A Bayesian Critique of a Rank-Based Method for Surrogate Marker Evaluation

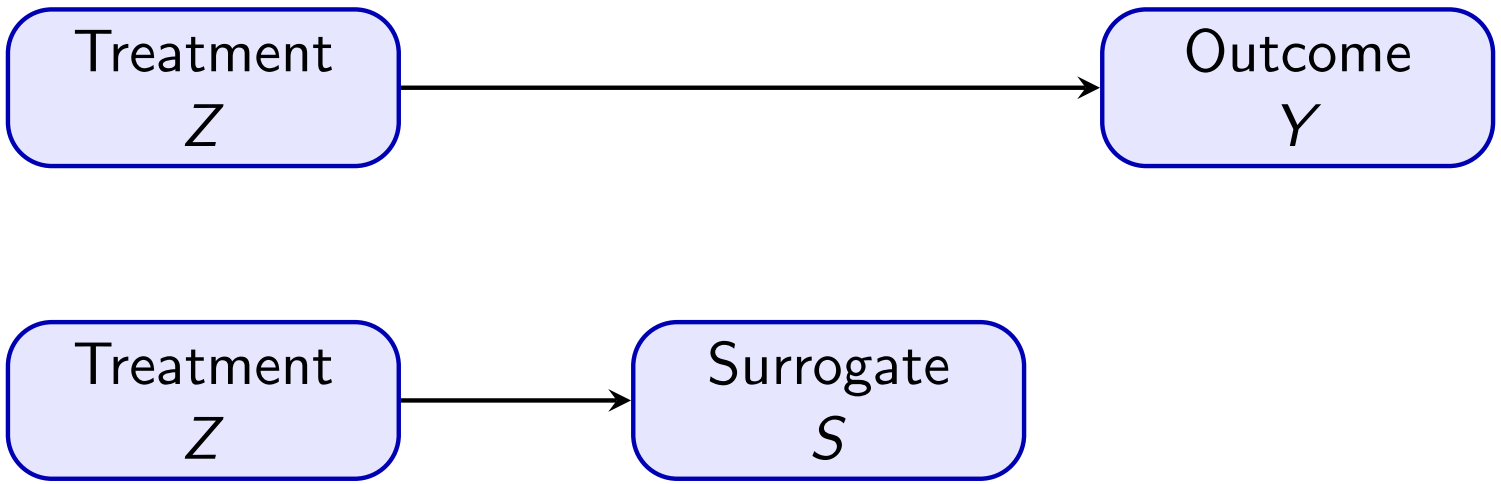
Pietro Carlotti Layla Parast

Department of Statistics and Data Sciences
University of Texas at Austin

Supported by R01DK118354

Surrogate Markers

Surrogate markers are often employed in **clinical trials** to replace primary outcomes when direct measurement is too difficult, expensive, or slow.



A Rank-Based Approach

Parast et al. (2024) recently proposed a **rank-based** approach for surrogate evaluation, which is fully **nonparametric** and **assumption-lean**.

While this approach may be useful in small-sample, model-agnostic settings because of its minimal assumptions, it still has some notable limitations:

- Limited **robustness** across heterogeneous data-generating processes.
- Reduced statistical efficiency, leading to lower **power**.
- Absence of a transparent **causal** interpretation.

Notation

We define the following variables:

- Y: Primary outcome
- S: Surrogate marker
- Z: Treatment indicator

Following Rubin (1974), we define the vector of potential outcomes as:

$$P = (P_1, P_0) = (Y_1, S_1, Y_0, S_0), \tag{1}$$

where the outcomes under treatment and control are

$$P_1 = (Y_1, S_1) \quad \text{and} \quad P_0 = (Y_0, S_0). \tag{2}$$

Setting

To clarify variable dependencies, we adopt this mathematical framework:

Data Generating Process

We assume the following data generating process:

$$
\begin{aligned}
(Y_i, S_i) &= \begin{cases} P_{1i}, & \text{if } Z_i = 1 \\ P_{0i}, & \text{if } Z_i = 0 \end{cases}, \quad i = 1, \ldots, n, \\
P_i &\overset{\text{ind}}{\sim} F_i, \quad i = 1, \ldots, n, \\
Z_i &\overset{\text{iid}}{\sim} \text{Bernoulli}(p), \quad i = 1, \ldots, n,
\end{aligned}
\tag{3}
$$

where F_i is a distribution on $\mathbb{R}^4$ and $p \in (0, 1)$.

Rationale of the Rank-Based Approach

The main idea of Parast et al. (2024) is that if S is a valid surrogate for Y, then the treatment Z should exert a **comparable** effect on both.

Given a measure of discrepancy δ between the treatment effects on Y and S, surrogate validity is assessed by testing whether δ is sufficiently small.

Parast et al. (2024) propose to use the following hypothesis test:

$$\begin{cases} H_0 : \delta \geq \varepsilon \\ H_1 : \delta < \varepsilon \end{cases} \tag{4}$$

where ε is a pre-specified threshold for surrogate validation.

The Mann-Whitney Parameter

Parast et al. (2024) propose to quantify treatment effects using

$$U_Y = \mathbb{P}(Y_{1i} > Y_{0j}) \quad \text{and} \quad U_S = \mathbb{P}(S_{1i} > S_{0j}). \tag{5}$$

The discrepancy parameter between U_Y and U_S is then defined as

$$\delta = U_Y - U_S. \tag{6}$$

Limitation of the Existing Approach

However, the approach of Parast et al. (2024) has some limitations:

- The Mann–Whitney test assumes potential outcomes are **i.i.d.**:
  - Potential outcomes often depend on unit-specific **covariates**.
- The test suffers from low **power**:
  - **Parametric** assumptions can boost surrogate detection in small samples.
- **Rank-based** statistics lack a clear causal interpretation:
  - The Mann–Whitney test checks group-level differences in distributions, which may not reflect individual-level **causal** effects.

Alternative Discrepancy Parameters

The Mann–Whitney parameter is the probability of higher treated outcomes, thus it captures **global** distributional differences between treatment arms.

A natural choice would be to compare potential outcomes for the same unit:

$$V_Y = \mathbb{P}(Y_{1i} > Y_{0i}) \quad \text{and} \quad V_S = \mathbb{P}(S_{1i} > S_{0i}). \tag{10}$$

Then, the discrepancy parameter between V_Y and V_S can be defined as

$$\theta = V_Y - V_S. \tag{11}$$

Since δ and θ can differ even under well-behaved data-generating mechanisms, a careful choice of the parameter is crucial for surrogate evaluation.

Bayesian Causal Inference

The key idea of Bayesian causal inference is to treat the unobserved potential outcomes as **latent variables** and to use the observed data to **impute** them.

By combining observed and imputed outcomes, we obtain the posterior distribution of θ, which can be used to assess surrogate validity.

Bayesian Imputation-Based Approach

Algorithm Bayesian Imputation-Based Approach

Require: Observed data $\{Y_i, S_i, Z_i, X_i\}_{i=1}^{n}$, prior parameters, iterations T.

for $t = 1$ to T **do**
  for $i = 1$ to n **do**
    if $Z_i = 1$ **then**
      Sample $P_{0i}^{(t)} \sim f\left(P_{0i}^{(t)} \mid P_{1i}, X_i, Z_i = 1\right)$
    else
      Sample $P_{1i}^{(t)} \sim f\left(P_{1i}^{(t)} \mid P_{0i}, X_i, Z_i = 0\right)$
    end if
  end for
end for

Bayesian Posterior Probability Decision Rule

We consider this hypothesis test for surrogate validity:

$$\begin{cases} H_0 : \theta \geq \eta \\ H_1 : \theta < \eta \end{cases} \tag{12}$$

where η is a pre-specified validation threshold. After discarding the first b burn-in samples, we compute a one-sided **credible interval** for θ with

$$\hat{\theta}_{1-\alpha} = \text{quantile}\left(\theta^{(b+1:T)}; 1-\alpha\right). \tag{13}$$

We then adopt the following **decision rule**:

- **Reject** H_0 if $\hat{\theta}_{1-\alpha} < \eta$.
- **Fail to reject** H_0 otherwise.

Parametric Model

The main challenge of this method lies in sampling the unobserved potential outcomes, as this step depends on the assumed data-generating mechanism.

In this work we focus on two **parametric** models:

1. A model *without* covariates, aligning with Parast et al. (2024).
2. A model *with* covariates, exploiting the strengths of our framework.

Parametric Model without Covariates

The first model assumes Gaussian potential outcomes:

$$
\begin{aligned}
(Y_i, S_i) &= \begin{cases} P_{1i}, & \text{if } Z_i = 1 \\ P_{0i}, & \text{if } Z_i = 0 \end{cases}, \quad i = 1, \ldots, n, \\
P_i &\overset{\text{ind}}{\sim} \mathcal{N}_4(\mu, \Sigma), \quad i = 1, \ldots, n, \\
\mu &\sim \mathcal{N}_4(\mu_0, \Sigma_0), \\
\Sigma &= \text{diag}(\sigma_{1:4})\, \Omega\, \text{diag}(\sigma_{1:4}), \\
\sigma_k &\overset{\text{iid}}{\sim} \text{Half-Normal}(0, s), \quad k = 1, 2, 3, 4, \\
\Omega &\sim \text{LKJ}(\tau).
\end{aligned} \tag{14}
$$

where $\{Z_i\}_{i=1}^n$ are observed variables.

Parametric Model with Covariates

The second model includes covariates through a linear regression structure:

$$
\begin{aligned}
(Y_i, S_i) &= \begin{cases} P_{1i}, & \text{if } Z_i = 1 \\ P_{0i}, & \text{if } Z_i = 0 \end{cases}, \quad i = 1, \ldots, n, \\
P_i &\overset{\text{ind}}{\sim} \mathcal{N}_4(BX_i, \Sigma), \quad i = 1, \ldots, n, \\
B &= \left[\beta_1\ \beta_2\ \beta_3\ \beta_4\right]^\top, \\
\beta_k &\overset{\text{iid}}{\sim} \mathcal{N}_d(\mu_\beta, \Sigma_\beta), \quad k = 1, 2, 3, 4, \\
\Sigma &= \text{diag}(\sigma_{1:4})\, \Omega\, \text{diag}(\sigma_{1:4}), \\
\sigma_k &\overset{\text{iid}}{\sim} \text{Half-Normal}(0, s), \quad k = 1, 2, 3, 4, \\
\Omega &\sim \text{LKJ}(\tau),
\end{aligned}
\tag{15}
$$

where $\{X_i, Z_i\}_{i=1}^n$ are observed variables.

Simulation Study Designs

To evaluate the performance of our proposed Bayesian approach, we conduct simulations under two distinct experimental configurations:

- **Baseline Scenario:** Replication of the settings in Parast et al. (2024) without covariates to establish a benchmark for comparison.
- **Augmented Scenario:** An extended framework incorporating covariates to demonstrate the advantages of the proposed Bayesian approach.

Settings without Covariates

We first consider the same settings explored by Parast et al. (2024):

Setting	Coverage		Power	
	Frequentist	Bayesian	Frequentist	Bayesian
Useless Surrogate	0.98	**1**	0	0
Perfect Surrogate	0.98	**1**	1	1
Imperfect Surrogate	0.9	**0.98**	0.54	**0.64**
Misspecified Model	0.96	**1**	**0.56**	0.06

Table: Coverage and power for the frequentist rank-based method and Bayesian imputation-based method in the 4 settings explored by Parast et al. (2024).

Our Bayesian approach **outperforms** the frequentist method in coverage and power, except for the **misspecified** model, where power drops notably.

Setting with Covariates

We then consider a more complex setting with a binary covariate X:

$$
\begin{aligned}
P_i \mid X_i &\overset{\text{ind}}{\sim} \begin{cases} \mathcal{N}_4(\mu^0, \Sigma^0) & \text{if } X_i = 0 \\ \mathcal{N}_4(\mu^1, \Sigma^1) & \text{if } X_i = 1 \end{cases}, \quad i = 1, \ldots, n \\
X_i &\overset{\text{iid}}{\sim} \text{Bern}(\tfrac{1}{2}), \quad i = 1, \ldots, n,
\end{aligned}
\tag{16}
$$

where the Gaussian parameters are set so that S is a **perfect surrogate** for Y with **treatment effect homogeneity** across X, and yet δ is far from 0.

Setting with Covariates

We now consider the more complex setting with a binary covariate.

Setting	**Coverage**		**Power**	
	Frequentist	Bayesian	Frequentist	Bayesian
Binary Covariate	0.925	**0.975**	0.075	**1**

Table: Coverage and power for the frequentist rank-based method and the Bayesian imputation-based method in the setting with covariates.

It appears that our proposed Bayesian approach yields **superior power** than the frequentist method, while maintaining a **comparable coverage**.

Summary

A brief summary of what we saw today:

1. We introduced the rank-based approach of Parast et al. (2024).
2. We argued that the discrepancy parameter δ may fail to capture the causal relationship between Y and S even in simple settings.
3. We proposed a Bayesian approach to derive the posterior of θ.
4. We showed that our approach matches the performance of Parast et al. (2024) in baseline settings while significantly improving power in covariate-adjusted scenarios.

Thank you for your attention!

Questions?

Gaussian Potential Outcomes with Covariates

Simplifying assumptions help clarify where problems arise.

Gaussian Data Generating Process with Covariates

Consider Gaussian potential outcomes with parameters varying with X:

$$P_i \mid X_i \overset{\text{ind.}}{\sim} \mathcal{N}_4\left(\mu^{X_i}, \Sigma^{X_i}\right), \quad i = 1, \ldots, n, \tag{17}$$

$$X_i \overset{\text{iid}}{\sim} f_X, \quad i = 1, \ldots, n, \tag{18}$$

where μ^x, Σ^x are the parameters for $X = x$ and f_X is a distribution on $\mathcal{X}$.

Decomposition of δ

By the law of total probability, we can express U_Y as

$$U_Y = \int_{\mathcal{X}} \int_{\mathcal{X}} U_Y^{x_i x_j} \, f_X(x_i) \, f_X(x_j) \, dx_i \, dx_j, \tag{19}$$

where

$$U_Y^{x_i x_j} = \mathbb{P}(Y_{1i} > Y_{0j} \mid X_i = x_i, X_j = x_j). \tag{20}$$

The same decomposition applies to U_S, and therefore δ can be expressed as

$$\delta = \int_{\mathcal{X}} \int_{\mathcal{X}} \delta^{x_i x_j} \, f_X(x_i) \, f_X(x_j) \, dx_i \, dx_j, \tag{21}$$

where $\delta^{x_i x_j} = U_Y^{x_i x_j} - U_S^{x_i x_j}$.

Decomposition of θ

In contrast, by the law of total probability, we can express V_Y as

$$V_Y = \int_{\mathcal{X}} V_Y^x \, f_X(x) \, dx, \tag{22}$$

where

$$V_Y^x = \mathbb{P}(Y_{1i} > Y_{0i} \mid X_i = x). \tag{23}$$

The same decomposition applies to V_S, and therefore θ can be expressed:

$$\theta = \int_{\mathcal{X}} \theta^x \, f_X(x) \, dx, \tag{24}$$

where $\theta^x = V_Y^x - V_S^x$.

Analytical Expressions for $\delta^{x_i x_j}$ and θ^x

Gaussianity provides closed-form solutions for $\delta^{x_i x_j}$ and θ^x:

$$\delta^{x_i x_j} = \Phi\left(\frac{\mu_{Y_1}^{x_i} - \mu_{Y_0}^{x_j}}{\sqrt{\sigma_{Y_1}^{x_i} + \sigma_{Y_0}^{x_j}}}\right) - \Phi\left(\frac{\mu_{S_1}^{x_i} - \mu_{S_0}^{x_j}}{\sqrt{\sigma_{S_1}^{x_i} + \sigma_{S_0}^{x_j}}}\right), \tag{25}$$

$$\theta^x = \Phi\left(\frac{\mu_{Y_1}^{x} - \mu_{Y_0}^{x}}{\sqrt{\sigma_{Y_1}^{x} + \sigma_{Y_0}^{x} - 2\sigma_{Y_1 Y_0}^{x}}}\right) - \Phi\left(\frac{\mu_{S_1}^{x} - \mu_{S_0}^{x}}{\sqrt{\sigma_{S_1}^{x} + \sigma_{S_0}^{x} - 2\sigma_{S_1 S_0}^{x}}}\right). \tag{26}$$

Uncorrelated Potential Outcomes

To further simplify the comparison, let us assume a spherical covariance:

$$\Sigma^x = \tfrac{1}{2}\, I_4. \tag{27}$$

Then, the expressions above simplify to

$$\delta^{x_i x_j} = \Phi\left(\mu_{Y_1}^{x_i} - \mu_{Y_0}^{x_j}\right) - \Phi\left(\mu_{S_1}^{x_i} - \mu_{S_0}^{x_j}\right), \tag{28}$$

$$\theta^x = \Phi\left(\mu_{Y_1}^{x} - \mu_{Y_0}^{x}\right) - \Phi\left(\mu_{S_1}^{x} - \mu_{S_0}^{x}\right). \tag{29}$$

which implies that $\delta^{xx} = \theta^x$.

A Binary Covariate Example

To simplify even more, assume that is binary X with symmetric probabilities:

$$X \sim \text{Bernoulli}\left(\tfrac{1}{2}\right). \tag{30}$$

Then, we can express δ and θ as

$$\delta = \frac{\delta^{00} + \delta^{11} + \delta^{01} + \delta^{10}}{4}, \quad \theta = \frac{\theta^0 + \theta^1}{2}. \tag{31}$$

Therefore, we can compute the distance between δ and θ as

$$|\delta - \theta| = \frac{|\delta^{00} + \delta^{11} - \delta^{01} - \delta^{10}|}{4}. \tag{32}$$

Treatment Effect Homogeneity for a Perfect Surrogate

Given a perfect surrogate S and homogeneous treatment effects across X:

$$\mu^x_{Y_1} - \mu^x_{Y_0} = \mu^x_{S_1} - \mu^x_{S_0} = \Delta > 0 \quad \text{for } x = 0, 1. \tag{33}$$

Then it follows that

$$|\delta - \theta| = \frac{|\Phi(\Delta + d_Y) - \Phi(\Delta + d_S) + \Phi(\Delta - d_Y) - \Phi(\Delta - d_S)|}{4}, \tag{34}$$

where $d_Y = \mu^1_{Y_0} - \mu^0_{Y_0}$ and $d_S = \mu^1_{S_0} - \mu^0_{S_0}$.

Analityical Properties of $|\delta - \theta|$

This means that $|\theta - \delta|$ is very large whenever

$$|d_Y| < \Delta \text{ and } |d_S| > \Delta \quad \text{or} \quad |d_Y| > \Delta \text{ and } |d_S| < \Delta. \tag{35}$$

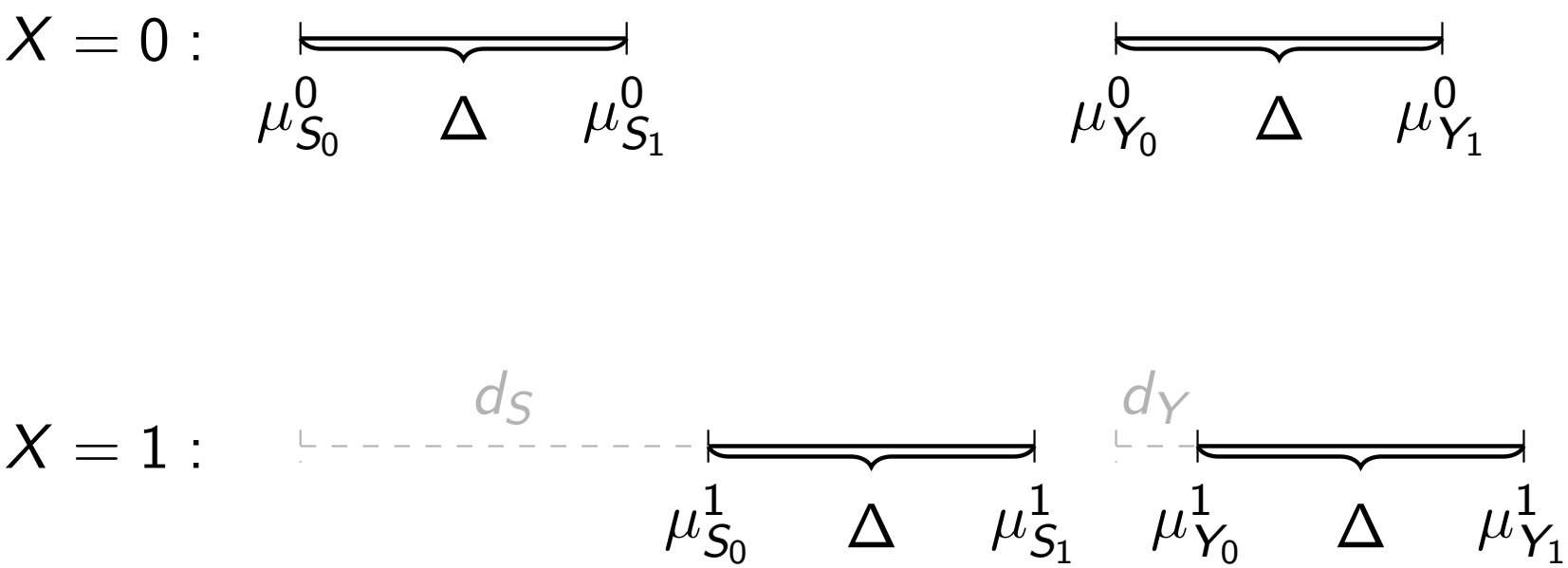
Moreover, we have that

$$\sup_{d_Y, d_S} \{|\theta - \delta|\} = \frac{1}{4}, \tag{36}$$

which is achieved when

$$d_Y = 0 \text{ and } |d_S| \to \infty \quad \text{or} \quad |d_Y| \to \infty \text{ and } d_S = 0. \tag{37}$$

A Visual Illustration



A Numerical Example

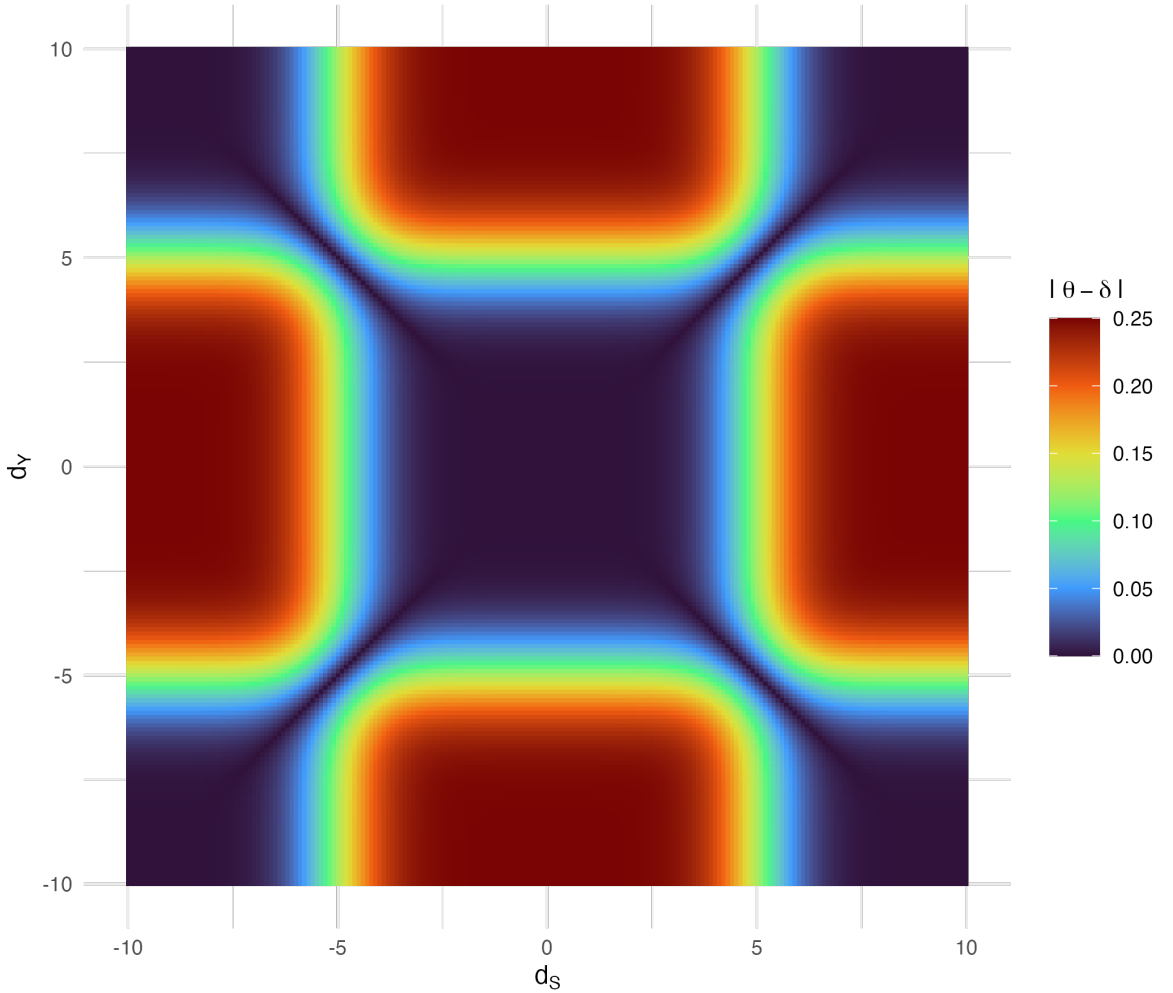


Figure: Scatter plot of the data.

Identifiability Issues

Both models suffer from an **identifiability** issue: some entries of Ω are unidentifiable because only one potential outcome is observed per unit:

$$\Omega = \begin{bmatrix} 1 & \rho_{Y_1 S_1} & \rho_{Y_1 Y_0} & \rho_{Y_1 S_0} \\ \rho_{Y_1 S_1} & 1 & \rho_{S_1 Y_0} & \rho_{S_1 S_0} \\ \rho_{Y_1 Y_0} & \rho_{S_1 Y_0} & 1 & \rho_{Y_0 S_0} \\ \rho_{Y_1 S_0} & \rho_{S_1 S_0} & \rho_{Y_0 S_0} & 1 \end{bmatrix}. \tag{38}$$

Green entries are identifiable from the data, but red entries, which directly enter the computation of θ, and orange entries, are both unidentifiable.

Asymptotic Behavior of the Posterior Distribution

Consider this well-known result by Schwartz (1965):

Schwartz Posterior Consistency Theorem

Consider the following set-up:

$$\begin{aligned} \textbf{True DGP:} \quad & D_i \overset{\text{iid}}{\sim} f_{\phi^*}, \quad i = 1, \ldots, n \\ \textbf{Bayesian Model:} \quad & D_i \mid \phi \overset{\text{iid}}{\sim} f_{\phi}, \quad i = 1, \ldots, n, \qquad \phi \sim \pi \end{aligned} \tag{39}$$

where f is a distribution parametrized by ϕ and π is a prior. Then, under mild regularity conditions, the posterior collapses asymptotically:

$$\pi_{D_{1:n}} \overset{a.s.}{\to} \delta_{\phi^*}, \tag{40}$$

where $\pi_{D_{1:n}}$ is the posterior distribution and δ_{ϕ^*} is a point mass at ϕ^*.

Asymptotic Behavior of Partially Identifiable Models

Let $\phi = (\phi_I, \phi_N)$, where ϕ_I denotes the identifiable components and ϕ_N the non-identifiable ones. Then, recall this result from Gustafson (2010):

Bayesian Posterior Consistency for Partially Identifiable Models

Consider the following set-up:

$$\begin{aligned} \textbf{True DGP:} \quad & D_i \overset{\text{iid}}{\sim} f_{\phi^*}, \quad i = 1, \ldots, n \\ \textbf{Bayesian Model:} \quad & D_i \mid \phi \overset{\text{iid}}{\sim} f_{\phi_I}, \quad i = 1, \ldots, n, \qquad \phi \sim \pi \end{aligned} \tag{41}$$

where f is a distribution parametrized by ϕ and π is a prior. Then, under mild regularity conditions, the posterior behaves as follows:

$$\pi_{D_{1:n}} \overset{a.s.}{\to} \pi(\phi_N \mid \phi_I = \phi_I^*)\delta_{\phi_I^*}, \tag{42}$$

where $\pi_{D_{1:n}}$ is the posterior distribution and $\delta_{\phi_I^*}$ is a point mass at ϕ_I^*.

Simulation Settings

In our experiments, we consider the following data generating process:

- **Baseline Scenario:** Original setting of Parast et al. (2024):
  - *Useless Surrogate*: Gaussian outcomes, S and Y independent
  - *Perfect Surrogate*: Gaussian outcomes, S and Y highly correlated
  - *Imperfect Surrogate*: Gaussian outcomes, S and Y less correlated
  - *Misspecified Model*: Non-Gaussian outcomes
- **Augmented Scenario:** Setting with a binary baseline covariate X:
  - Treatment effect homogeneity across X
  - S and Y highly correlated within each X group

Useless Surrogate Setting

In the useless surrogate setting, we generate data as follows:

$$Y_{1i} \overset{\text{iid}}{\sim} \mathcal{N}(5, 3) \qquad Y_{0i} \overset{\text{iid}}{\sim} \mathcal{N}(3, 3)$$

$$S_{1i} \overset{\text{iid}}{\sim} \mathcal{N}\left(\tfrac{2}{3}, 1\right) \qquad S_{0i} \overset{\text{iid}}{\sim} \mathcal{N}\left(\tfrac{2}{3}, 1\right)$$

All potential outcomes are independent and there is no treatment effect on S, which is therefore a useless surrogate for Y.

Perfect Surrogate Setting

In the perfect surrogate setting, we generate data as follows:

$$Y_{1i} \overset{\text{iid}}{\sim} \mathcal{N}(6, 3) \qquad Y_{0i} \overset{\text{iid}}{\sim} \mathcal{N}\left(\tfrac{5}{2}, 3\right)$$
$$S_{1i} = Y_{1i} + \mathcal{N}\left(0, \tfrac{1}{10}\right) \qquad S_{0i} = Y_{0i} + \mathcal{N}\left(0, \tfrac{1}{10}\right)$$

S is just a noisy version of Y with a small noise variance.

Imperfect Surrogate Setting

In the imperfect surrogate setting, we generate data as follows:

$$S_{1i} \overset{\text{iid}}{\sim} \mathcal{N}(5, 3) \qquad S_{0i} \overset{\text{iid}}{\sim} \mathcal{N}(3, 3)$$
$$Y_{1i} = S_{1i} + \mathcal{N}\left(\tfrac{3}{2}, \tfrac{3}{5}\right) \qquad Y_{0i} = S_{0i} + \mathcal{N}\left(0, \tfrac{3}{5}\right)$$

Y is a noisy version of S with a larger noise variance.

Misspecified Model Setting

In this setting, we the data is not Gaussian, but rather generated as follows:

$$S_{1i} \overset{\text{iid}}{\sim} \text{log-Normal}\left(\tfrac{5}{2}, \tfrac{3}{2}\right)$$
$$Y_{1i} = 2 + \tfrac{6}{5}\sqrt{S_{1i}} + \tfrac{3}{10}\exp\left\{\tfrac{S_{1i}}{500}\right\} + \text{log-Normal}\left(0, \tfrac{3}{10}\right)$$

$$S_{0i} \overset{\text{iid}}{\sim} \text{log-Normal}\left(\tfrac{1}{2}, \tfrac{3}{2}\right)$$
$$Y_{0i} = \tfrac{4}{5}\sqrt{S_{0i}} + \tfrac{1}{5}\exp\left\{\tfrac{S_{0i}}{50}\right\} + \text{log-Normal}\left(0, \tfrac{3}{10}\right)$$

Y is a non-linear function of S plus some noise.

Criteria for the Validation Threshold

Following Parast et al. (2024), we choose the validation threshold η based on the power of the test to detect the treatment effect on S.

We argue that if S is a **valid** surrogate if the following conditions hold:

1. **Similarity:** V_S is close to V_Y,
2. **Detectability:** The test for V_S has a high power,

where the power is evaluated against the following hypotheses:

$$\begin{cases} H_0 : V_S = \frac{1}{2} \\ H_1 : V_S > \frac{1}{2} \end{cases} \tag{43}$$

Validation Threshold

In mathematical terms, the validity conditions can be written as:

$$\begin{cases} 0 < V_Y - V_S < \eta, \\ V_S > v_S > \frac{1}{2}, \end{cases} \tag{44}$$

where v_S is the minimum value of V_S which guarantees a power of $1 - \beta$. Then, the smallest admissible threshold η is obtained at the **boundary**:

$$\eta = \max\{v_Y - v_S, 0\}. \tag{45}$$

where v_Y denotes the hypothesized treatment effect on Y.

Bayes Factor Approach

To detect the treatment effect on S one could use the **Bayes factor** arising from this Bayesian model taking V_S as its unknown parameter:

Bayes Factor for Testing V_S

Consider the following model:

$$n\,\widehat{V}_S \mid V_S \sim \text{Binomial}(n, V_S), \quad 2\,V_S - 1 \sim \text{Beta}(a, b), \tag{46}$$

where $\widehat{V}_S = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(S_{1i} > S_{0i})$. Thus, the Bayes factor is

$$BF_n = \frac{1 - F_{\text{Beta}\left(a+n\,\widehat{V}_S,\, b+n-n\,\widehat{V}_S\right)}\left(\frac{1}{2}\right)}{1 - F_{\text{Beta}(a,b)}\left(\frac{1}{2}\right)}\, 2^n, \tag{47}$$

where $F_{\text{Beta}(c,d)}$ is the CDF of a Beta distribution with parameters c, d.

Rejection Threshold for the Bayes Factor Test

In a **frequentist** view, to ensure a type I error probability of α, we need to compute the value of $BF_{n,\alpha}$ that satisfies the following equation:

$$\mathbb{P}(BF_n > BF_{n,\alpha} \mid H_0) = \alpha. \tag{48}$$

Thus, $BF_{n,\alpha}$ is the $(1-\alpha)$-quantile of the distribution of BF_n under H_0. As BF_n takes at most $n+1$ values, its computation is **numerically feasible**.

Computation of v_S

Given the quantile $BF_{n,\alpha}$, we can compute v_s as the value of V_S such that

$$\mathbb{P}\left(BF_n > BF_{n,\alpha} \mid V_S = v_S\right) = 1 - \beta. \tag{49}$$

Notice that since for $v_S \in \left(\frac{1}{2}, 1\right)$

$$v_S \mapsto \mathbb{P}\left(BF_n > BF_{n,\alpha} \mid V_S = v_S\right) - 1 + \beta \tag{50}$$

is a **continuous** and **monotone** function, its unique root can be easily obtained using standard numerical root-finding algorithms.

Definition of the Bayes Factor

The Bayes factor in Equation (47) is defined as

$$BF_n = \frac{L\left(\widehat{V}_S \mid H_1\right)}{L\left(\widehat{V}_S \mid H_0\right)}, \tag{51}$$

where $L\left(\widehat{V}_S \mid H_0\right)$ and $L\left(\widehat{V}_S \mid H_1\right)$ are the marginal likelihoods of $\widehat{V}_S$ under the null and the alternative hypothesis, respectively.

Computation of $L\left(\widehat{V}_S \mid H_0\right)$

The marginal likelihood of $\widehat{V}_S$ under the null hypothesis is

$$\begin{aligned} L\left(\widehat{V}_S \mid H_0\right) &= \text{Binomial}\left(n\,\widehat{V}_S \mid n, \tfrac{1}{2}\right) \\ &= \binom{n}{n\,\widehat{V}_S} \left(\tfrac{1}{2}\right)^{n\widehat{V}_S} \left(1-\tfrac{1}{2}\right)^{n-n\,\widehat{V}_S} \\ &= \binom{n}{n\,\widehat{V}_S} \frac{1}{2^n}. \end{aligned} \tag{52}$$

Computation of $L\left(\widehat{V}_S \mid H_1\right)$

The marginal likelihood of $\widehat{V}_S$ under the alternative hypothesis is

$$
\begin{aligned}
L\left(\widehat{V}_S \mid H_1\right) &= \int_{\frac{1}{2}}^{1} \text{Binomial}\left(n\,\widehat{V}_S \mid n, v_S\right) \frac{\text{Beta}(v_S \mid a, b)}{\int_{\frac{1}{2}}^{1} \text{Beta}(v \mid a, b) dv} dv_S \\
&= \int_{\frac{1}{2}}^{1} \binom{n}{n\,\widehat{V}_S} v_S^{n\widehat{V}_S} (1 - v_S)^{n - n\,\widehat{V}_S} \frac{v_S^{a-1}(1 - v_S)^{b-1}}{\int_{\frac{1}{2}}^{1} v^{a-1}(1 - v)^{b-1} dv} dv_S \\
&= \binom{n}{n\,\widehat{V}_S} \frac{\int_{\frac{1}{2}}^{1} v_S^{a+n\widehat{V}_S-1}(1 - v_S)^{b+n-n\widehat{V}_S-1} dv_S}{\int_{\frac{1}{2}}^{1} v^{a-1}(1 - v)^{b-1} dv} \\
&= \binom{n}{n\,\widehat{V}_S} \frac{1 - F_{\text{Beta}\left(a+n\,\widehat{V}_S, b+n-n\,\widehat{V}_S\right)}\left(\frac{1}{2}\right)}{1 - F_{\text{Beta}(a,b)}\left(\frac{1}{2}\right)}.
\end{aligned}
\tag{53}
$$

Computation of $L\left(\widehat{V}_S \mid H_1\right)$

Thus, the Bayes factor is

$$BF_n = \frac{L\left(\widehat{V}_S \mid H_1\right)}{L\left(\widehat{V}_S \mid H_0\right)} = \frac{1 - F_{\text{Beta}\left(a+n\,\widehat{V}_S, b+n-n\,\widehat{V}_S\right)}\left(\frac{1}{2}\right)}{1 - F_{\text{Beta}(a,b)}\left(\frac{1}{2}\right)}\, 2^n. \tag{54}$$

References I

Fay, M. P., Brittain, E. H., Shih, J. H., Follmann, D. A., and Gabriel, E. E. (2018). Causal estimands and confidence intervals associated with wilcoxon-mann-whitney tests in randomized experiments. *Statistics in Medicine*, 37(20):2923–2937.

Gustafson, P. (2010). Bayesian inference for partially identified models. *The international journal of biostatistics*, 6(2).

Parast, L., Cai, T., and Tian, L. (2024). A rank-based approach to evaluate a surrogate marker in a small sample setting. *Biometrics*, 80(1):ujad035.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

References II

Schwartz, L. (1965). On bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 4(1):10–26.