BART For Causal Inference

Antonio R. Linero

Learning Objectives

- What are the benefits of framing questions causally?
- What are the traps of applying machine learning to causal inference?
  - Prior Dogmatism
  - Treatment Effect Heterogeneity Priors
- What are BCFs and why do we parameterize them the way we do?
- How do we summarize the posterior?
- What is the workflow for applying BART to causal inference?

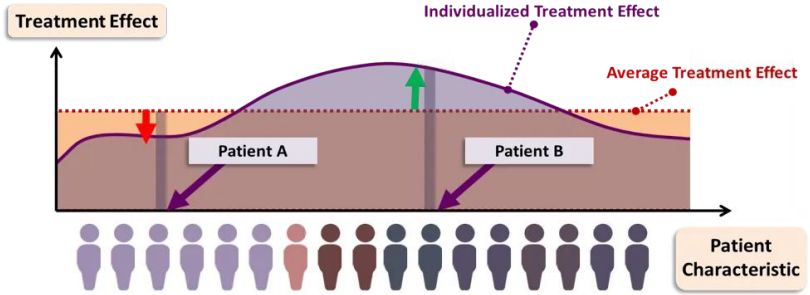
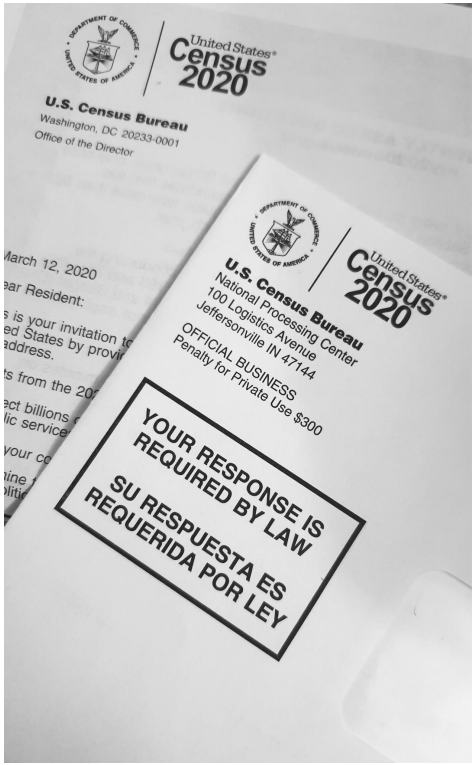


Image from https://www.vanderschaar-lab.com/individualized-treatment-effect-inference/

What is Causal Inference

Types of Studies

RANDOMIZED



NOT RANDOMIZED

Why Do Causal Inference?

- **Forces us to ask good/clear questions**
  - Causal questions are *model free*
  - Causal questions can directly lead to policy recommendations, or actions

- **Forces us to clarify our assumptions**
  - Encourages transparency
  - Forces us to think carefully about identification, sensitivity analysis, and being explicit about confounders

Why Focus on Treatment Effect Heterogeneity

- **In RCTs**: Not all treatments are useful for all people, and we might want to target treatments to individuals who will benefit the most.

- **In Observational Studies:** Can motivate different/better policies, but also **causal effects in observational studies tend to be more robust when they are large!** So finding treatments highly effective in subgroups can motivate better follow up studies and more robust conclusions.

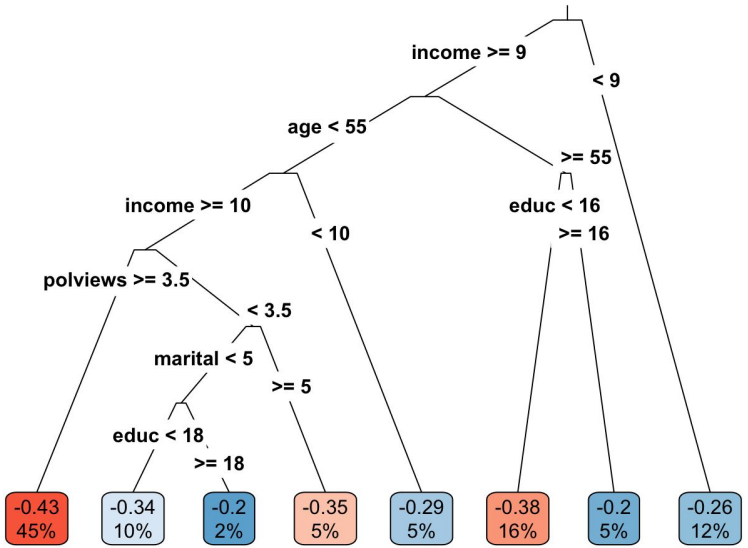
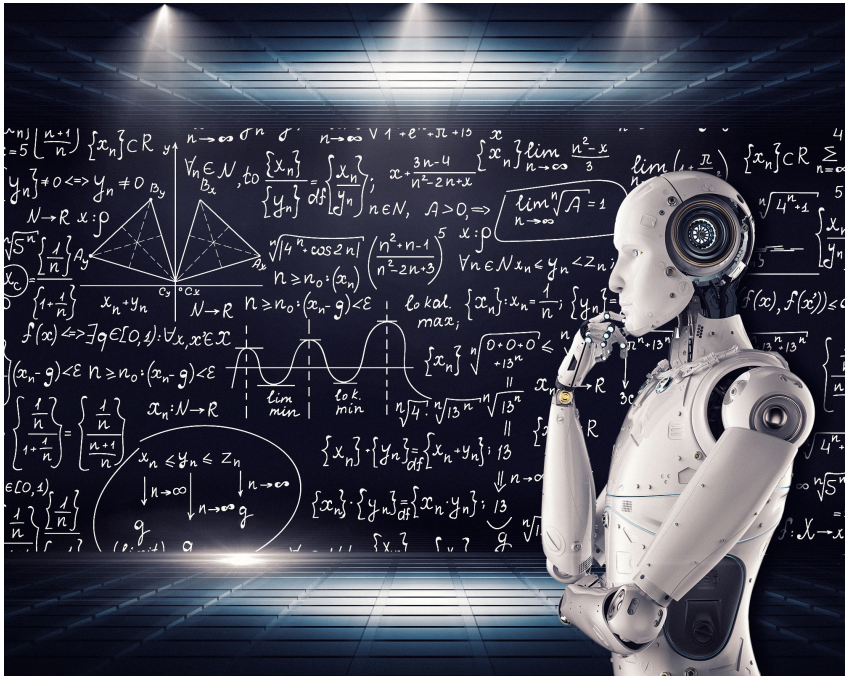


Image from
https://bookdown.org/stanfordgsbsilab/ml-ci-tutorial/hte-i-binary-treatment.html

Why Machine Learning?



- Reduces risks of model misspecification
- Avoids needing to specify interactions
- **But comes with its own catches due to regularization!**

Causal Framework

$A_i \sim \text{Bernoulli}\{e(X_i)\}$ Exposure Model

$\{Y_i(0), Y_i(1)\} \sim F_{X_i}$ Outcome Model

$\text{Observe } \{Y_i(A_i), A_i, X_i\}$ Always missing one…

Our Goal: Estimate the Treatment Effect

$$\tau(x) = \mathbb{E}\{Y_i(1) - Y_i(0) \mid X_i = x\}$$

Causal Assumptions

No Interference:

$$Y_i(\mathbf{a}) = Y_i(a_i)$$

Positivity:

$$\delta \leq e(x) \leq \delta^{-1} \qquad \text{for all } x$$

Unconfoundedness:

$$A_i \perp \{Y_i(0), Y_i(1)\} \mid X_i$$

Software Options

- **bcf** package on CRAN

- **SoftBart** package on CRAN (*softbart_vc_regression)*

First Gotcha: Prior Dogmatism

Ignorable Priors are Bad

The selection model is **ignorable** if (in addition to the Axioms)

$$\pi(d\mu, d\tau, de) = \pi(d\mu, d\tau)\,\pi(de)$$

People like ignorable priors because they avoid the need to model the propensity score.

In causal inference, methods that do not use the propensity score are sometimes called *outcome regression* approaches.

Evidence: Ignorable Priors are Bad (Dogmatism)

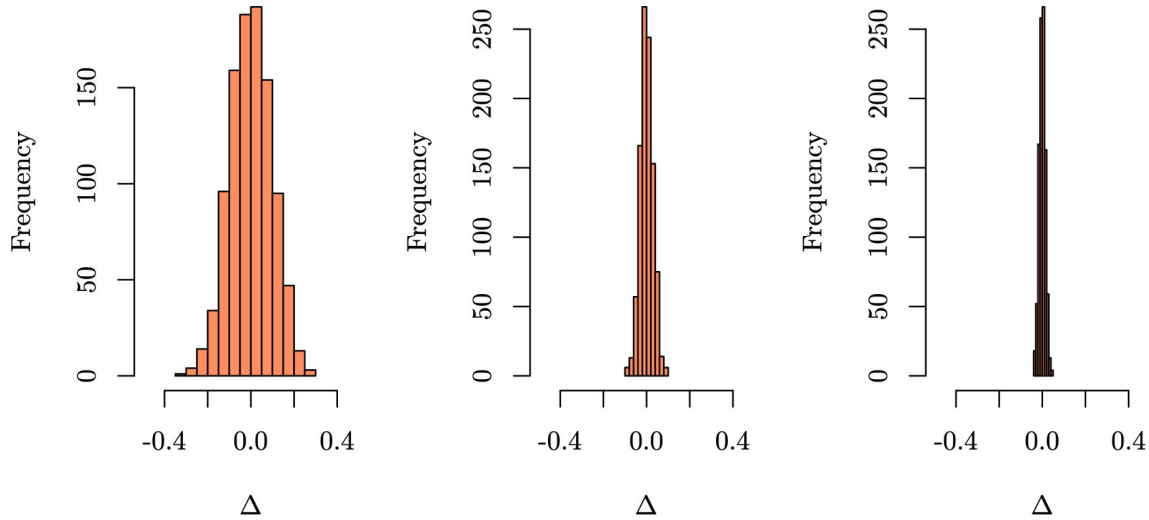


Figure 2: Prior distribution of Δ for the BART model in Section 2.2 for $P \in \{1, 10, 50\}$.

Linero (2024)

$$\Delta = \text{amount of selection bias}$$

Propensity score controls is essential *subjectively*

Second Gotcha: Heterogeneity Priors

What Not to Do: The T-Learner

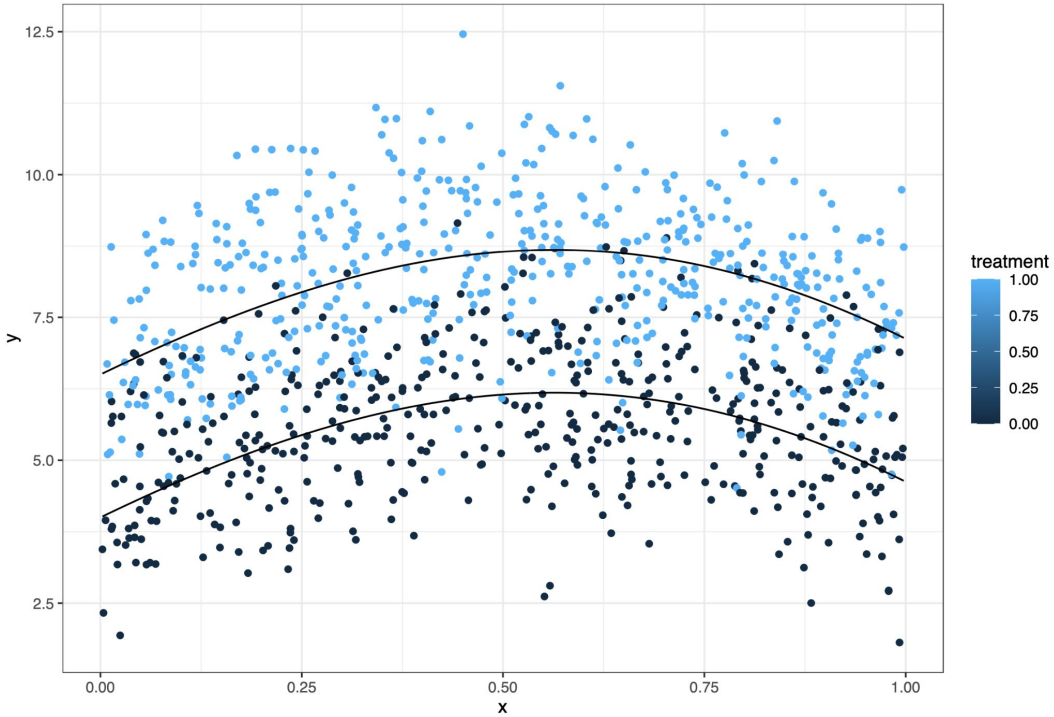
$$\mu_0(x) \sim N(0, a)$$
$$\mu_1(x) \sim N(0, b)$$
$$\tau(x) = \mu_1(x) - \mu_0(x) \sim N(0, a + b)$$

Large treatment effects

$$\mathrm{Var}\{\tau(X)\} \asymp \frac{\mathrm{Var}\{\mu_1(X) + \mu_0(X)\}}{2}$$

Heterogeneity scales with complexity of prognostic effect

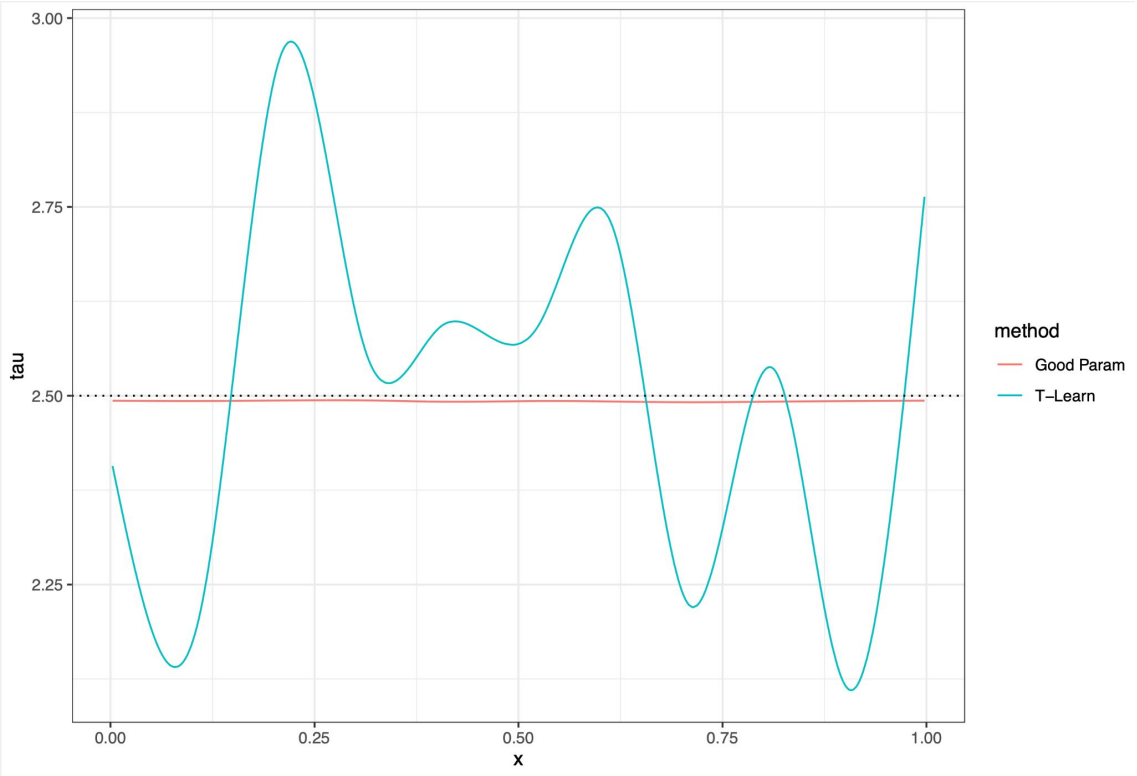
Bad Priors Make Bad Posteriors



Bad Priors Make Bad Posteriors

Simulation:

- Complicated prognostic
- Constant treatment effect
- Penalized splines used



Bayesian Causal Forests

What Should a Good HTE Prior Look Like?

In my areas of application, we expect $\tau(X_i)$'s to be:

1. Small (relative to $\mu(X_i)$ and $\text{Var}(Y_i)$)
2. Relatively homogeneous ($\text{Var}\{\tau(X_i)\}$ is small)
3. Likely associated with treatment choice.

A parameterization that makes sense here is something like:

$$\tau(x) = \tau_0 + \sigma_\tau \tau^\star(x)$$

with $\sigma_\tau \tau^\star(x) \ll \tau_0$ with high prior probability.

A Bayesian Causal Forest (Hahn et al. 2020)

$$Y_i(a) = \mu(X_i) + \{a - \hat{e}(X_i)\}\,\tau(X_i) + \epsilon_i(a)$$

Prognostic effect

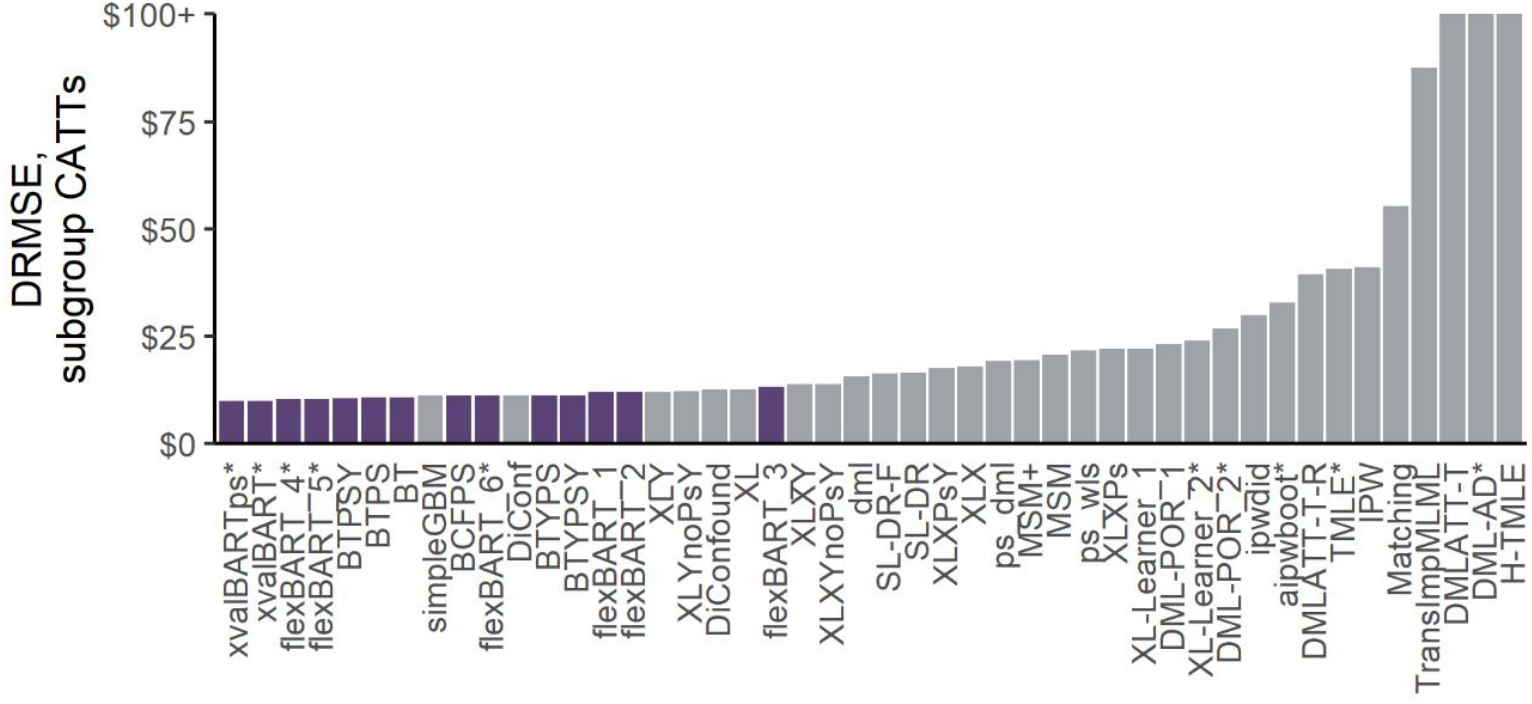
Prevents Dogmatism

Treatment effect

Put BARTS on all unknown functions!

(Can also include PS in prognostic effect)

BART-based methods led the pack, including **flexBART** from U Wisconsin



Heterogeneity Prior

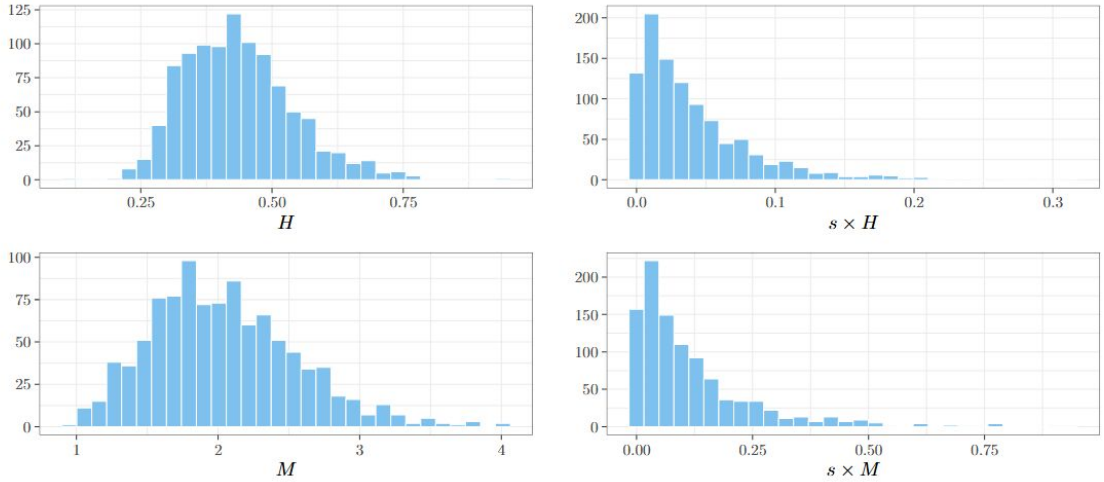


Figure 1: Prior distribution of the root mean squared heterogeneity H and maximal heterogeneity M , either with $\sigma_\tau = 1$ or $s_\tau = 0.1$ (denoted by $s \times H$ and $s \times M$).

Theorem 1. *For the BART prior described in the Supplementary Material, we have* $\mathbb{E}(H^2) = \sigma_\tau^2(1 - e^{-\lambda/3})$ *where* λ *is the average depth of a given leaf node under the prior.*

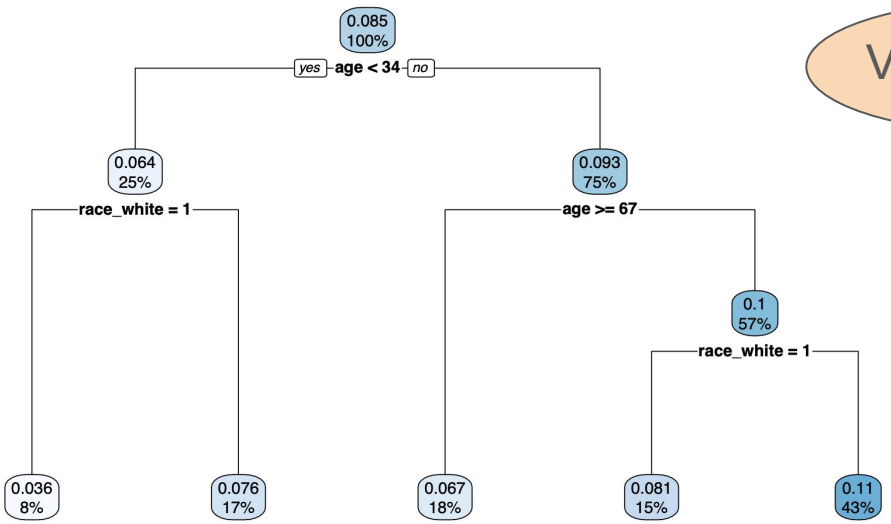
Posterior Summarization

Question: How do I extract scientific insight from the models?

Possible Goals:

- Subgroup identification (who benefited most from treatment?)
- Find an interpretable surrogate for $\tau(x)$

Subgroup Identification



Virtual Twins

Figure 3: Posterior summarization of the indirect effect using a single regression tree.

Best Tree-Based Approximation to HTE

Projection Summary

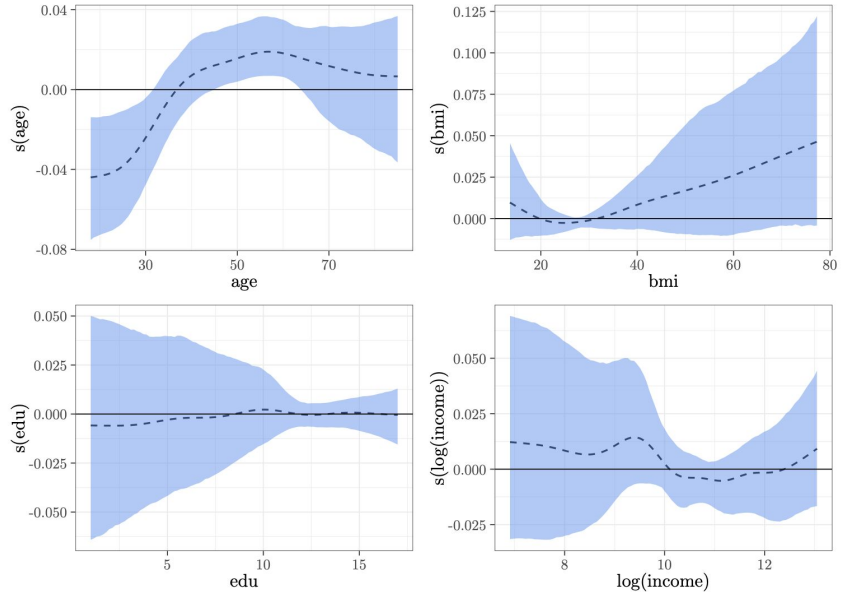
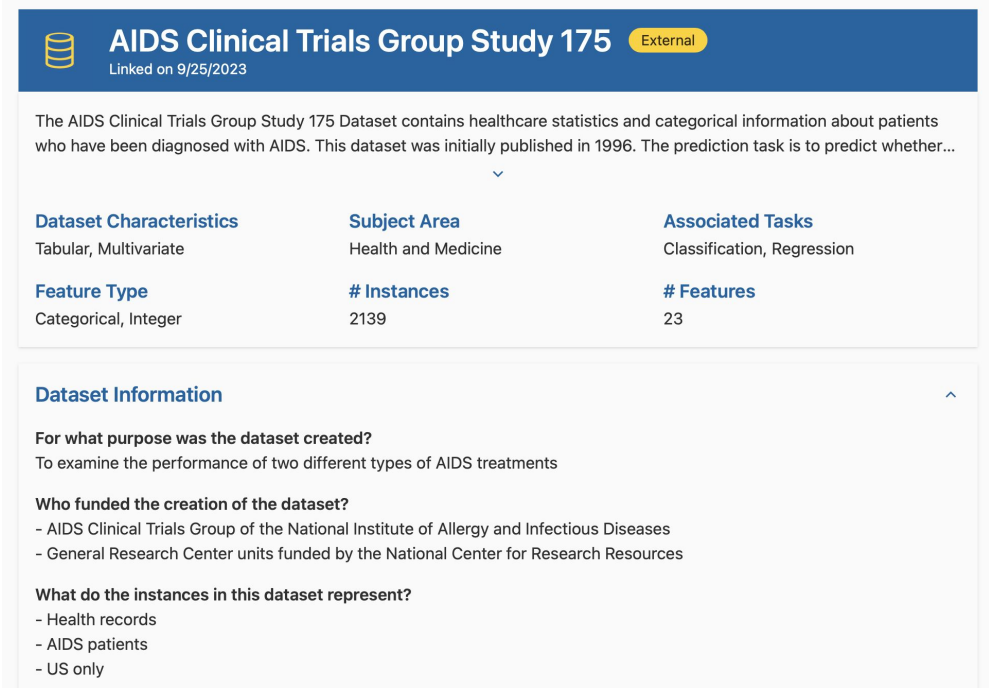


Figure 5: Posterior summarization of the indirect effect using a GAM for the continuous variables. The projection of the posterior mean is given by the dashed line while the shaded area gives a posterior 95% credible band of the projection.

Case Study: an RCT

Question: is treatment more effective at reducing CD4 count change in certain subpopulations?



AIDS Clinical Trials Group Study 175

External

Linked on 9/25/2023

The AIDS Clinical Trials Group Study 175 Dataset contains healthcare statistics and categorical information about patients who have been diagnosed with AIDS. This dataset was initially published in 1996. The prediction task is to predict whether...

Dataset Characteristics	Subject Area	Associated Tasks
Tabular, Multivariate	Health and Medicine	Classification, Regression

Feature Type	# Instances	# Features
Categorical, Integer	2139	23

Dataset Information

For what purpose was the dataset created?
To examine the performance of two different types of AIDS treatments

Who funded the creation of the dataset?
- AIDS Clinical Trials Group of the National Institute of Allergy and Infectious Diseases
- General Research Center units funded by the National Center for Research Resources

What do the instances in this dataset represent?
- Health records
- AIDS patients
- US only

Go Over Code