Bayesian Inference for Randomized Experiments with Noncompliance and Nonignorable Missing Data

Andrea Mercatanti

Bank of Italy
Economic and Financial Statistics Department

Durham, 5 May 2014

Introduction

- A range of models for randomized experiments with noncompliance and nonignorable missing data, in the setting of binary assignment, binary treatment received, and binary outcomes is proposed.
- The conditions for model identification stem from the analysis of the contingency table for the observable data.
- Identified models are proposed for three scenarios: missingness in all the three variables, with and without a binary pretreatment variable, and missingness in the outcome and in the treatment received without pretreatment variables.
- A Bayesian approach is developed for inference. The method is illustrated by a simulated comparative example.

Introduction

- The theoretical model of a randomized experiment with noncompliance is widely adopted in biomedical and social sciences for inferring causal effects.
- The original application stems from non-coercive experimental settings, where the treatment received by an individual can differ from the treatment randomly assigned to him by the experimenter.
- The concept was later generalized to observational settings where the treatment is not randomized.
- The nonrandom nature of the treatment implies different distributions of the potential outcomes between treatment groups: self-selection.

Introduction

- Under the Potential Outcome Approach (Rubin; 1974, 1978) the Average Treatment Effect, ATE, is defined as the average difference between the outcomes potentially observed under the range of possible values for the assignment to treatment.

$$ATE = E(Y(D = 1) - Y(D = 0)).$$

- Self-selection makes the comparison of the outcome distributions between treatment groups a biased estimate for the treatment effect.
- One approach to obtain unbiased estimates for causal effects is to find a variable that can be viewed as a nonmanipulated, natural, random assignment to the treatment to which the units not necessarily comply, and then to apply inferential methods for randomized experiment with noncompliance.

Introduction

Examples of nonmanipulated assignment to treatments in the literature on return to schooling (effect of education on earnings):

- Coorth of birth (Ichino and Winter-Ebmer, 2004)
- College proximity (Card, 1995; Kling, 2001)

Introduction

- When the model of an experiment with noncompliance is adopted in nonexperimental studies, information is generally collected from surveys where missing data due to item nonresponses to questionnaires is a diffuse problem.
- One type of information that has traditionally been difficult to obtain in survey is income data. For example: the U.S. Current Population Surveys suffers from 20 to 25% of nonresponse rate on many income items.
- Other type of information with high rate on nonresponse are those related to stigmatized activities like drinking behavior or use of drug.
- More generally, nonresponses are common in surveys whenever the population consists of units such as individual people, households or businesses.

Introduction

- In general, standard methods for complete data cannot be immediately used to analyze the dataset with missing data.
- Moreover, possible biases can arise because the respondents are often systematically different from the nonrespondents, and these biases are difficult to eliminate since the reasons for nonresponse are usually not known.
- The existing literature on missing data in randomized experiment with noncompliance has mainly focused on the scenarios with missingness only in the outcome.
- However, missing data in the treatment, and/or in the assignment to treatment are common in practice, where usual ignorability conditions for the missing data model are considered too restrictive.

Notation

- Z : the assignment to a binary treatment
  D : the treatment
  Y : the binary outcome
- $[Y(Z=0), Y(Z=1), D(Z=0), D(Z=1)]$: the potential quantities
- The population can be classified into four sub-groups, compliance types, based on their potential treatment status to both levels of assignment: $U = (D(0), D(1))$.

Notation

- $U = a$ (always-takers): $D(Z = 1) = D(Z = 0) = 1$: units who always take the treatment;
- $U = n$ (never-takers): $D(Z = 1) = D(Z = 0) = 0$: units who never take the treatment;
- $U = c$ (compliers): $D(Z = 1) = 1, D(Z = 0) = 0$: units who are induced to take the treatment by the assignment;
- $U = d$ (defiers): $D(Z = 1) = 0, D(Z = 0) = 1$: units who do the opposite of the assignment.

Notation

- To account for nonresponses we introduce the additional notation.
- **R** is the vector (R_y, R_d, R_z), with R_τ the binary missing data indicator for $\tau = y, d, z$:
  $R_\tau = 1$ if τ is observed, 0 otherwise.
- D^*, Z^*, and Y^* are the observable quantities:
  - $D^* = D$ if $R = 1$, $D^* = *$ if $R = 0$
  - $Z^* = Z$ if $R = 1$, $Z^* = *$ if $R = 0$
  - $Y^* = Y$ if $R = 1$, $Y^* = *$ if $R = 0$
- **X** is the set (Y, D, Z)
  - $\mathbf{X}_{obs}$ the observed part of **X**
  - $\mathbf{X}_{mis}$ the unobserved part of **X**

Missing Data Models

- Missing Completely At Random: $P(\mathbf{R}|\mathbf{X}) = P(\mathbf{R})$
  that is the probability of missingness is the same for each unit.
- Missing At Random: $P(\mathbf{R}|\mathbf{X}) = P(\mathbf{R}|\mathbf{X}_{obs})$
  the probabilities of missingness depends only on the observed part of $\mathbf{X}$.
- Non-ignorable: $P(\mathbf{R}|\mathbf{X}) \neq P(\mathbf{R}|\mathbf{X}_{obs}) \neq P(\mathbf{R}|\mathbf{X}_{mis}) \neq P(\mathbf{R})$

Main assumptions

We maintain hereafter the standard assumptions to identify causal effects in randomized experiments with noncompliance (Imbens and Rubin, 1997).

A.1 *Stable unit treatment value assumption* (SUTVA). There are no interference between units.

A.2 *Randomization of treatment assignment*. $P(Z|Y(0), Y(1), D(0), D(1)) = P(Z)$.

A.3 *Monotonicity*. $D_i(1) \geq D_i(0)$ for all i, ruling out the group of defiers.

A.4 *Exclusion restriction* (ER). $Y_i(1) = Y_i(0)$ for all noncompliers, implying that the assignment to treatment has no direct effect on the outcome for noncompliers.

Identification

Under assumptions A.1-A.4 the model for the observable data $(\mathbf{R},\ Y^*,\ D^*,\ Z^*)$ can be written as:

$$
\begin{aligned}
& P(\mathbf{R}, Y^*, D^*, Z^*) \\
= & \sum_{Y,D,Z} P(\mathbf{R}, Y, D, Z) \cdot I[P(Y, D, Z | Y^*, D^*, Z^*) > 0] \\
= & \sum_{Y,U,Z} P(\mathbf{R}, Y, U, Z) \cdot I[P(Y, U, Z | Y^*, D^*, Z^*) > 0] \\
= & \sum_{Y,U,Z} P(\mathbf{R} | Y,\ U,\ Z) \cdot P(Y,\ U,\ Z) \cdot \\
& I[P(Y,\ U,\ Z \mid Y^*,\ D^*,\ Z^*) > 0].
\end{aligned}
$$

Identification

- Dealing with binary variables allows to relate the parameter identification to the analysis of the contingency table for the observable data (Y^*, D^*, Z^*).
- The identification rule to apply is the general one for contingency tables that prescribed to set the number of parameters for regular models at most equal to the number of the entries of the contingency table minus one.

Identification

- $X_1 : \{0, 1\}; X_2 : \{0, 1\}$

	$X_2 = 1$	$X_2 = 0$
$X_1 = 1$	$1, 1$	$1, 0$
$X_1 = 0$	$0, 1$	$0, 0$

- $logit(1, 1) = \alpha_1$
  $logit(1, 0) = \alpha_2$
  $logit(0, 1) = \alpha_3$
- $logit(X_1 = 1) = \alpha$
  $logit(X_2 = 1 | X_1) = \beta_1 + \beta_2 X_1$

Missingness in Y, D, and Z without covariates

When the possibility of nonresponses exists for each of the three binary variables, Y, D, and Z, then the contingency table for the observables Y^*, D^*, and Z^*, will show $3 \times 9 = 27$ cells in total, and consequently at most $27 - 1 = 26$ parameters will be allowed for $P(\mathbf{R}, Y^*, D^*, Z^*)$.

		D^*, Z^*								
		$1,0$	$0,1$	$1,1$	$0,0$	$1,*$	$0,*$	$*,1$	$*,0$	$*,*$
Y^*	1									
	0									
	$*$									

Missingness in Y, D, and Z without covariates

The specification for $P(Y, U, Z)$ from Imbens and Rubin (1997) has 7 parameters:

$$
\begin{aligned}
& P(Y,\ U,\ Z;\ \pi, \omega, \theta) \\
= \quad & \pi^{Z}\,(1-\pi)^{1-Z} \\
& {\omega_a}^{I(U=a)}\,{\omega_n}^{I(U=n)}(1-\omega_a-\omega_n)^{I(U=c)} \\
& \times {\theta_a}^{Y\,I(U=a)}\,(1-\theta_a)^{(1-Y)\,I(U=a)} \\
& \times {\theta_n}^{Y\,I(U=n)}\,(1-\theta_n)^{(1-Y)\,I(U=n)} \\
& \times {\theta_{c1}}^{Y\,I(U=c)\,Z}\,(1-\theta_{c1})^{(1-Y)\,I(U=c)\,Z} \\
& \times {\theta_{c0}}^{Y\,I(U=c)\,(1-Z)}(1-\theta_{c0})^{(1-Y)\,I(U=c)\,(1-Z)}.
\end{aligned}
$$

$P(Y, U, Z)$ shows 7 parameters so that at most $26 - 7 = 19$ are allowed for $P(\mathbf{R}|\ Y,\ U,\ Z)$ whose domain is composed by the 8 possible combinations $(R_Y,\ R_D,\ R_Z) : \{(1,1,1),\ (1,1,0),\ ...\}$.

Missingness in Y, D, and Z without covariates

- Any attempt to comply with the limit of 19 parameters via a multinomial logit model for the 8 categories of the missing data model leads to very few parameters for each logit equation and consequently to unplausible models.
- We impose that the three marginal missing data model $P(R_v|Y,U,Z)$, $v = Y, D, Z$ are mutually independent:

$$\begin{aligned} & P(\mathbf{R}|Y,U,Z) \\ = \; & P(R_Y|Y,U,Z) \cdot P(R_D|Y,U,Z) \cdot P(R_Z|Y,U,Z). \end{aligned}$$

Missingness in Y, D, and Z without covariates

The following model specification for $P(\mathbf{R}|Y, U, Z)$ implies identifiability of $(\mathbf{R}, Y^*, D^*, Z^*)$ in that it complies with the restriction on the maximum number of parameters:

$$\begin{aligned} & \text{logit}[P(R_v = 1|Y, U, Z)] \\ = \; & \alpha_{0v} + \alpha_{1v} I(U = a) + \alpha_{2v} I(U = n) + \alpha_{3v} I(U = a, Z = 1) \\ & + \alpha_{4v} I(U = n, Z = 1) + \alpha_{5v} I(U = c, Z = 1), \end{aligned}$$

for $v = D, Z$, and

$$\begin{aligned} & \text{logit}[P(R_Y = 1|Y, U, Z)] \\ = \; & \alpha_{0Y} + \alpha_{1Y} I(U = a) + \alpha_{2Y} I(U = n) + \alpha_{3Y} I(U = a, Z = 1) \\ & + \alpha_{4Y} I(U = n, Z = 1) + \alpha_{5Y} I(U = c, Z = 1) + \alpha_{6Y} Y. \end{aligned}$$

Missingness in Y, D, and Z without covariates

- The model reflect nonignorable conditions for the missing data because the probability of missingness for τ is affected by the value of τ itself.
- The proposed model is based on weaker conditions than other proposals recently appeared in the literature for dealing with missingness only on the outcome:

- Compared to Frangakis and Rubin (1999) no kind of response exclusion restrictions is imposed.
- Compared to Frangakis and Rubin (1999) and Mealli et al. (2004), it does not imposes ignorability of Y conditionally on the compliance status.
- Compared to Imai (2009), it is based on weaker conditions because the way that treatment affects missingness depends on the assignment to treatment.
- Given that the dependency between missingness and the assignment to treatment depends on the treatment received, the model is weaker than Small and Cheng (2008).

Missingness in Y, D, and Z with a binary pretreatment variable

- The double need to comply with the maximum number of parameters to ensure $P(\mathbf{R}, Y^*, D^*, Z^*)$ is identifiable and to avoid the proposal of unplausible models directed us to restrict the three marginal missing data models $P(R_v|Y, U, Z)$, $v = Y, D, Z$, to be mutually independent.
- The restriction can be relaxed by introducing a binary and always-observed pretreatment variable X. This implies a larger contingency table for the observables and a large numbers of parameters allowed for $P(\mathbf{R}, Y^*, D^*, Z^*)$ to be identified.
- The contingency table for (Y^*, D^*, Z^*) has 3 rows and 18 columns that permit at most $(3 \times 18) - 1 = 53$ parameters for $P(\mathbf{R}, Y^*, D^*, Z^*, X)$.

Missingness in Y, D, and Z with a binary covariate

Following Chen et al. (2008) we assume that X does not enter in the missing data model so that:
$P(\mathbf{R},\ Y^*,\ D^*,\ Z^*,\ X) = P(\mathbf{R} \mid Y,\ U,\ Z) \cdot P(Y, U, Z, X)$, where

$$
\begin{aligned}
& P(Y, U, Z, X) \\
= \quad & \pi^{Z} (1-\pi)^{1-Z} \, \omega_{1a}{}^{I(X=1,\,a)} \, \omega_{0a}{}^{I(X=0,\,a)} \, \omega_{1n}{}^{I(X=1,\,n)} \, \omega_{0n}{}^{I(X=0,\,n)} \\
& \times \omega_{1c}{}^{I(X=1,\,c)} \, (1 - \omega_{1a} - \omega_{0a} - \omega_{1n} - \omega_{0n} - \omega_{1c})^{I(X=0,\,c)} \\
& \times \theta_{1a}{}^{Y\, I(X=1,\,a)} \, \theta_{0a}{}^{Y\, I(X=0,\,a)} \\
& \times (1-\theta_{1a})^{(1-Y)\, I(X=1,\,a)} \, (1-\theta_{0a})^{(1-Y)\, I(X=0,\,a)} \\
& \times \theta_{1n}{}^{Y\, I(X=1,\,n)} \, \theta_{0n}{}^{Y\, I(X=0,\,n)} \\
& \times (1-\theta_{1n})^{(1-Y)\, I(X=1,\,n)} \, (1-\theta_{0n})^{(1-Y)\, I(X=0,\,n)} \\
& \times \theta_{1c1}{}^{Y\, I(X=1,\,c)\, Z} \, \theta_{0c1}{}^{Y\, I(X=0,\,c)\, Z} \\
& \times \ldots
\end{aligned}
$$

Missingness in Y, D, and Z with a binary covariate

- Given the larger number of free parameters, $P(\mathbf{R}|Y, U, Z)$ can be now modeled as a multinomial logit for the 8 combinations $\mathbf{R} = (R_Y, R_D, R_Z)$. Taking into account that 14 parameters are involved in $P(Y, U, Z, X)$, the following two plausible models for $P(\mathbf{R}|Y, U, Z)$ can be delineated.
- The first one complies with the natural choice for a nonignorable model for missing data, that is the simple dependency on the values of the variables subjected to missingness:

$$
\begin{aligned}
& \text{logit}[P(R_Y,\ R_D,\ R_Z)] \\
= \quad & \alpha_{0R_yR_dR_z} + \alpha_{1R_yR_dR_z}\ Y \\
& + \alpha_{2R_yR_dR_z}\ I(U = a,\ Z = 0) + \alpha_{3R_yR_dR_z}\ I(U = n,\ Z = 1) \\
& + \alpha_{4R_yR_dR_z}\ [I(U = a,\ Z = 1) + I(U = c,\ Z = 1)].
\end{aligned}
$$

Missingness in Y, D, and Z with a binary covariate

To note that the treatment D indirectly enters into $\text{logit}[P(R_Y, R_D, R_Z)]$ by the compliance status that is a function of the couple of potential quantities $D(Z = z)$. The logit can be coherently rewritten as:

$$\begin{aligned} & \text{logit}[P(R_Y,\, R_D,\, R_Z)] \\ = \;& \alpha_{0R_yR_dR_z} + \alpha_{1R_yR_dR_z}\, Y + \alpha_{2R_yR_dR_z}\, I(D = 1,\, Z = 0) \\ & + \alpha_{3R_yR_dR_z}\, I(D = 0,\, Z = 1) + \alpha_{4R_yR_dR_z}\, I(D = 1,\, Z = 1). \end{aligned}$$

Missingness in Y, D, and Z with a binary covariate

The second specification imposes response exclusion restrictions for two compliance statuses; for example, response exclusion restriction for noncompliers:

$$\begin{aligned} & \text{logit}[P(R_Y, R_D, R_Z)] \\ = \; & \alpha_{0R_yR_dR_z} + \alpha_{1R_yR_dR_z} Y + \alpha_{2R_yR_dR_z} I(U = a) \\ & + \alpha_{3R_yR_dR_z} I(U = n) + \alpha_{4R_yR_dR_z} I(U = c, Z = 1). \end{aligned}$$

Missingness in Y and D without covariates

- In economic and social sciences the model of a randomized experiment with noncompliance is often used in observational studies, where most common choices for the natural assignment to treatment Z are registry information (such as date of birth, location, etc.), usually unaffected by the problem of nonresponses.
- A plausible assumption is that there is no missingness in the treatment assignment: $P(R_z = 1) = 1$.
- The contingency table for the observable data now allows $(3 \times 6) - 1 = 17$ free parameters for $P(\mathbf{R}, Y^*, D^*, Z^*, X)$.

Missingness in Y and D without covariates

- $P(Y, U, Z)$ remains as previously specified, while an identified missing data model can be specified by a multinomial logit model where each logit depends on the values of the two variables subjected to missingness Y and D:

$$\begin{aligned} & \text{logit}[P(R_Y, R_D)] \\ = \quad & \alpha_{0R_yR_d} + \alpha_{1R_yR_d}\, Y + \alpha_{2R_yR_d}\, [I(U = a) + I(U = c, Z = 1)]. \end{aligned}$$

- Again, the treatment indirectly enters into $\text{logit}[P(\mathbf{R}|Y, U, Z)]$ by the compliance status so that the logits can be rewritten as:

$$\text{logit}[P(R_Y, R_D)] = \alpha_{0R_yR_d} + \alpha_{1R_yR_d}\, Y + \alpha_{2R_yR_d}\, D.$$

Estimation: Bayesian method

- In principle, a ML estimation via the EM algorithm could be attractive also in case of nonignorable condition for the missing data mechanism because if Y, U and Z were known for all units, $P(Y, U, Z)$ would not involve mixture components.
- However, and contrary to the cases of ignorable missing data mechanisms, the augmented log-likelihood function would not be linear in the missing information.
- Consequently the EM algorithm could not work by simply filling-in missing data and then updating the parameter estimates, so that the expected augmented log-likelihood function should be computed at each iteration of the algorithm.
- It is computationally more convenient to conduct Bayesian inference.

Simulations

Model with nonresponses in Y and D.

- The simulation study shows the biases we incurred when simpler but wrong models such as those based on MAR or MCAR are applied.
- The posterior distribution can be sensitive to the choice of prior distribution, because of the mixture-structure of the observed likelihood. As shown by Hirano et al. (2000) in a similar context, standard diffuse improper prior can lead to improper posteriors.
- We adopt a proper prior distribution that correspond to adding 48 observations: one for each of the 48 combinations of the variables (Z, U, Y, R_Y, R_D). Example: for $(Z = 1, U = n, Y = 0, R_Y = 1, R_D = 1)$ the following term is added to the likelihood:

$$\pi \cdot \omega_n \cdot (1-\theta_n) \cdot \exp(\alpha_{011}) \cdot (1+\exp(\alpha_{011})+\exp(\alpha_{010})+\exp(\alpha_{001}))^{-1}.$$

Simulations

- The same arguments applies to the priors for the analysis under MAR and MCAR leading to 12 and 30 added observations to the likelihood respectively.
- MCMC:
  - the Gibbs is not particularly attractive given there are no conjugate forms for the prior distribution and no standard forms for the conditional posterior distributions under nonignorable missing data.
  - Metropolis-Hastings algorithm: OK! MH: a general term for a family of Markov Chain method useful for drawing samples from posterior distributions. It is an adaptation of a random walk that uses an acceptance/rejection rule to converge to the specified posterior.

How does the MH algorithm works?

Sample a proposal θ^c from a jumping distribution, $J(\theta^c \mid \theta_{t-1})$, at time t.

Calculate the quantity:

$$r = \frac{P(\theta^c|\mathbf{X})}{P(\theta_{t-1}|\mathbf{X})} \frac{J(\theta_{t-1} \mid \theta^c)}{J(\theta^c \mid \theta_{t-1})}.$$

If $J(\cdot)$ symmetric, then $\frac{J(\theta_{t-1} \mid \theta^c)}{J(\theta^c \mid \theta_{t-1})} = 1$ and $r = \frac{P(\theta^c|\mathbf{X})}{P(\theta_{t-1}|\mathbf{X})}$.

Set $\theta_t = \theta^c$ with probability $\min(r, 1)$.

Simulations: the hypothetical population

- Hypothetical population: parameters values.

$\pi = P(Z = 1)$	0.50	α_{011}	-2
$\omega_a = P(U = a)$	0.30	α_{111}	2
$\omega_n = P(U = n)$	0.45	α_{211}	0.5
$1 - \omega_a - \omega_n = P(U = c)$	0.25	α_{010}	0.5
μ_a	0.90	α_{110}	1.5
μ_n	0.20	α_{210}	1.5
μ_{c1}	0.70	α_{001}	-2
μ_{c0}	0.40	α_{101}	1.5
$\mu_{c1} - \mu_{c0}$	0.30	α_{201}	2

Simulations: the hypothetical population

- Hypothetical population: complete set of conditional probabilities of missingness.

$P(Ry = 1, Rd = 1 \mid Y = 1, D = 1)$	0.823
$P(Ry = 1, Rd = 1 \mid Y = 1, D = 0)$	0.739
$P(Ry = 1, Rd = 1 \mid Y = 0, D = 1)$	0.769
$P(Ry = 1, Rd = 1 \mid Y = 0, D = 0)$	0.565
$P(Ry = 1, Rd = 0 \mid Y = 1, D = 1)$	0.040
$P(Ry = 1, Rd = 0 \mid Y = 1, D = 0)$	0.100
$P(Ry = 1, Rd = 0 \mid Y = 0, D = 1)$	0.023
$P(Ry = 1, Rd = 0 \mid Y = 0, D = 0)$	0.046
$P(Ry = 0, Rd = 1 \mid Y = 1, D = 1)$	0.111
$P(Ry = 0, Rd = 1 \mid Y = 1, D = 0)$	0.061
$P(Ry = 0, Rd = 1 \mid Y = 0, D = 1)$	0.104
$P(Ry = 0, Rd = 1 \mid Y = 0, D = 0)$	0.046

Simulations: the hypothetical population

- Hypothetical population: marginal and conditional probabilities of missingness.

$P(R_Y = 0)$	0.2544
$P(R_D = 0)$	0.2292
$P(R_D = 0 \mid D = 0)$	0.3429
$P(R_D = 0 \mid D = 1)$	0.0755
$P(R_Y = 0 \mid Y = 0)$	0.3646
$P(R_Y = 0 \mid Y = 1)$	0.1431

Simulations

- A sample of size 5000 was drawn from the hypothetical population. Four Markov Chains each of 200,000 iterations were run. Starting values were drawn from an overdisperded normal distribution. Parameters were updated in batch with a Acceptance Rate of $\approx 20\%$.
- Convergence assessed by the potential scale reduction indicator, $\hat{R}$, proposed by Gelman et al. (2004):

$$\hat{R} = \sqrt{\frac{W + (B/n)}{W}} < 1.1 \text{ for each parameter.}$$

- Posterior inferences from the second half of the simulated sequences.

Simulations:results

Posterior means and standard deviations for some parameters from the application of the correct model alongside those calculated under MAR and MCAR.

		Nonignorable		MAR		MCAR	
π	0.50	0.506	(0.007)	0.506	(0.020)	0.521	(0.009)
ω_a	0.30	0.328	(0.015)	0.379	(0.025)	0.369	(0.012)
ω_n	0.45	0.435	(0.017)	0.387	(0.032)	0.396	(0.012)
ω_c	0.25	0.237	(0.014)	0.232	(0.038)	0.234	(0.016)
$\mu_{c1} - \mu_{c0}$	0.30	0.287	(0.071)	0.239	(0.075)	0.225	(0.063)

Simulations: results

Conditional probabilities of the missing data indicators: real values alongside the posterior means and standard deviations from the application of the correct model.

	Real value	Posterior value	
$P(Ry = 1, Rd = 1 \mid Y = 1, D = 1)$	0.823	0.794	(0.028)
$P(Ry = 1, Rd = 1 \mid Y = 1, D = 0)$	0.739	0.706	(0.069)
$P(Ry = 1, Rd = 1 \mid Y = 0, D = 1)$	0.769	0.730	(0.057)
$P(Ry = 1, Rd = 1 \mid Y = 0, D = 0)$	0.565	0.588	(0.044)
$P(Ry = 1, Rd = 0 \mid Y = 1, D = 1)$	0.041	0.047	(0.009)
$P(Ry = 1, Rd = 0 \mid Y = 1, D = 0)$	0.100	0.077	(0.019)
$P(Ry = 1, Rd = 0 \mid Y = 0, D = 1)$	0.023	0.033	(0.013)
$P(Ry = 1, Rd = 0 \mid Y = 0, D = 0)$	0.046	0.052	(0.006)
$P(Ry = 0, Rd = 1 \mid Y = 1, D = 1)$	0.111	0.095	(0.012)
$P(Ry = 0, Rd = 1 \mid Y = 1, D = 0)$	0.061	0.044	(0.017)
$P(Ry = 0, Rd = 1 \mid Y = 0, D = 1)$	0.104	0.130	(0.046)
$P(Ry = 0, Rd = 1 \mid Y = 0, D = 0)$	0.046	0.056	(0.010)

Conclusions

- Based on the analysis of contingency tables, we have investigated the identification issue in the models for studies with nonignorable missing data in the response, as well as in the assignment to treatment and/or in the treatment received.
- Simulation results suggest that the causal estimates were sensitive to the assumption of the missing data model, which merits special attention in practice.
- Extentions to:
  - incorporate continuous outcomes.
  - sensitivity analysis regarding possible deviations from the identifiability conditions and varying proportions of missing data.

- Thank you for your attention!