Statistical Analysis of Randomized Experiments with Nonignorable Missing Binary Outcomes

Kosuke Imai

Department of Politics
Princeton University

January 12, 2007

Overview

- Use of randomized experiments for causal inference.
- Missing outcomes threaten the validity of causal inference.
- A growing literature on the topic:
  - Method of bounds (e.g., Horowitz and Manksi, 2000).
  - Semiparametric methods (e.g., Scharfstein et al. 1999).
  - Ignorability (e.g., Yau and Little, 2001).
  - Latent ignorability (e.g., Frangakis and Rubin, 1999).
- Nonignorable missing outcomes:
  - Political science: self-reported voting behavior.
  - Economics: self-reported income.
  - Medicine: self-reported health status.
- The research project:
  - Alternative identification and estimation strategies.
  - With and without noncompliance.
  - New sensitivity analyses.
  - Applications in political science, psychology, and public health.

A Motivating Example: German Election Experiment

- Internet randomized experiment during the 2005 election.
  - Treatment group: asked if they intend to vote, whether in person or by mail, and the main obstacle they face.
  - Control group: asked if they intend to vote, but not how.
  - Outcome: self-reported turnout.
- Psychological theory on intentions (e.g., Gollwitzer, 1999):
  - *Goal intentions*: "I am going to vote!"
  - *Implementation intentions*: "Since I will be busy on the election day, I am going to vote by mail!"
  - Theoretical and empirical evidence: implementation intentions can more effectively increase the probability of achieving one's goal by automating goal implementation through anticipatory decisions (e.g., drug intake, breast self-examination, regular exercises).

Data and Nonresponse Problem

- Data:

	size	fraction of female	birth year (mean)	fraction of vote intenders	nonresponse rate
treatment	547	0.54	1970.86	0.94	0.20
control	572	0.54	1971.08	0.93	0.25

- Different nonresponse rates (statistically significant at 10% level using χ^2 test).
- Possibility of nonignorable nonresponse: the act of voting itself may increase their interest in politics and hence the probability of their participation in the post-election survey.

Framework for Standard Randomized Experiments

- Causal inference via potential outcomes (e.g., Holland 1986).
  - Experimental unit: $i = 1, 2, \ldots, n$.
  - Binary treatments: $T_i \in \{0, 1\}$.
  - Potential outcomes: $Y_i(T_i)$.
  - Observed outcome: $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$.
  - Potential response indicators: $R_i(T_i)$.
  - Observed response indicator: $R_i = T_i R_i(1) + (1 - T_i) R_i(0)$.
  - Pre-treatment covariates: X_i.
- No interference among units (Cox 1958; Rubin 1990).
- Randomized treatment: $(Y_i(1), Y_i(0), R_i(1), R_i(0)) \perp\!\!\!\perp T_i$ for all i.
- Estimands:
  - Average Treatment Effect (ATE):
    $\tau_{ATE} \equiv E[Y_i(1) - Y_i(0)] = E[Y_i \mid T_i = 1] - E[Y_i \mid T_i = 0]$.
  - Conditional Average Treatment Effect (CATE):
    $\tau_{CATE} \equiv \frac{1}{n} \sum_{i=1}^{n} E[Y_i(1) - Y_i(0) \mid X_i]$.

Identification Problem in the Binary Case

- Assume $Y_i(0), Y_i(1) \in \{0, 1\}$.
- Define,

$$
\begin{aligned}
p_{jk} &\equiv \Pr(Y_i = 1 \mid T_i = j, R_i = k), \\
\pi_{jk} &\equiv \Pr(T_i = j, R_i = k),
\end{aligned}
$$

- Then, the ATE can be written as,

$$
\tau_{ATE} = \frac{p_{10}\pi_{10} + p_{11}\pi_{11}}{\pi_{10} + \pi_{11}} - \frac{p_{00}\pi_{00} + p_{01}\pi_{01}}{\pi_{00} + \pi_{01}},
$$

  where p_{00} and p_{10} are not identifiable from the data.
- Since $p_{j0} \in [0, 1]$, the sharp bounds (Horowitz & Manski, 2000) are given by,

$$
\tau_{ATE} \in \left[\frac{p_{11}\pi_{11}(\pi_{00} + \pi_{01}) - (\pi_{00} + p_{01}\pi_{01})(\pi_{10} + \pi_{11})}{(\pi_{10} + \pi_{11})(\pi_{00} + \pi_{01})}, \right.
\left. \frac{(\pi_{10} + p_{11}\pi_{11})(\pi_{00} + \pi_{01}) - p_{01}\pi_{01}(\pi_{10} + \pi_{11})}{(\pi_{10} + \pi_{11})(\pi_{00} + \pi_{01})} \right].
$$

Identification Strategies

- **Ignorability Assumption** (Little & Rubin, 1987): the outcome variable is *missing at random* (MAR) given the treatment status and observed covariates. For $j \in \{0,1\}$ and $x \in \mathcal{X}$,

$$\begin{aligned} & \Pr(R_i(j) = 1 \mid T_i = j, Y_i(j) = 1, X_i = x) \\ = \; & \Pr(R_i(j) = 1 \mid T_i = j, Y_i(j) = 0, X_i = x), \end{aligned}$$

- The proposed assumption: missing-data mechanism directly depends on the realized value of the outcome variable itself, but is conditionally independent of the treatment status.
- Reasonable if the treatment does not *directly* cause nonresponse.
- **Nonignorability (NI) Assumption**: For $k \in \{0,1\}$ and $x \in \mathcal{X}$,

$$\begin{aligned} & \Pr(R_i(j) = 1 \mid T_i = 0, Y_i(0) = k, X_i = x) \\ = \; & \Pr(R_i(j) = 1 \mid T_i = 1, Y_i(1) = k, X_i = x). \end{aligned}$$

- Identification of the ATE is established via Bayes rule (PROPOSITION 1).

Inference under the Nonignorability Assumption

1. Without observed covariates (given a particular value of a covariate), the ML estimator of the ATE is available in a closed form (PROPOSITION 2).
2. A parametric approach with the covariates (estimation of CACE):
   - Specify the following parametric models (e.g., logistic regression),

$$\begin{aligned} q_j(x) &= \Pr(Y_i = 1 \mid T_i = j, X_i = x), \\ r_{jk}(x) &= \Pr(R_i = 1 \mid T_i = j, Y_i = k, X_i = x), \end{aligned}$$

   - Complete-data likelihood function:

$$\begin{aligned} & \prod_{i=1}^{n} \left[r_{\cdot 1}(X_i)^{R_i} \{1 - r_{\cdot 1}(X_i)\}^{1-R_i} \right]^{Y_i} \left[r_{\cdot 0}(X_i)^{R_i} \{1 - r_{\cdot 0}(X_i)\}^{1-R_i} \right]^{1-Y_i} \\ & \times \left[q_1(X_i)^{Y_i} \{1 - q_1(X_i)\}^{1-Y_i} \right]^{T_i} \left[q_0(X_i)^{Y_i} \{1 - q_0(X_i)\}^{1-Y_i} \right]^{1-T_i}, \end{aligned}$$

   where $r_{\cdot k}(x) = r_{1k}(x) = r_{0k}(x)$ for $x \in \mathcal{X}$ under the NI assumption.
   - Computation: *EM* algorithm, Gibbs sampler with prior distributions.

Multi-valued Outcome and Treatment Variables

- Setup:
  - J-valued treatment variable: $T_i \in \mathcal{T} \equiv \{0, 1, \ldots, J-1\}$.
  - K-valued outcome variable: $Y(T_i) \in \mathcal{Y} \equiv \{0, 1, \ldots, K-1\}$.
  - Average Treatment Effects: $\tau_{ATE}^{(j)} \equiv E[Y_i(j) - Y_i(j-1)]$.
- The NI assumption:

$$
\begin{aligned}
& \Pr(R_i(j) = 1 \mid T_i = j, Y_i(j) = k, X_i = x) \\
= \; & \Pr(R_i(j') = 1 \mid T_i = j', Y_i(j') = k, X_i = x).
\end{aligned}
$$

- Identification: there are $J(K-1)$ unknown probabilities while the assumption implies $J(J-1)K/2$ constraints. Thus, the identification is possible so long as $J \geq 3 - 2/K$.
- A general parametric approach: For example, we may assume,

$$
\Pr(R_i = 1 \mid T_i = j, Y_i = y, X_i = x) = \frac{\exp(\alpha + \beta y + \gamma x)}{1 + \exp(\alpha + \beta y + \gamma x)},
$$

  for every $j \in \mathcal{T}, x \in \mathcal{X}$, and $y \in \mathcal{Y}$.

Sensitivity Analysis with No Covariate

- Motivation: since neither MAR nor NI assumptions are directly verifiable from the data, it is of interest to examine the sensitivity of one's conclusion to the key identifying assumption.
- Sensitivity analysis based on the following parameter,

$$
\theta_k^{NI} \equiv \frac{\Pr(R_i(1) = 1 \mid T_i = 1, Y_i(1) = k)}{\Pr(R_i(0) = 1 \mid T_i = 0, Y_i(0) = k)},
$$

  for $k = 0, 1$ where the range of the parameter is given by,

$$
\begin{aligned}
\frac{(1 - p_{11})\pi_{11}}{(1 - p_{11})\pi_{11} + \pi_{10}} \leq \; & \theta_0^{NI} \leq \frac{(1 - p_{01})\pi_{01} + \pi_{00}}{(1 - p_{01})\pi_{01}}, \\
\frac{p_{11}\pi_{11}}{p_{11}\pi_{11} + \pi_{10}} \leq \; & \theta_1^{NI} \leq \frac{p_{01}\pi_{01} + \pi_{00}}{p_{01}\pi_{01}}.
\end{aligned}
$$

- τ_{ATE} is now a function of θ_k^{NI} and identifiable parameters.
- See how τ_{ATE} varies along with the value of θ_k.

Sensitivity Analysis with Observed Covariates

- Consider the following logistic regression:

$$\Pr(R_i = 1 \mid T_i = j, Y_i = k, X_i = x) \quad = \quad \frac{\exp(\alpha_{jk} + \beta x)}{1 + \exp(\alpha_{jk} + \beta x)},$$

- The sensitivity analysis can be based on the odds ratio for the conditional probabilities of missingness,

$$\Gamma_k^{NI} \quad = \quad \frac{r_{1k}(x;\eta_{1k})/[1 - r_{1k}(x;\eta_{1k})]}{r_{0k}(x;\eta_{0k})/[1 - r_{0k}(x;\eta_{0k})]} \quad = \quad \exp(\alpha_{1k} - \alpha_{0k}),$$

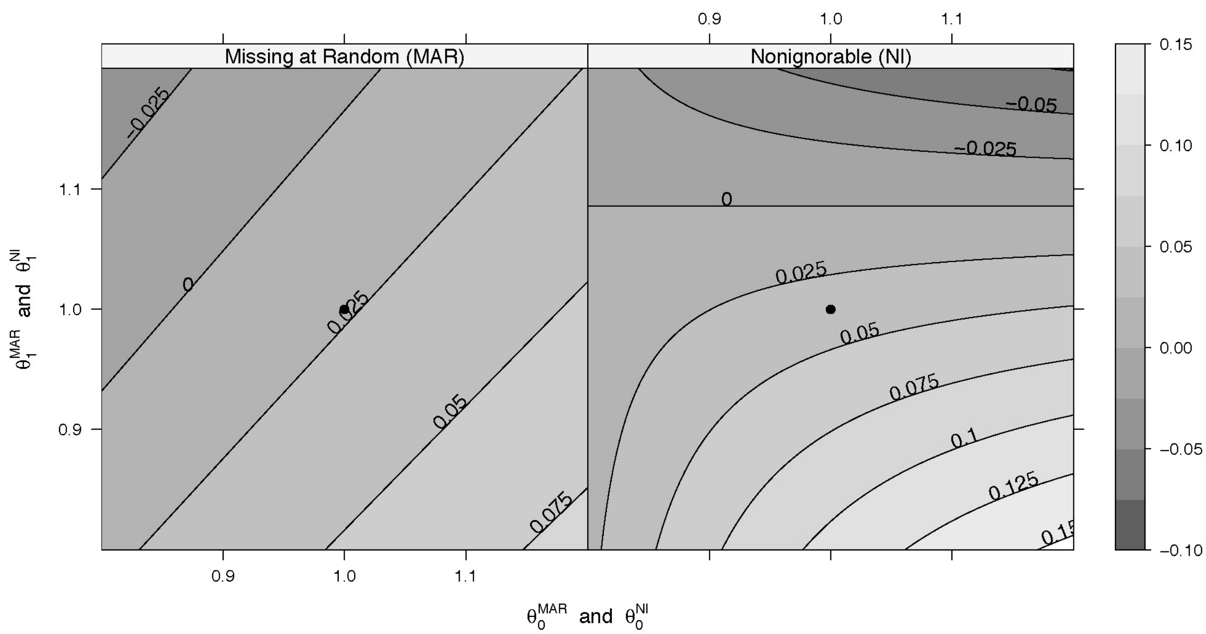
  where $\Gamma_k^{NI} \geq 0$ for $k \in \{0, 1\}$.
- Computation: *EM* algorithm with the following constraint $\alpha_{1k} = \log \Gamma_k^{NI} + \alpha_{0k}$, or Bayesian analysis incorporating this constraint.

Analysis of the German Election Experiment

- Model:
  1. Turnout model: $q_j(X_i) = \Pr(Y_i = 1 \mid T_i = j, X_i = x) = \exp(\alpha_j + x^\top \beta)/[1 + \exp(\alpha_j + x^\top \beta)]$.
  2. Response model: $r_{\cdot k}(X_i) = \Pr(R_i = 1 \mid Y_i = k, X_i = x) = \exp(\gamma_k + x^\top \delta)/[1 + \exp(\gamma_k + x^\top \delta)]$.
- ML estimates (using *EM* algorithm) with bootstrap standard errors.
- Results:

	point estimate	standard error	95% CI lower	95% CI upper
Missing at Random (MAR)				
No covariate	0.021	0.026	−0.030	0.073
With covariates	0.014	0.025	−0.035	0.063
Nonignorable (NI)				
No covariate	0.035	0.051	−0.049	0.119
With covariates	0.046	0.036	−0.011	0.129

Sensitivity Analysis without Covariates



Sensitivity Analysis with Covariates

- Results under the NI assumption:

	$\Gamma_1^{NI} = \frac{1}{3}$	$\Gamma_1^{NI} = 1$	$\Gamma_1^{NI} = 3$
$\Gamma_0^{NI} = \frac{1}{3}$	0.046 (0.027) $[-0.006, 0.100]$	0.003 (0.020) $[-0.032, 0.046]$	-0.075 (0.027) $[-0.128, -0.024]$
$\Gamma_0^{NI} = 1$	0.045 (0.029) $[-0.015, 0.097]$	0.046 (0.036) $[-0.011, 0.129]$	0.004 (0.039) $[-0.073, 0.080]$
$\Gamma_0^{NI} = 3$	0.134 (0.029) $[0.080, 0.192]$	0.047 (0.033) $[-0.020, 0.111]$	0.046 (0.028) $[-0.009, 0.101]$

- The ML estimates appear to be somewhat sensitive, but the scenarios corresponding to $(\Gamma_0^{NI}, \Gamma_1^{NI}) = (3, 1/3), (1/3, 3)$ may be highly unlikely.

Randomized Experiments with Noncompliance

- Randomized "encouragement" design:
  - Binary encouragement: $Z_i \in \{0, 1\}$.
  - Potential binary treatments: $T_i(Z_i) \in \{0, 1\}$.
  - Observed treatment: $T_i = Z_i T_i(1) + (1 - Z_i) T_i(0)$.
  - Potential outcomes: $Y_i(Z_i)$.
  - Observed outcome: $Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$.
  - Potential response indicators: $R_i(Z_i)$.
  - Observed response indicator: $R_i = Z_i R_i(1) + (1 - Z_i) R_i(0)$.
- Randomization of encouragement:

$$(Y_i(1), Y_i(0), T_i(1), T_i(0), R_i(1), R_i(0)) \perp\!\!\!\perp Z_i,$$

- Intention-To-Treat (ITT) effect: $\tau_{ITT} \equiv E[Y_i(T_i(1), 1) - Y_i(T_i(0), 0)]$.

Instrumental Variables (Angrist, Imbens & Rubin, 1996)

- Noncompliance
  - Complier: $T_i(1) = 1$ and $T_i(0) = 0$.
  - Noncomplier:
    1. Always-taker ($C_i = c$): $T_i(1) = T_i(0) = 1$.
    2. Never-taker ($C_i = n$): $T_i(1) = T_i(0) = 0$.
    3. Defier ($C_i = d$): $T_i(1) = 0$ and $T_i(0) = 1$.
- Assumptions:
  1. Monotonicity (no defier): $T_i(1) \geq T_i(0)$.
  2. Exclusion restriction for noncompliers: $Y_i(1) = Y_i(0)$ for $C_i = a, n$ (i.e., zero ITT effect for always-takers and never-takers).
- Complier Average Causal Effect (IV estimand):

$$\tau_{CACE} \equiv E[Y_i(1) - Y_i(0) \mid C_i = c] = \frac{E[Y_i(1) - Y_i(0)]}{E[T_i(1) - T_i(0)]}.$$

Identification Strategies

- **Ignorability** (Yau & Little, 2001): For $j = 0, 1$ and $l = 0, 1$,

$$\begin{aligned} & \Pr(R_i(l) = 1 \mid Y_i(l) = 1, T_i(l) = j, Z_i = l, X_i = x) \\ = \; & \Pr(R_i(l) = 1 \mid Y_i(l) = 0, T_i(l) = j, Z_i = l, X_i = x). \end{aligned}$$

- **Latent Ignorability** (Frangakis & Rubin, 1999):
  1. Latent ignorability: For $l = 0, 1$ and $t \in \{c, n, a\}$,

$$\begin{aligned} & \Pr(R_i(l) = 1 \mid Y_i(l) = 1, Z_i = l, C_i = t, X_i = x) \\ = \; & \Pr(R_i(l) = 1 \mid Y_i(l) = 0, Z_i = l, C_i = t, X_i = x). \end{aligned}$$

  2. Compound exclusion restriction for noncompliers: $Y_i(0) = Y_i(1)$, and $R_i(1) = R_i(0)$, for $C_i = n, a$.
- **Nonignorability**: For $j = 0, 1$, and $k = 0, 1$,

$$\begin{aligned} & \Pr(R_i(1) = 1 \mid T_i(1) = j, Y_i(1) = k, Z_i = 1, X_i = x) \\ = \; & \Pr(R_i(0) = 1 \mid T_i(0) = j, Y_i(0) = k, Z_i = 0, X_i = x). \end{aligned}$$

Theoretical Results in the Binary Case

- Apply the same analytical strategy as before.
- Define,

$$\begin{aligned} p_{jkl} & \equiv \Pr(Y_i = 1 \mid T_i = j, R_i = k, Z_i = l), \\ \pi_{jkl} & \equiv \Pr(T_i = j, R_i = k, Z_i = l). \end{aligned}$$

- Rewrite the ITT effect as,

$$\tau_{ITT} = \frac{\sum_{j=0}^{1} \sum_{k=0}^{1} p_{jk1} \pi_{jk1}}{\sum_{j=0}^{1} \sum_{k=0}^{1} \pi_{jk1}} - \frac{\sum_{j=0}^{1} \sum_{k=0}^{1} p_{jk0} \pi_{jk0}}{\sum_{j=0}^{1} \sum_{k=0}^{1} \pi_{jk0}},$$

  where π_{jkl} and p_{j1l} are identifiable, but p_{j0l} is not.
- Thus, the identification of τ_{ITT} requires four constraints (Proposition 3).

Concluding Remarks

- Missing outcomes in randomized experiments are frequently encountered in practice.
- Possibility of nonignorable missing-data mechanism.
- Identification and estimation strategies are proposed for:
  - standard randomized experiments.
  - randomized experiments with noncompliance.
- The proposed sensitivity analyses are useful to examine the robustness of one's conclusion.
  - The method of bounds gives the identification region without any assumption.
  - The assumptions such as MAR and NI are not directly identifiable from the observed data, but point-identify the quantity of interest.
  - Sensitivity analysis complements these two approaches.