcome if the subject had been assigned to treatment level a.

Causal effects

- average causal effect: $E[Y^1 - Y^0]$
- conditional average causal effect: $E[(Y^1 - Y^0)|V]$, $V \subset L$
- quantile causal effect: $F_1^{-1}(p) - F_0^{-1}(p)$

Uncheckable causal assumptions I

1. Consistency: $Y^a = Y$ among subjects with $A = a$, for all a.
   - implies that $p(Y^a|A = a, L) = p(Y|A = a, L)$.
   - allows us to estimate parameters in $Y^a|L$ model using the observed data
2. Positivity: $p(a|L) > 0$
   - each treatment level has non-zero probability for every confounder level.
3. Ignorability: $\{Y^a : \forall a \in \mathcal{A}\} \perp A|L\}$.
   - implies that $p(Y^a|A = a, L) = p(Y^a|A = a', L)$.
   - 'no unmeasured confounders' assumption.

Uncheckable causal assumptions II

- Three assumptions imply (this is an identifying restriction)

$$F(y^a|L) = F(y^a|A = a, L) = F(y|A = a, L)$$

- can introduce sensitivity parameter (to ignorability) here as

$$F(y^a|L) = F(y + \Delta * a|A = a, L)$$

  - Δ: average difference in Y^a (for $a = 0, 1$) among subjects assigned A $= 1$ compared with subjects assigned A $= 0$, who have the same covariates L
  - can assign an informative prior distribution for Δ (representing our expectation about unmeasured confounding, as well as our uncertainty about it)

Uncheckable causal assumptions III

- to calibrate Δ (using the observed data), we first calculate the total variance in Y explained by L (but not A); denote this by R^2
- then assume that $|\Delta| = |E(Y^a|A = 1, L) - E(Y^a|A = 0, L)| < \sqrt{var(Y)(1 - R^2)k}$
- unmeasured confounding would account for less than k% of the remaining variance; specify a prior distribution for k (sensitivity parameter)
- can do similar development with pseudo R^2 for non-continuous outcomes

Observed data models using BNP I

- model the joint distribution $p(Y, A, L)$
  - implicitly allows for ignorable missingness in L (uncheckable but not focus here)
- for simplicity, let $X_i = (A_i^T, L_i^T)^T$ so we model $p(Y, X)$ (or just condition on A)

Observed data models using BNP II

Model the joint distribution of (Y, X) using an enriched Dirichlet process (EDP) mixture (Wade et al, 2011, 2014)

$$\begin{aligned} Y_i|X_i, \theta_i &\sim p(y|x, \theta_i) \\ X_{i,r}|\omega_i &\sim p(x_r|\omega_i), \text{ indep} \\ (\theta_i, \omega_i)|P &\sim P \\ P &\sim EDP(\alpha_\theta, \alpha_\omega, P_0). \end{aligned} \tag{1}$$

The notation $P \sim EDP(\alpha_\theta, \alpha_\omega, P_0)$ means that $P_\theta \sim DP(\alpha_\theta, P_{0,\theta})$ and $P_{\omega|\theta} \sim DP(\alpha_\omega, P_{0,\omega|\theta})$ with base measure $P_0 = P_{0,\theta} \times P_{0,\omega|\theta}$.

Observed data models using BNP III

- each subject i has their own parameters θ_i and ω_i
  - However, because P is discrete, some clusters of subjects will have the same θ_i and ω_i
- The number of clusters depends on the concentration parameters α_θ and α_ω (low values indicate fewer clusters)
  - typical DP models have a single concentration parameter
- The *enrichment* of the usual DP is to have nested concentration parameters
  - allows for more x-clusters than y-clusters, which is important because the dimension of x will typically be much larger than that of y.
  - keeps cluster membership dependent on both $y|x$ and x through the nesting of the random partition.

Observed data models using BNP IV

We assume a local generalized linear model for $p(y|x, \theta_i)$,

$$p(y|x, \theta_i) = \exp\left\{ \frac{Y_i \eta_i - b(\eta_i)}{a(\phi_i)} + c(y_i, \phi_i) \right\}$$

where $g\{b'(\eta_i)\} = X_i \beta_i$ and $g\{\}$ is a link function.
If Y is binary

$$Y_i|X_i, \theta_i \sim \text{Bern}\{\text{logit}^{-1}(X\beta_i)\}$$

where $\theta_i = \beta_i$ and X is the design matrix involving A and L

Observed data models using BNP V

- assumes covariates **X** are locally independent.
  - That is given ω_i, covariates are independent. Two subjects in the same subcluster would have similar values of X.
- local independence assumption
  - makes it easy to include many continuous and discrete confounders, because the joint distribution is just a product of marginal distributions.
  - makes computations considerably faster because covariance matrices for the joint distribution of confounders are not needed.
- while assume that locally the generalized linear model is correctly specified for y and x and that the x's are independent from each other, globally all of the variables are dependent with potentially non-linear relationships.

Observed data models using BNP VI

- The conditional distribution implied by the joint model is $p(y|x) = \sum_{j=1}^{\infty} w_j(x) K(y|x, \theta_j)$, where

$$w_j(x) = \frac{\sum_{l=1}^{\infty} \gamma_{l|j} K(x|\omega_{l|j})}{\sum_{h=1}^{\infty} \gamma_h \sum_{l=1}^{\infty} \gamma_{l|h} K(x|\omega_{l|h})}.$$

  - Notice that the weights $w_j(x)$ depend on x. Therefore, even though $K(y|x, \theta_j)$ is a generalized linear model, $p(y|x)$ is a computationally tractable, flexible, non-linear, non-additive model.

Posterior computations

- Gibbs sampler for obtaining draws from the posterior distribution of the parameters (extension of Neal (2000) Algorithm 8 to accomodate nested clustering)
- Data augmentation - sample from conditional distribution of missing covariates at each iteration (valid under ignorable missingness)
- MC integration (over L) for each posterior draw to compute any functional of the distribution of the potential outcomes (can be done in parallel) - G computation step

$$p(y^a) = \int p(y|a, l)p(L)dL$$

HIV Data I

- Antiretroviral therapy (ART) is recommended for all human immunodeficiency virus (HIV) / chronic hepatitis C virus (HCV)-coinfected patients.
- ART regimens often include drugs from the nucleoside reverse transcriptase inhibitor (NRTI) class.
  - concern that some drugs in the NRTI class (didanosine, stavudine, zidovudine, and zalcitabine) might cause depletion of mitochondrial DNA, leading to liver injury.

HIV Data II

- We apply the proposed BNP approach to compare outcome Y (death within 2 years) among those prescribed mitochondrial toxic NRTI (mtNRTI)-containing ART regimen to those prescribed other NRTI-containing ART regimen.
- data from a study of HIV/HCV patients who newly initiated ART within the Veterans Aging Cohort Study (VACS)
- The study population included co-infected patients who newly initiated an ART-regimen that include NRTIs (either mtNRTIs or other NRTIs) from 2002 to 2009.
- total of $n = 1747$ patients included in the study

HIV Data III

- $A = 1$: initiating an ART regimen that included an mtNRTI; $A = 0$ initiating an ART regiment that included some other NRTI.
- outcome: all-cause mortality (focused on the event occurring within 2 years of ART initiation)
- There were 76 deaths out of 836 patients in the mtNRTI group, and 89 deaths out of 911 patients in the other NRTI group.
- causal parameter of interest is the relative risk:

$$\psi_{rr} = E(Y^1)/E(Y^0)$$

HIV Data IV

- Variables that were included in the model as confounders (L) included
  - baseline demographics and clinical variables: age at baseline (years), race/ethnicity, body mass index, diabetes mellitus, alcohol dependence/abuse, injection/non-injection drug abuse, year of ART initiation, and exposure to other antiretrovirals associated with hepatotoxicity (i.e., abacavir, nevirapine, saquinavir, tipranavir).
  - baseline laboratory variables: CD4 count, HIV RNA, alanine aminotransferase (ALT), aspartate aminotransferase (AST), and fibrosis-4 (FIB-4) score.

HIV Data V

- The percentage of missing data for each variable is as follows: ALT 1.3%, AST 2.5%, CD4 1.8%, FIB-4 3.1%.
- The percentage of patients with at least one missing variable is 4.8%.

Case Study I: Results I

- The posterior median and 95% CI of the average causal relative risk (RR), ψ_{rr}, were

$$1.16(0.87, 1.54)$$

  - 16% increased risk of death within 2 years comparing mtNRTI-containing ART regimens with other NRTI-containing ART regimens.

Case Study I: Conclusions I

- proposed a Bayesian nonparametric approach for causal inference (for large p) that can handle discrete or continuous outcomes and categorical treatment.
- simulations (now presented here) show overall good performance of the BNP approach compared with non-Bayesian and/or parametric approaches

Case Study I: Conclusions II

- While the full distribution of outcome, treatment, and confounders is modeled, the proposed BNP approach allows
  - for flexible modeling of these distributions
  - estimation of any functionals of the potential outcome distribution
  - high-dimensional confounding.
  - 'imputation' of missing covariates under ignorable missingness
  - uncertainty about uncheckable assumptions can be accounted for via informative priors on 'easy to calibrate' sensitivity parameter

Case study II: Semi-competing risks

- Semi-competing risks occur in studies where observation of a nonterminal event (e.g., progression) may be pre-empted by a terminal event (e.g., death), but not vice versa.
- In randomized clinical trials to evaluate treatments of life-threatening diseases, patients are often observed for specific types of disease progression and survival.
- Often, the primary outcome is patient survival, resulting in data analyses focusing on the terminal event using standard survival analysis tools
- However, there may also be interest in understanding the effect of treatment on nonterminal outcomes such as progression or readmission

Motivating Example

- randomized trial for the treatment of malignant brain tumors
  - one of the important progression endpoints is based on deterioration of the cerebellum
  - biologically plausible that a patient could die without cerebellar deterioration
  - thus, analyzing the effect of treatment on progression needs to account for the fact that progression is not well-defined after death.

Relevant literature

- Varadhan et al. (2014): nice review - classify models into two broad categories:
  - models for the distribution of the observable data, e.g., cause-specific hazards, sub-distribution functions (Fix and Neyman, 1951; Hougaard, 1999; Xu et al., 2010; Lee et al. 2015)
  - models for the distribution of the latent failure times (Robins, 1995; Lin et al., 1996; Wang, 2003; Peng and Fine, 2007; Chen, 2012; Hsieh and Huang, 2012)
- inference has focused on specific model parameters (e.g., regression coefficients).
- With the exception of Robins (1995ab) none of the approaches have discussed *causal interpretability* of the target parameters.

Approach to problem

- Here, interested in estimating the causal effect of treatment on the non-terminal endpoint from a randomized trial generating semi-competing risk data.
- Using the potential outcomes framework, we propose a principal stratification estimand (Frangakis and Rubin, 2002) to quantify the causal effect.
- Introduce assumptions that utilize baseline covariates to identify this estimand from the distribution of the observable data

Notation

- $z = 0, 1$ represents control and treatment group (used A in previous example; here Z since randomized)
- Y_P^z: progression time under treatment z.
- Y_D^z: death time under treatment z.
- C^z: censoring time under treatment z.
- Fundamental to our setting is that $Y_P^z \not> Y_D^z$ (i.e., progression cannot happen after death).

Causal estimand

The causal estimand of interest:

$$\tau(u) = \frac{Pr[Y_P^1 < u \mid Y_D^0 \geq u, Y_D^1 \geq u]}{Pr[Y_P^0 < u \mid Y_D^0 \geq u, Y_D^1 \geq u]},$$

where $\tau(\cdot)$ is a smooth function of u.

- Among patients who survive to time u under both treatments, this estimand contrasts the risk of progression prior to time u for treatment 1 relative to treatment 0.
- example of a principal stratum causal effect

Observed data

- Z denote treatment assignment
- $\boldsymbol{X}$ denote a vector of the baseline covariates (replaced $\boldsymbol{L}$, the confounders from previous case study)
- the observed event times and event indicators.
  - $Y_P = Y_P^Z$, $Y_D = Y_D^Z$ and $C = C^Z$.
  - $T_1 = Y_P \wedge Y_D \wedge C$,
  - $\delta = I(Y_P < Y_D \wedge C)$,
  - $T_2 = Y_D \wedge C$,
  - $\xi = I(Y_D < C)$
- The observed data for each patient are $\boldsymbol{O} = (T_1, T_2, \delta, \xi, Z, \boldsymbol{X})$.

Assumption 1

Assumption 1: Treatment is randomized, i.e.,

$$Z \perp (Y_P^z, Y_D^z, C^z, \boldsymbol{X}); \quad z = 0, 1,$$

and $0 < Pr[Z = 1] < 1$.
This holds by design in randomized trials

Assumption 2

Assumption 2: Censoring is non-informative in the sense that

$$C^z \perp (Y_P^z, Y_D^z) \,|\, \boldsymbol{X} = \boldsymbol{x}; \quad z = 0, 1,$$

and $Pr[C^z > Y_P^z, C^z > Y_D^z | \boldsymbol{X} = \boldsymbol{x}] > 0$ for all $\boldsymbol{x}$.

Identification Results 1

Let $\lambda^z_{\boldsymbol{X}}$ and $G^z_{\boldsymbol{X}}$ denote the conditional hazard function and conditional distribution function of Y^z_D given $\boldsymbol{X} = \boldsymbol{x}$, respectively. Under Assumptions 1 and 2, $\lambda^z_{\boldsymbol{X}}$ and $G^z_{\boldsymbol{X}}$ are identified

$$\lambda^z_{\boldsymbol{X}}(t) = \lim_{dt \to 0} \left\{ \frac{Pr[t \leq T_2 < t + dt, \xi = 1 \,|\, T_2 \geq t, \boldsymbol{X} = \boldsymbol{x}, Z = z]}{dt} \right\}$$

and

$$G^z_{\boldsymbol{X}}(t) = 1 - \exp\left\{ - \int_0^t \lambda^z_{\boldsymbol{X}}(t) dt \right\}.$$

Identification Results 2

The conditional sub-distribution function of Y_P^z given Y_D^z and $\boldsymbol{X} = \boldsymbol{x}$, $V_{\boldsymbol{X}}^z$, is identified

$$V_{\boldsymbol{X}}^z(s|t) \quad = \quad Pr[T_1 \leq s, \delta = 1 \,|\, T_2 = t, \xi = 1, \boldsymbol{X} = \boldsymbol{x}, Z = z],$$

where $s \leq t$.

- Together $G_{\boldsymbol{X}}^z(t)$ and $V_{\boldsymbol{X}}^z(s|t)$ identify the joint subdistribution $V_{\boldsymbol{X}}^z(s,t)$ for (Y_P^z, Y_D^z) given $\boldsymbol{X} = \boldsymbol{x}$.

Assumption 3

Assumption 3: The conditional joint distribution function of (Y_D^0, Y_D^1) given $\boldsymbol{X} = \boldsymbol{x}$, $G_{\boldsymbol{X}}$, follows a Gaussian copula model, i.e.,

$$G_{\boldsymbol{X}}(v, w; \rho) = \Phi_{2,\rho}[\Phi^{-1}\{G^0_{\boldsymbol{X}}(v)\}, \Phi^{-1}\{G^1_{\boldsymbol{X}}(w)\}],$$

where Φ is a standard normal c.d.f. and $\Phi_{2,\rho}$ is a bivariate normal c.d.f. with mean 0, marginal variances 1, and correlation ρ.

- ρ is a sensitivity parameter
- easy to interpret and bounded in $[-1, 1]$
- for fixed ρ, $G_{\boldsymbol{X}}$ is identified since $G^0_{\boldsymbol{X}}$ and $G^1_{\boldsymbol{X}}$ are identified
- similar assumptions have been used in the causal mediation literature (Daniels et al. 2012)

Assumption 4

Assumption 4: Progression time under treatment z is conditionally independent of death time under treatment $1-z$ given death time under treatment z and covariates $\boldsymbol{X} = \boldsymbol{x}$, i.e.,

$$Y_P^z \perp Y_D^{1-z} \,\big|\, Y_D^z, \boldsymbol{X} = \boldsymbol{x}; \quad z = 0, 1.$$

Identification Results 3

Lemma: Under Assumptions 1-4, $\tau(\cdot)$ is identified from the distribution of the observed data as follows:

$$\tau(u) = \frac{\int_{\boldsymbol{X}} \int_{s<u} \int_{v \geq u} \int_{t \geq u} dV^1_{\boldsymbol{X}}(s|t) dG_{\boldsymbol{X}}(v,t) dK(\boldsymbol{x})}{\int_{\boldsymbol{X}} \int_{s<u} \int_{v \geq u} \int_{t \geq u} dV^0_{\boldsymbol{X}}(s|t) dG_{\boldsymbol{X}}(v,t) dK(\boldsymbol{x})},$$

where $K(\boldsymbol{x})$ is the empirical distribution of $\boldsymbol{X}$.

BNP model for the observed data distribution

- specify independent Dependent Dirichlet Process-Gaussian Process prior (DDP-GP) for each treatment group z, on the unknown conditional (on $\boldsymbol{X} = \boldsymbol{x}$) probability measure ($H^z_{\boldsymbol{X}}$) of (Y^z_P, Y^z_D).
- Since $V^z_{\boldsymbol{X}}(s|t) = H^z_{\boldsymbol{X}}(Y^z_P \leq s, Y^z_P \leq Y^z_D | Y^z_D = t)$ $(s \leq t)$ and $G^z_{\boldsymbol{X}}(t) = H^z_{\boldsymbol{X}}(Y^z_D \leq t)$,
  - the prior on $H^z_{\boldsymbol{X}}$ induces priors on $V^z_{\boldsymbol{X}}(s|t)$ and $G^z_{\boldsymbol{X}}(t)$ (identified under Assumptions 1 and 2)
  - together with the Gaussian copula for $G_{\boldsymbol{X}}$ implies a prior on the estimand $\tau(\cdot)$.
  - the prior on $H^z_{\boldsymbol{X}}$ also induces priors on non-identified quantities (i.e., progression after death) which have no impact on our analysis.

More on DDP-GP I

- prior on the conditional (on covariates $\boldsymbol{X} = \boldsymbol{x}$) distribution, $H_{\boldsymbol{x}}$
- use a Dependent Dirichlet process (DDP) mixture of normals,

$$dH_{\boldsymbol{x}}(\boldsymbol{v}) = \sum_h w_h \phi\left(\boldsymbol{v}; \boldsymbol{\theta}_h(\boldsymbol{x}), \boldsymbol{\Sigma}\right) d\boldsymbol{v}.$$

  where $w_h = \nu_h \prod_{l<h}(1-\nu_l)$ with $\nu_h \sim \text{Beta}(1, \alpha)$
- what about $\{\boldsymbol{\theta}_h(\boldsymbol{x}) : \boldsymbol{x}\}$?

More on DDP-GP II

- GP prior: $\{\theta_{hj}(\boldsymbol{x}) : \boldsymbol{x}\} \sim GP(\mu_{hj}(\cdot), R_j(\cdot,\cdot))$
  - model the mean function $\mu_{hj}(\cdot)$ as a linear regression on covariates $\mu_{hj}(\boldsymbol{x}_l; \boldsymbol{\beta}_{hj}) = \boldsymbol{x}_l \boldsymbol{\beta}_{hj}$, with covariance process specified as

$$R_j(\boldsymbol{x}_l, \boldsymbol{x}_{l'}) = \exp\left\{-\sum_{d=1}^{D}(x_{ld} - x_{l'd})^2\right\} + \delta_{ll'}\epsilon^2,$$

    where D is the dimension of the covariate vector, $\delta_{ll'} = I(l = l')$ and ϵ is a small constant used to ensure that the covariance function is positive definite.
  - continuous covariates standardized
- final model/prior: $\{H_{\boldsymbol{X}}\} \sim DDPGP(\alpha, \boldsymbol{\Sigma}, GP(\mu_j(\cdot), R_j(\cdot,\cdot)), j = 1, \ldots, J)$.

Brain Cancer Data example I

- randomized (placebo-controlled) phase II trial (Brem et al, 1995)
- 222 recurrent gliomas patients, who were scheduled for tumor resection
- Eligible patients had a single focus of tumor in the cerebrum, had a Karnofsky score greater than 60, had completed radiation therapy, had not taken nitrosoureas within 6 weeks of enrollment, and had not had systematic chemotherapy within 4 weeks of enrollment.
- The data includes 11 baseline prognostic measures and a baseline evaluation of cerebellar function.

Brain Cancer Data example II

- Patient were randomized to receive surgically implanted biodegradable polymer discs *with or without* 3.85% of carmustine.
- The follow-up duration was 1 year.
- Of the 219 patients with complete baseline measures
  - 204 were observed to die
  - 100 were observed to progress prior to death
  - Of the 15 patients who did not die, 4 were observed to have cerebellar progression.
- **Goal:** estimate the causal effect of treatment on time to cerebellar progression.

Causal inference results I

- posterior inference for the causal estimand, $\tau(u)$.
- sensitivity parameter, ρ
  - fix ρ at 0.2, 0.5, and 0.8.
  - prior $\rho \sim \text{Beta}(0.1875, 0.0625)$ [mean and variance, 0.75 and 0.15]

Causal inference results II

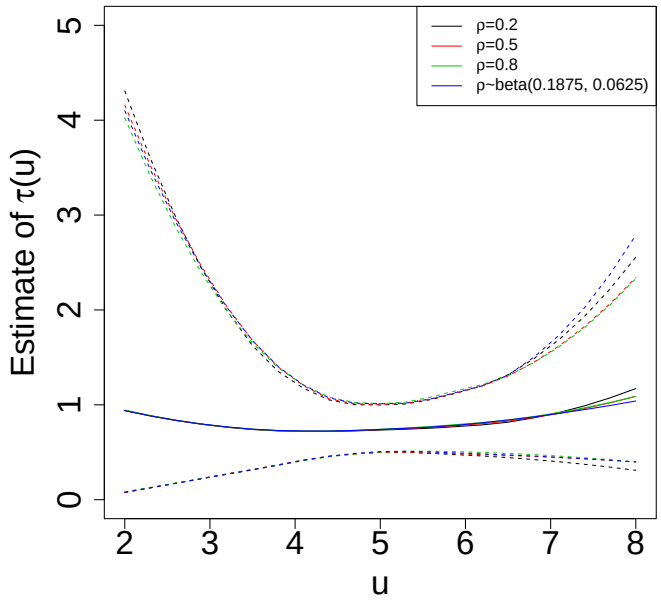

Figure: Posterior estimated $\tau(u)$ versus u for different ρ's. The solid lines represent the posterior estimated $\tau(u)$, and the dashed lines represent 95% (pointwise) credible intervals.

Causal inference results III

- Different values of ρ result in very similar posterior estimated $\tau(u)$
- When the survival time is no more than 55 days ($u \leq 4$), the posterior mean $\hat{\tau}(u) < 1$ and decreases as u increases
  - suggests that among patients who survive up to 55 days under both control and treatment, the risk of progression prior to time u is higher for the control relative to the treatment
  - however, the 95% credible interval covers 1.

Causal inference results IV

- For those that would survive under both arms beyond 55 days
  - the relative risk of progression for treatment versus control approaches and then exceeds one around 1800 days ($u = 7.5$), indicating a negative effect of the treatment on progression.
  - however, again the 95% credible interval covers 1.

Conclusions

- proposed a Bayesian approach for causal inference in setting of semi-competing risks
  - BNP for the observed data distribution
  - an interpretable causal estimand
  - one of uncheckable assumptions parameterized by a (easy to interpret) sensitivity parameter

Ongoing work

- Case study I:
  - EDP approach allows α_ω to be a function of θ
    - In our analyses we only included a single α_ω parameter
    - explore more complex models
  - extension to the time-varying confounding setting
  - extension to settings with many covariates that are not actually confounders: explore zero-inflated or shrinkage priors for the coefficients in the BNP model
- Case Study II
  - how to best determine values of the sensitivity parameter, ρ
  - weaken/remove Assumption 4

Other missing data and causal settings where use this approach

- Nonignorable missingness (monotone and non-monotone) using DPMs; Linero and D, 2015; Linero and D, 2017; Linero, 2017
- Nonignorable missingness with auxiliary covariates using DDP-GP; Zhou, D. and Mueller, 2018
- Comparative effectiveness in EHRs (using GP, EDP, and BART & DPM); Roy, et al., D 2016; **Roy, D, et al., 2017**; Xu, D. et al., 2017
- Causal inference for semi-competing risks using DDP-GP; **Xu, Mueller, Scharfstein, and D. (2018)**
- Causal inference with time-varying exposure and confounders with nonignorable missingness using BART; Josefsson and D. 2018