Principal Stratification

BIOSTAT 258 Midterm Presentation

D. Rubin   C. Frankagakis

Presenters: Haobin (Tony) Chen, Christian Testa, Fuyu Guo

April 16th 2024

Topic of Discussion

Our article:

Today we'll be discussing the Biometrics 2002 article *Principal Stratification in Causal Inference* by Constantine E. Frangakis and Donald B. Rubin.

Biometrics
JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY
International Biometric Society
AN INTERNATIONAL SOCIETY DEVOTED TO THE DEVELOPMENT AND APPLICATION OF STATISTICAL AND MATHEMATICAL THEORY AND METHODS IN THE BIOSCIENCES
OXFORD UNIVERSITY PRESS

Keywords: Biomarker, Causal inference, Censoring by death, Missing data, Posttreatment variable, Principal stratification, Quality of life, Rubin causal model, Surrogate

Table of Contents

§1 Background and §2 Goals and Standard Approaches for Post-treatment Variables

§3 Principal stratification and principal causal effects

§4 Examples of Principal Effects in Action

§5 Defining Surrogate Endpoints Using Causal Effects

- 5.1 Two Goals of Surrogate Endpoints and Previous Approaches
- 5.2 Definition of Principle Surrogate and Property of Causal Necessity
- 5.3 Principal Stratification and Property of Statistical Generalizability

§6 Remarks and Extensions

Aftermath

Recommended References

Background

Causal effects

Consider a group of units $i = 1, \dots, n$, where each unit can be assigned a standard treatment $z = 1$ or a new treatment $z = 2$. The measurement is an outcome variable Y, e.g., survival status at a specific time. Let $Y_i(z)$ be the value of Y if unit i is assigned treatment z. Then a **causal effect** of Z on the measurement Y is defined as be comparison between the sets of potential outcomes on a **common set of units**, i.e.,

$$\{Y_i(1) : i \in \mathcal{I}_1\} \quad \text{and} \quad \{Y_i(2) : i \in \mathcal{I}_2\}$$

where $\mathcal{I}_1 = \mathcal{I}_2$. *That is, we are comparing the same group of units.*

Subgroup comparison prior to treatment

Usually, in observational studies, researchers are interested in the subgroup comparison defined by pre-treatment variables, e.g.,

$$\{Y_i(1) : L_i \in \mathcal{L}\} \quad \text{and} \quad \{Y_i(2) : L_i \in \mathcal{L}\}$$

For example, we can compare the survival rates between two treatments among males, where l_i represents the pre-treatment sex. The potential outcomes are not all observable, and we can only use observed data for the comparison. For example, we may compare

$$\Pr(Y_i = 1 | L_i = l, A_i = 1) \quad \text{to} \quad \Pr(Y_i = 1 | L_i = l, A_i = 0)$$

This comparison can be causal, given the conditional exchangeability assumption.

$$A_i \perp\!\!\!\perp Y_i(1), Y_i(0) | L_i = l$$

Comparison by post-treatment variable

Usually, a variable S_i is measured after the treatment, i.e., post-treatment. Some examples:

- In RCT, S_i is a measure of an individual's compliance with the assigned treatment.
- In large cohort studies, S_i is the indicator for missing outcomes.
- In large cohort studies, S_i is the censoring indicator.
- In RCTs where the outcome of interest takes a long time to occur, to save time and expensive, researchers may choose S_i as a surrogate endpoint.

Adjusting for post-treatment variables

Sometimes, comparisons after adjusting for post-treatment S_i are of concern, e.g., conditioning on individuals who complied with the assigned treatment can help assess the efficacy.

The net treatment effect of assignment Z, adjusting for the post-treatment variable S^{obs}

A standard way for the adjustment is to compare,

$$\begin{gathered}\Pr\left\{Y_i^{\text{obs}} = 1 \mid S_i^{\text{obs}} = s, Z_i = 1\right\} \\ \Pr\left\{Y_i^{\text{obs}} = 1 \mid S_i^{\text{obs}} = s, Z_i = 2\right\}\end{gathered} \tag{1}$$

Immediately, it follows from the consistency assumption,

$$\Pr\left\{Y_i^{\text{obs}} = 1 \mid S_i^{\text{obs}} = s, Z_i = z\right\} = \Pr\left\{Y_i(z) = 1 \mid S_i(z) = s, Z_i = z\right\}$$

Post-treatment selection bias

In a completely randomized trial, where $Z \perp\!\!\!\perp (S(1), S(0), Y(1), Y(0))$, the comparison (1) is equivalent to the comparison by the exchangeability assumption,

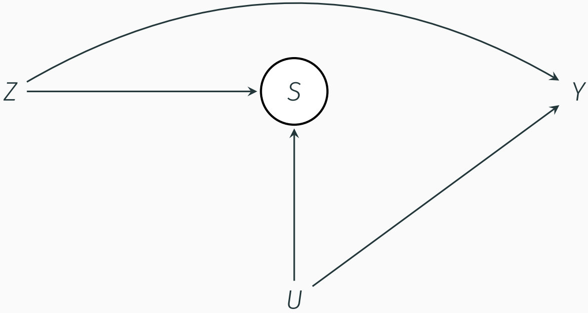
$$\Pr\{Y_i(1) = 1 \mid S_i(1) = s\} \quad \text{and} \ \Pr\{Y_i(2) = 1 \mid S_i(2) = s\}$$

This comparison can represent a causal effect if the Z does not affect S, but can be problematic otherwise. Individuals the groups $\{i : S_i(1) = s\}$ and $\{i : S_i(2) = s\}$ are different.

- For example, the new treatment $Z = 2$ can reduce the blood pressure S, compared to the standard treatment $Z = 1$. We also know older people tend to have higher blood pressure. If we know that Person A in the new treatment group has the same $S = s$ as Person in the standard treatment, then Person A is more likely to be older than Person B.

Post-treatment selection bias

Since, $\{i : S_i(1) = s\}$ and $\{i : S_i(2) = s\}$ are different groups, by the definition, comparison (1) does not represent a causal effect. This concern is referred to as post-treatment selection bias in Epidemiology.



Z: treatment, S: blood pressure, Y: 2-year mortality, U: unknown variables for S and Y.

Table of Contents

§1 Background and §2 Goals and Standard Approaches for Post-treatment Variables

§3 Principal stratification and principal causal effects

§4 Examples of Principal Effects in Action

§5 Defining Surrogate Endpoints Using Causal Effects

- 5.1 Two Goals of Surrogate Endpoints and Previous Approaches
- 5.2 Definition of Principle Surrogate and Property of Causal Necessity
- 5.3 Principal Stratification and Property of Statistical Generalizability

§6 Remarks and Extensions

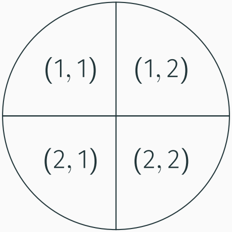
Aftermath

Recommended References

Principal stratification

Definition

1. The basic principal stratification P_0 w.r.t. posttreatment variable S is the partition of units $i = 1, \ldots, n$ such that, within any set of P_0, all units have the same vector $(S_i(1), S_i(2))$.
2. A principal stratification P w.r.t. posttreatment variable S is a partition of the units whose sets are union of sets in the basic principal stratification P_0.

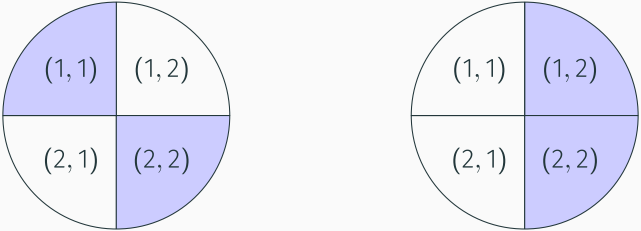


For example, the above pie plot represents a basic principal stratification, whose $(S_i(1), S_i(2))$ are $(1, 1), (1, 2), (2, 1), (2, 2)$

Examples for a principle stratification

Example 1. Partition people into stratification where the treatment has no on S, i.e., $S_i(2) = S_i(1)$, i.e.,

Example 2. Partition people into stratification where $S_i(2)$ will always take 1 and 2, if the assigned treatment is $Z = 2$.



(a) Example 1: $S_i(1) = S_i(2)$

(b) Example 2: $S_i(2) = 1$ and $S_i(2) = 2$

Principal effect w.r.t principal stratification

Definition

Let P be a principal stratification w.r.t. the posttreatment variable S nd let S_i^P indicate the stratum of P to which unit i belongs. Then a principal effect w.r.t. that principal stratification is defined as a comparison of potential outcomes under standard versus new treatment within a principal stratum ς in P, i.e., a comparison between the sets

$$\left\{Y_i(1) : S_i^P = \varsigma\right\} \quad \text{and} \quad \left\{Y_i(2) : S_i^P = \varsigma\right\} \tag{2}$$

Note that by definition, the values of $S_i(1), S_i(2)$ are affected by the treatment Z, which leads to

Properties

Property 1: the stratum S_i^P is unaffected by treatment for an principal stratifcation P.

Property 2: Any principal effect (2) is a causal effect because the comparison are on the same set of people.

- Non-causal comparison

$$\Pr\{Y_i(1) \mid S_i(1) = s\} \quad \text{and } \Pr\{Y_i(2) \mid S_i(2) = s\}$$

- Causal comparison

$$\Pr\left\{Y_i(1) \mid S_i^P = \varsigma\right\} \quad \text{and } \Pr\left\{Y_i(2) \mid S_i^P = \varsigma\right\}$$

Inference based on the observed data

Suppose we aim to estimate the principal effects, but have only observed data $H^{\text{obs}} = (Y^{\text{obs}}, S^{\text{obs}}, Z)$. There are two missing potential variables, $S^{\text{mis}} = \{S_i(z) : \text{all } i; z \neq Z_i\}$ and $Y^{\text{mis}} = \{Y_i(z) : \text{all } i; z \neq Z_i\}$. The observed likelihood is a marginal one that integrates over $(S^{\text{mis}}, Y^{\text{mis}})$

$$
\begin{aligned}
& L\left(H^{\mathrm{obs}}; \theta^S, \theta^Y\right) \\
& = \iint \mathrm{pr}\left\{Z \mid (Y(1), Y(2)), S^{P_0}\right\} \times \mathrm{pr}\left(S^{P_0} \mid \theta^S\right) \\
& \times \mathrm{pr}\left\{(Y(1), Y(2)) \mid S^{P_0}; \theta^Y\right\} dY^{\mathrm{mis}} dS^{\mathrm{mis}}
\end{aligned}
$$

where θ^Y, θ^S represents the causal effect of Z on Y and S respectively.

Table of Contents

§1 Background and §2 Goals and Standard Approaches for Post-treatment Variables

§3 Principal stratification and principal causal effects

§4 Examples of Principal Effects in Action

§5 Defining Surrogate Endpoints Using Causal Effects

- 5.1 Two Goals of Surrogate Endpoints and Previous Approaches
- 5.2 Definition of Principle Surrogate and Property of Causal Necessity
- 5.3 Principal Stratification and Property of Statistical Generalizability

§6 Remarks and Extensions

Aftermath

Recommended References

Three Examples

Three important types of problems that can be handled through the framework of principal effects are

(i) treatment noncompliance

(ii) missing outcomes following treatment noncompliance, and

(iii) censoring by death.

Complier Average Causal Effect (CACE)

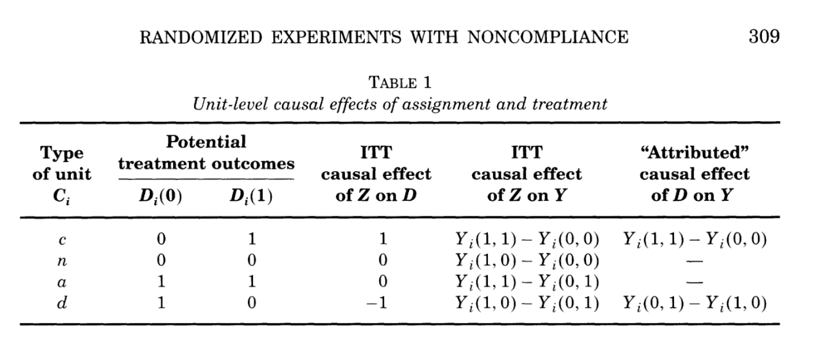
Consider, for example, Imbens and Rubin’s 1997 re-analysis of data from a study on Vitamin A study where interest was in the effect of **taking vitamin A vs. not taking vitamin A**; in other words, the causal effect when everyone assigned to treatment takes treatment and everyone assigned to nontreatment.

The CACE a special case of a principal effect: this is a contrast of potential outcomes among the **”compliers”**, noting that ”compliers” are a valid principal strata.

*In general, regulatory agencies and practitioners do not trust the exchangeability assumptions necessary for uncontrolled compliance. *Why would you approve a drug based on the effect of* $\underline{\textit{assignment}}$ *to* $Z = 1$ *or* $Z = 0$ *when you could look at the actual effect?*

CACE Table

From Imbens and Rubin 1997, *Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance*:



RANDOMIZED EXPERIMENTS WITH NONCOMPLIANCE 309

TABLE 1
Unit-level causal effects of assignment and treatment

Type of unit	**Potential treatment outcomes**		**ITT causal effect of Z on D**	**ITT causal effect of Z on Y**	**"Attributed" causal effect of D on Y**
C_i	$D_i(0)$	$D_i(1)$			
c	0	1	1	$Y_i(1,1) - Y_i(0,0)$	$Y_i(1,1) - Y_i(0,0)$
n	0	0	0	$Y_i(1,0) - Y_i(0,0)$	—
a	1	1	0	$Y_i(1,1) - Y_i(0,1)$	—
d	1	0	−1	$Y_i(1,0) - Y_i(0,1)$	$Y_i(0,1) - Y_i(1,0)$

- D is the realized outcome (disease)
- C is the principal strata (complier, never-taker, always-taker, defier)
- ITT = intention-to-treat
- Z is treatment
- Y is a potential outcome (inspired by Neyman's *'potential yield'*)
- $Y(Z, \_)$ represents their potential outcome under assignment to treatment and what they did after.

Alternatives to CACE

Note that the **CACE is a subgroup analysis**. It would be nice to use more of the data to estimate the complier-effect. However, other estimands require simulating what would have happened if, under some treatment level $Z = z$,

1. All subjects (including noncompliers) were somehow forced to take the treatment;
2. All subjects (including noncompliers) would have been forced to take the control.

Simulating such scenarios poses issues. Part of the problem arises from the fact that ”forcing them” into treatment or nontreatment is not a function of the controllable factor Z alone, and therefore the authors claim, does not lead to well-defined causal effects.

Missing Outcomes Following Treatment Noncompliance

Consider the setting where **school-choice programs** are being evaluated through a randomized intervention to provide school vouchers to children of low-income parents. Posttreatment Problems:

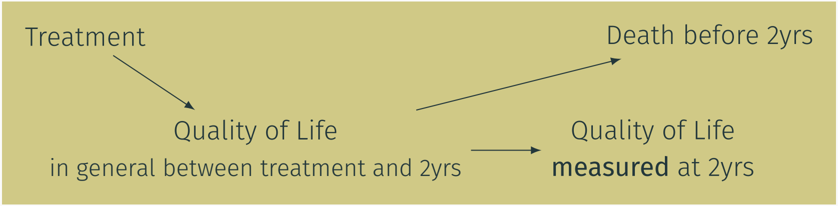
1. Actual use of the vouchers was uncontrolled;
2. Not every child took subsequent standardized tests (their choice of outcome) to evaluate performance.

Frangakis and Rubin (1997, 1999) show that intention-to-treat analysis does not estimate intention-to-treat effects *when there is outcome missingness dependent on post-treatment variables.*

The recommended, better approach is to use **principal strata** defined both by compliance and missingness.

Censoring Due to Death Dependent on Posttreatment

Consider the time-ordered events:



If QOL (Quality of Life at 2yrs) is censored due to death, we should not try to impute it because it's not a "missing value;" it's just a truly null value — a value that never existed in the first place.

Progress is made on the problem in Rubin's 2006 *Causal Inference Through Potential Outcomes and Principal Stratification: Application to Studies with "Censoring"* by considering the principal stratification {those who always live, those who always die, live under treatment + die under control, die under treatment + live under control}.

Table of Contents

§1 Background and §2 Goals and Standard Approaches for Post-treatment Variables

§3 Principal stratification and principal causal effects

§4 Examples of Principal Effects in Action

§5 Defining Surrogate Endpoints Using Causal Effects

- 5.1 Two Goals of Surrogate Endpoints and Previous Approaches
- 5.2 Definition of Principle Surrogate and Property of Causal Necessity
- 5.3 Principal Stratification and Property of Statistical Generalizability

§6 Remarks and Extensions

Aftermath

Recommended References

Table of Contents

§5 Defining Surrogate Endpoints Using Causal Effects

5.1 Two Goals of Surrogate Endpoints and Previous Approaches

5.2 Definition of Principle Surrogate and Property of Causal Necessity

5.3 Principal Stratification and Property of Statistical Generalizability

Principle Stratum & Surrogacy

Define a randomized intervention $Z_i \in \{0, 1\}$, mortality outcome $Y_i(z)$ as primary endpoint, and a surrogate variable for endpoint $S_i(z)$.

Property of surrogacy

1. (Causal necessity): Z has causal effect on Y only if Z has an effect on S.
2. (Statistical generalizability): S^{obs} should be correlated with Y^{obs} in application, even when Y^{obs} is not immediately observed.

- Treatment can't act on outcome without manipulating surrogate.
- It's important to have predictability since observing mortality endpoint is unattainable.

Principle Stratum & Surrogacy

Some past efforts to describe surrogacy:

- "$Y^{obs} = Y_i(Z = z_i) \perp\!\!\!\perp Z_i \mid S_i^{obs}$" (Prentice 1989)
  - Regression parameter coefficient, R^2, net treatment comparison

Definition (Traditional statistical surrogacy)

$$P\left(Y_i^{obs} \middle| Z_i = 1, S_i^{obs} = s\right) = P\left(Y_i^{obs} \middle| Z_i = 2, S_i^{obs} = s\right)$$

$\implies S$ is a *statistical surrogate* for comparing effect of Z on Y.

Limitation: Fail to satisfy the causal necessity property – possible that unit whose treatment doesn't affect *statistical surrogate* but affects outcome $\longrightarrow$ does not satisfy causal necessity.

Solution: Define an improved *principal surrogate*.

Table of Contents

§5 Defining Surrogate Endpoints Using Causal Effects

- 5.1 Two Goals of Surrogate Endpoints and Previous Approaches
- 5.2 Definition of Principle Surrogate and Property of Causal Necessity
- 5.3 Principal Stratification and Property of Statistical Generalizability

Principle Stratum & Surrogacy

Definition (Principal surrogate)

$$\{Y_i(1) : S_i(1) = S_i(2) = s\} = \{Y_i(2) : S_i(1) = S_i(2) = s\}, \forall\, s$$

$\implies S$ is a principal surrogate for Z and Y.

Guarantees that causal effect of treatment on only exists when causal effect of treatment on surrogate exists (definition $\sim$ reads contrapositively).

Under randomization, principal surrogate implies:

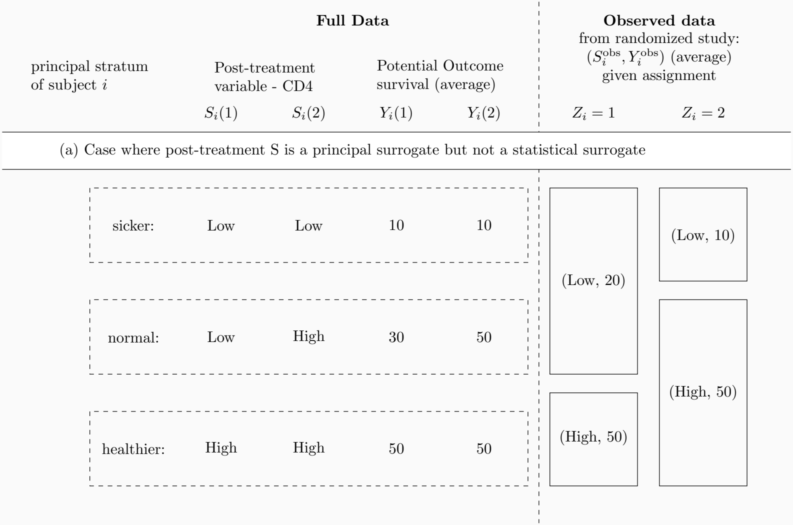
$$P\left(Y_i^{obs} \middle| Z_i = 1, S_i(1) = S_i(2) = s\right) = P\left(Y_i^{obs} \middle| Z_i = 2, S_i(1) = S_i(2) = s\right)$$

and have corollaries:

1. S is principal surrogate $\implies S$ is generally not a statistical surrogate;
2. S is statistical surrogate $\implies S$ is generally not a principal surrogate;

Example with HIV treatment, CD4 counts, and mortality.

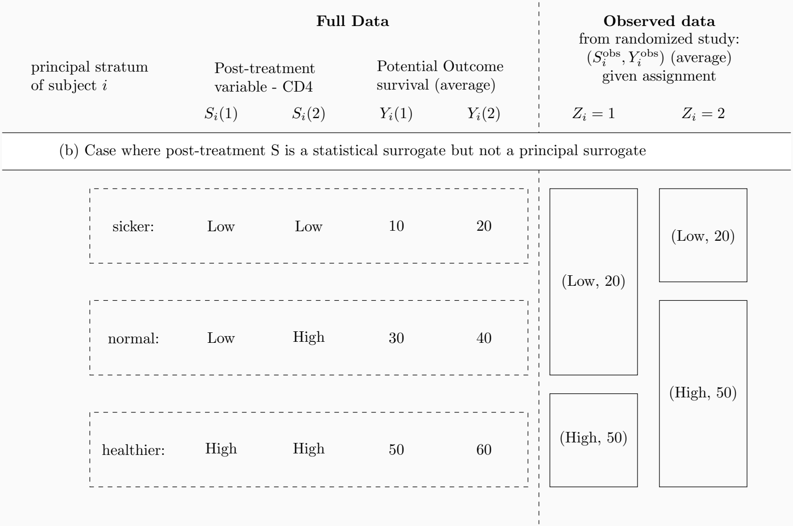
Figure 1a — Distinction between Principal and Statistical Surrogates



	Full Data				**Observed data** from randomized study: $(S_i^{\text{obs}}, Y_i^{\text{obs}})$ (average) given assignment	
principal stratum of subject i	Post-treatment variable - CD4		Potential Outcome survival (average)			
	$S_i(1)$	$S_i(2)$	$Y_i(1)$	$Y_i(2)$	$Z_i = 1$	$Z_i = 2$
(a) Case where post-treatment S is a principal surrogate but not a statistical surrogate						
sicker:	Low	Low	10	10	(Low, 20)	(Low, 10)
normal:	Low	High	30	50		(High, 50)
healthier:	High	High	50	50	(High, 50)	

Observe that if $S_i(1) = S_i(2) = s$, there is no effect of treatment on the outcome.

Figure 1b — Distinction between Principal and Statistical Surrogates



	Full Data				**Observed data** from randomized study: $(S_i^{\text{obs}}, Y_i^{\text{obs}})$ (average) given assignment	
principal stratum of subject i	Post-treatment variable - CD4		Potential Outcome survival (average)			
	$S_i(1)$	$S_i(2)$	$Y_i(1)$	$Y_i(2)$	$Z_i = 1$	$Z_i = 2$
(b) Case where post-treatment S is a statistical surrogate but not a principal surrogate						
sicker:	Low	Low	10	20	(Low, 20)	(Low, 20)
normal:	Low	High	30	40		(High, 50)
healthier:	High	High	50	60	(High, 50)	

Controlling for a fixed level of $S_i^{obs} = s$, the observed treatment effect goes away.

Associative and Dissociative Effects

Suggestion: Evaluate the effects of treatment on outcome that are associative and dissociative with effects on the posttreatment variable.

Treatment-outcome effects that are dissociative with effect on the surrogate are comparisons between:

$$\{Y(1) : S(1) = S(2)\} \quad \text{and} \quad \{Y(0) : S(1) = S(2)\}.$$

On the other hand, treatment-outcome effects that are associative with effect on the surrogate are comparisons between

$$\{Y(1) : S(1) \neq S(2)\} \quad \text{and} \quad \{Y(0) : S(1) \neq S(2)\}.$$

Large dissociative effect ⇒ conclude large effect for subjects with CD4 unaffected.
Large associative effect ⇒ conclude there is a large effect for subjects whose CD4 was affected by treatment.

Table of Contents

§5 Defining Surrogate Endpoints Using Causal Effects

5.1 Two Goals of Surrogate Endpoints and Previous Approaches

5.2 Definition of Principle Surrogate and Property of Causal Necessity

5.3 Principal Stratification and Property of Statistical Generalizability

Principal Stratification for Surrogate Endpoints: Setup

Consider: How can principal stratification help us predict the outcomes in a randomized study that wants to use surrogate endpoints? E.g., to not wait for the expensive Y^{obs}.

Let us consider the following probability models for a validation study V and an application study A:

$$P^V\{(S(1), S(2))\}, \quad P^V\{Y^{\text{obs}}|S(1), S(2), Z\}, \tag{V1}$$

$$P^A\{(S(1), S(2))\}, \quad P^A\{Y^{\text{obs}}|S(1), S(2), Z\}. \tag{A1}$$

Assume that all of the above are available except for $P^A(Y^{\text{obs}}|S(1), S(2), Z)$.

Principal Stratification for Surrogates: Predictive Distribution

Before the outcomes Y_i^{obs} are known in the application study, we can predict them according to the following distribution:

$$P^A(Y^{\mathrm{obs}}|S^{\mathrm{obs}}, Z) = \frac{\int P^A\{Y^{\mathrm{obs}}|S(1), S(2), Z\}P^A\{S(1), S(2)\}\mathrm{d}S^{\mathrm{mis}}}{\int P^A\{S(1), S(2)\}\mathrm{d}S^{\mathrm{mis}}}.$$

However, this is not available to us because we do not have $P^A(Y^{\mathrm{obs}}|S(1), S(2), Z)$.

*A common practice is to use $P^V(Y|S^{\mathrm{obs}}, Z)$ but this replaces both probability distributions of A1 with those of V1.

Principal Stratification for Surrogates: Problem

The *problem* with this common practice is that the application study can differ from the validation study in either **distribution of principal strata** or the **potential outcomes given the principal strata**.

This makes the predictive distribution incorrect for the application study.

This could explain why the empirical distribution $P^V(Y^{\text{obs}}|S^{\text{obs}}, Z)$ may differ substantially from one validation study to another.

Instead, consider that the right side of A1 matches the right side of V1 than it is that both sides match.

$\Rightarrow$ Strategy: replace $P^A(Y^{\text{obs}}|S(1), S(2))$ with $P^V(Y^{\text{obs}}|S(1), S(2))$

Principal Stratification for Surrogates: Solution

By replacing $P^A(Y^{\text{obs}}|S(1), S(2))$ with $P^V(Y^{\text{obs}}|S(1), S(2))$, we obtain the synthetic predictive distrbiution:

$$P^{\text{SYN}}(Y^{\text{obs}}|S^{\text{obs}}, Z) = \frac{\int P^V\{Y^{\text{obs}}|S(1), S(2), Z\}P^A\{S(1), S(2)\}\mathrm{d}S^{\text{mis}}}{\int P^A\{S(1), S(2)\}\mathrm{d}S^{\text{mis}}}.$$

The authors claim that this should be a more plausible approximation to the correct predictive distribution in the application study than just replacing $P^A\{Y^{\text{obs}}|S(1), S(2), Z\}$ with $P^V\{Y^{\text{obs}}|S(1), S(2), Z\}$.

Estimation in Principal Stratification

Recall unconfoundedness, $p(Z_i|Y_i(1), Y_i(0), D_i(1), D_i(0), X_i; \vec{\theta}) = p(Z_i|X_i)$ is the propensity score model,

$$
\begin{aligned}
L(\vec{\theta}) &= \prod_{i=1}^{n} p\Big(Y_i(1), Y_i(0), D_i(0), D_i(1), Z_i, X_i; \vec{\theta}\Big) \\
&= \prod_{i=1}^{n} \underbrace{p(Z_i|Y_i(1), Y_i(0), X_i; \vec{\theta})}_{\text{const.}} \underbrace{p(Y_i(1), Y_i(0)|S_i, X_i; \vec{\theta})}_{\text{unconfounded}} p(S_i|X_i; \vec{\theta}) p(X_i|\vec{\theta}) \\
&\propto \prod_{i=1}^{n} \underbrace{p(Y_i(1), Y_i(0)|S_i, X_i; \vec{\theta})}_{outcome} \underbrace{p(S_i|X_i; \vec{\theta})}_{strata} p(X_i|\vec{\theta}) \\
&\propto \prod_{i} p(Y_i(0)|S_i, X_i, \vec{\theta})^{I(Z_i=0)} p(Y_i(1)|S_i, X_i, \vec{\theta})^{I(Z_i=1)} p(S_i|X_i; \vec{\theta})
\end{aligned}
$$

assuming potential outcome independence. Also propensity score model is ignorable.

Estimation in Principal Stratification

More granularly, individual stratum composition:

$$L(\vec{\theta}) \propto \prod_{i=1}^{n} \sum_{s \in S: D_i = D(s, Z_i)} p(S_i = s | X_i, \vec{\theta}) p(Y_i | S_i = s, Z_i, X_i, \vec{\theta})$$

S contains all possible strata; $D(s, z)$ is post-treatment variable under s and treatment z (Liu, 2023; Li, 2022). Resembles the form of a Gaussian mixture model: "$\pi \mathcal{N}_1 + (1 - \pi)\mathcal{N}_2$".

Impose (multivariate) prior distribution $p(\vec{\theta})$ on parameter vectors that index two probability models, allowing us to harvest $p(\vec{\theta} | Z, D, Y, X)$, the posterior.

In reality, compute posterior iteratively on: 1) $p(D^{mis} | Y^{obs}, D^{obs}, Z, X; \theta)$ and 2) $p(\theta | Y^{obs}, D^{mis,obs}, Z, X; \theta)$

The *principal causal effect* is identified as:

$$PCE = \mathbb{E}\Big[\mathbb{E}[Y_i | Z_i = 1, S_i = s, X_i] \mid S_i = s\Big] - \mathbb{E}\Big[\mathbb{E}[Y_i | Z_i = 2, S_i = s, X_i] \mid S_i = s\Big]$$

Estimation in Principal Stratification

As a missing data (since all causality $\equiv$ missingness) & specifically a latent mixture problem, *EM algorithm* is also highly appropriate.
Same likelihood:

$$L(\vec{\theta}) \propto \prod_{i=1}^{n} \sum_{s \in S: D_i = D(s, Z_i)} p(S_i = s | X_i, \vec{\theta}) p(Y_i | S_i = s, Z_i, X_i, \vec{\theta})$$

- E-step: Impute S_i^t with $\mathbb{E}[S_i | Y^{obs}, D^{obs}, X_i, \vec{\theta}^t]$
- M-step: Get gradient and optimize:

$$\arg\max_{\vec{\theta}} \; L(\vec{\theta} | Y^{obs}, S, X)$$

Table of Contents

§1 Background and §2 Goals and Standard Approaches for Post-treatment Variables

§3 Principal stratification and principal causal effects

§4 Examples of Principal Effects in Action

§5 Defining Surrogate Endpoints Using Causal Effects

- 5.1 Two Goals of Surrogate Endpoints and Previous Approaches
- 5.2 Definition of Principle Surrogate and Property of Causal Necessity
- 5.3 Principal Stratification and Property of Statistical Generalizability

§6 Remarks and Extensions

Aftermath

Recommended References

Authors’ Remarks and Extensions

For outcomes adjusted for posttreatment variables, we focused on estimands before estimation by formulating principal causal effects.

Basically any application of principal strata will require imputing missing information on the principal strata, i.e., the unobserved counterfactual of the post-treatment covariate under different treatment.

Explicit restrictions such as latent ignorability* of outcome missingness or the compound exclusion restriction** can be more scientifically plausible than the implicit assumptions of standard approaches.

A key contribution is formalizing what types of assumptions are needed for estimating causal effects adjusted for post-treatment variables.

Authors' Remarks and Extensions: Necessary Footnotes

* As opposed to traditional ignorability assumptions that suppose conditioning on observed covariates, the potential outcomes are independent of the treatment $Y(1), Y(2) \perp Z \mid L$, the latent ignorability assumption instead supposes that conditional on observed covariates and some possibly unobserved latent factors U, we have that $Y(1), Y(2) \perp Z \mid L, U$.

One way this is written is $E[Y(1)|Z = 1, S = 1] = E[Y(1)|Z = 1, S = 0]$ and $E[Y(0)|Z = 0, S = 0] = E[Y(0)|Z = 1, S = 1]$.

** The compound exclusion restriction is commonly seen in instrumental variables, and states that the instrumental variable Z does not directly affect the outcome Y except through the treatment and that there are no unmeasured confounders of the $Z \sim Y$ relationship other than through the treatment. In this context, this means that the causal effect for the always-takers and never-takers is assumed to be 0.

Authors’ Proposed Extensions

Though we’ve considered the binary outcome and post-treatment variable case, the authors say the framework is immediately applicable to post-treatment variables that are multivariate, time-dependent, or continuous as well as continuous treatments.

Authors' Summary

Summary

"Continued use of the current frameworks in problems with posttreatment variables (e.g., surrogate endpoints) in principle makes incorrect attributions of effects of treatments."

'Until now, we had always thought that the roles of **biology and statistics did not mix** in these complex problems. But principal causal effects set the framework for allowing **biological assumptions** in statistical methods and vice versa.'

"We hope that this article provokes the development and dissemination of *more principled frameworks*."

Table of Contents

§1 Background and §2 Goals and Standard Approaches for Post-treatment Variables

§3 Principal stratification and principal causal effects

§4 Examples of Principal Effects in Action

§5 Defining Surrogate Endpoints Using Causal Effects

- 5.1 Two Goals of Surrogate Endpoints and Previous Approaches
- 5.2 Definition of Principle Surrogate and Property of Causal Necessity
- 5.3 Principal Stratification and Property of Statistical Generalizability

§6 Remarks and Extensions

Aftermath

Recommended References

Aftermath

- Very little practical recommendations *in this paper* for how to estimate $P(S(1), S(2))$; one of the key assumptions is that we have access to $P(S(1), S(2))$.
  - They highlight this in their statement *“we focused on estimands before estimation by formulating principal effects.”*
  -
- Pearl 2011 (*Transportability across studies: a formal approach*) heavily criticizes the idea of “principal strata”
  - “Their definition, called “principal surrogacy” requires that causal effects of X on Y may exist if and only if causal effects of X on Z exist, … but stops short of delineating the set of new conditions under which this requirement should be sustained.”
  - “Rubin (2004) went further and proposed to do away with “deceptive” concepts such as direct and indirect effects and replace them with “principal surrogacy.””

Aftermath

Mealli and Mattei respond in 2012 in Int. J. of Biostatistics:

> A principal stratification with respect to a post-treatment variable is a partition of units into latent classes defined by the joint potential values of that post-treatment variable under each of the treatments being compared. From this standpoint, some previous works (e.g., Robins (1986, 1998), Robins and Greenland (1989a,b, 1994), [...], Heckman and Vytlacil (2001)), [...] can be viewed as examples of principal stratification. By definition, principal strata are not affected by treatment assignment, therefore a principal stratification can be used as any classification of units, to **define meaningful causal estimands** conditional on principal strata, to **discover treatment effect heterogeneities**, to **state identifying assumptions as behavioral assumptions on the principal strata**.

* Emphasis added

Table of Contents

§1 Background and §2 Goals and Standard Approaches for Post-treatment Variables

§3 Principal stratification and principal causal effects

§4 Examples of Principal Effects in Action

§5 Defining Surrogate Endpoints Using Causal Effects

- 5.1 Two Goals of Surrogate Endpoints and Previous Approaches
- 5.2 Definition of Principle Surrogate and Property of Causal Necessity
- 5.3 Principal Stratification and Property of Statistical Generalizability

§6 Remarks and Extensions

Aftermath

Recommended References

Recommended References (1/2)

- Frangakis C. E., Rubin D. B. Principal stratification in causal inference. Biometrics. 2002 *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4137767/*
- Zhang, J. L., & Rubin, D. B. Estimation of Causal Effects via Principal Stratification When Some Outcomes Are Truncated by "Death." Journal of Educational and Behavioral Statistics. 2003 28(4), 353–368. *http://www.jstor.org/stable/3701340*
- Rubin, D. B. Causal Inference Through Potential Outcomes and Principal Stratification: Application to Studies with "Censoring" Due to Death. 2006 *https://arxiv.org/abs/math/0612783*
- Imbens G. W., Rubin D.B. Annals of Statistics 1997. Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance. *https://www.jstor.org/stable/2242722*
- Mealli F, Mattei A. A refreshing account of principal stratification. Int J Biostat (2012) *https://pubmed.ncbi.nlm.nih.gov/22611592/*
- Grossi G, Mariani M, Mattei A, Mealli F. Bayesian principal stratification with longitudinal data and truncation by death (2023). *https://arxiv.org/pdf/2401.00196.pdf*

Recommended References (2/2)

- Fan Li's Slides on Post-treatment Confounding: Principal Stratification *http://www2.stat.duke.edu/~fl35/teaching/640/Chapter6.2_principal%20stratification.pdf*
- Liu B., Li, F. PStrata: A package for Principal Stratification. Submitted to Journal of Statistical Software (2023) *https://arxiv.org/abs/2304.02740*
- Pearl, J. Principal stratification — A goal or a tool? *https://ftp.cs.ucla.edu/pub/stat_ser/r382.pdf*
- Tan X., Abberbock J., Rastogi P., Tang G. Identifying Principal Stratum Causal Effects Conditional on a Post-treatment Intermediate Response. Proceeding of Machine Learning Research (2022) *https://proceedings.mlr.press/v177/tan22a/tan22a.pdf*

Supplemental Materials

Supplemental: Relationship to Mediation? (1/2)

It's very natural to ask what is the relationship between this principal stratification framework and the framework of mediation?

See Mealli and Mattei (2012). The Principal Strata Direct Effect is the principal causal effect for the stratum where the intermediate variable $S(1) = S(2) = s$, i.e., $\text{PSDE}(s) = E[Y(2) - Y(1) \mid S(1) = S(2) = s]$. Only in strata where the intermediate variable is unaffected by the treatment can we learn something about the direct effect of treatment.

Causal mediation, on the other hand, focuses on disentangling direct and indirect effects. $\text{NDE}(x) = E[Y(2, S(x)) - Y(1, S(x))]$, $\text{NIE}(x) = E[Y(x, S(2)) - Y(x, S(1))]$, $x = 0, 1$ such that the average total causal effect $\text{ATE} = \text{NDE}(x) + \text{NIE}(1 - x)$.

Conversely, principal stratification does not allow decomposition of the total effect into direct and indirect effects without additional assumptions.

Supplemental: Relationship to Mediation? (2/2)

Still from Mealli and Mattei (2012) (paraphrased):

If $\mathrm{PSDE}(s) = 0$ for each $s \in \mathcal{S}$, then there is no evidence of direct effect of treatment after controlling for the mediator, because the causal effect only exists in the presence of an effect on the intermediate variable.

This does not mean that there is no natural direct effect of the treatment: The principal causal effects for units in principal strata where post-treatment is affected by treatment ('associative effects') combine natural direct and indirect effects.

Supplemental: Relationship to Instrumental Variables?

Principal Stratification and Instrumental Variables are quite intrinsically linked in the sense that Z (assignment to treatment) plays the role of an instrument and S is actually taking the treatment.

The assumption of monotonicity is saying there are no defiers, and the exclusion-restriction is an assumption on the causal effects for always-takers and never-takers.

Supplemental: Approaches for Use in Practice

PStrata implements a Bayesian approach (in Stan) for building S and Y-models.

It appears to have been extended to interesting settings like longitudinal data, survival analysis, multi-level data, etc.

* That said, in theory, the approach can also be used with the EM-algorithm or with a latent-mixture-modeling approach.

Supplemental: What are the framework's limitations?

- Same criticisms as the Rubin Causal Model in general: is it okay to use unobservable counterfactuals?
  - However, if the post-treatment variables are not *manipulable* we may not have the best evidence to base counterfactuals where we assign treatment one way and post-treatment variables another way (as if the unit had received a different level of treatment).
- You need a good model for the principal strata units fall into.
- Since the principal strata are never observable, you need strong identifying assumptions.
- May need a lot of data if there are lots of strata for which we need to estimate principal effects.