

Exploiting multivariate outcomes in Bayesian inference for causal effects with noncompliance

Alessandra Mattei, Fabrizia Mealli and Barbara Pacini

Abstract A Bayesian approach to causal inference in the presence of noncompliance to assigned randomized treatment is considered. It exploits multivariate outcomes for improving estimation of weakly identified models. We maintain the monotonicity of compliance assumption, while relaxing the usually invoked exclusion restriction assumption for never-takers. Using artificial data sets, we analyze the properties of the posterior distribution of causal estimands to evaluate the potential gains of jointly modelling more than one outcome. The approach can be used to assess robustness with respect to deviations from structural identifying assumptions. It can also be extended to the analysis of observational studies with instrumental variables where exclusion restriction assumptions are usually questionable.

Key words: Bayesian statistics, Causal inference, Principal stratification, Multivariate outcome, Noncompliance

1 Introduction

Randomized experiments are generally considered the gold standard for causal inference. Yet even randomized experiments may suffer from a number of complications. A common complication when drawing inference about causal effects is that compliance is rarely perfect: noncompliance occurs when the actual treatment that units receive differs from their nominal assignment. Noncompliance is relatively

Alessandra Mattei

Dipartimento di Statistica “G. Parenti”, Università di Firenze e-mail: mattei@ds.unifi.it

Fabrizia Mealli

Dipartimento di Statistica “G. Parenti”, Università di Firenze e-mail: mealli@ds.unifi.it

Barbara Pacini

Dipartimento di Statistica e Matematica, Università di Pisa e-mail: barbara.pacini@sp.unipi.it

frequent in experiments involving human subjects, complicating traditional theories of inference that require adherence to the random treatment assignment. A standard approach is to estimate the intention-to-treat (ITT) effect, focusing on the causal effect of assignment of treatment rather than the causal effect of receipt of treatment. Nevertheless, interest often centers on the effect of the treatment itself, and the interpretation of an ITT analysis is sometimes based on an implicit assumption, which is generally but not always plausible (Hirano et al., 2000), that the effect of assignment is indicative of the effect of the treatment.

Randomized experiments with noncompliance are closely related to econometrics instrumental variable (IV) settings, where the instrument plays the parallel role of treatment assignment.

Angrist et al. (1996) show that causal effects of assignment among compliers (subjects who would take treatment if offered and would not take it if not offered) are identifiable under the assumption that there are no subjects who would take the active treatment if randomized to the control arm but would not take it if assigned to the treatment arm (no-defier, monotonicity assumption) and under the additional assumption of a null effect of assignment on the outcome for those whose treatment status is not affected by assignment (exclusion restriction). The effect on compliers is usually interpreted as the effect of receipt of the treatment, under the implicit assumption that the effect of assignment is solely due to actual treatment receipt.

The stratification of units in subpopulations characterized by their potential compliance is a special case of Principal Stratification (PS, Frangakis and Rubin, 2002), which is a more general framework to deal with *adjusting* for post-treatment variables.

Principal strata are a latent classification of units and are defined by the joint values of one (or more) post-treatment variable under all possible levels of treatment assignment.

Identification of causal effects for particular strata is usually possible only under some assumptions aimed at reducing the number of strata (such as monotonicity assumptions) or at imposing certain features of the distribution of outcomes within or among strata (such as exclusion restrictions). These assumptions essentially allow one to uniquely disentangle the observed distribution of outcomes into a mixture of distributions associated to the latent strata (Imbens and Rubin, 1997a).

Depending on the empirical setting, some of these substantive assumptions may be questionable. For example, the exclusion restriction appears plausible in blind or double-blind experiments, but less so in open-label ones.

These structural assumptions can be relaxed by using a fully Bayesian model, thereby avoiding the need for full identification. In encouragement designs with compliance as the only partially uncontrolled factor, and where there are full outcome data, Imbens and Rubin (1997b) extended the Bayesian approach to causal inference of Rubin (1978) to handle simple randomized experiments with noncompliance. Hirano et al. (2000) further extended the approach to handle fully observed covariates and considering the consequences of violations of exclusion restrictions by the comparison of results based on different combinations of these assumptions. Without these assumptions, inference, although straightforward in the Bayesian ap-

proach, can be imprecise even in large samples. Models may result as *weakly identified*: they have a proper posterior distribution, but do not have unique maximum likelihood estimates.

Adding information may be pursued as a general strategy for identification purposes. Covariates, for example, generally improve the precision of causal estimates (i.e., reduce posterior variability) because they improve the prediction of missing potential outcomes. In addition they can be used to achieve full identification when some of the usually invoked assumptions are relaxed. For example, Jo (2002) and Frangakis (2006) relax the exclusion restriction, and show how identification can be achieved by imposing some plausible behavioral hypotheses within or among groups defined by the values of the covariates, which translate into restrictions on coefficients (exclusion of some interaction terms or equality of coefficients across strata for some covariates).

Here, we investigate the role of information coming from additional outcome variables, which may help reducing the uncertainty about the treatment effects on the outcome of primary interest. We concentrate on the simple case of a binary treatment assignment, to which the units comply or do not comply (the compliance behavior is assumed to be all or none), and a bivariate continuous outcome.

The framework we adopt uses potential outcomes to define causal effects regardless of the mode of inference. We pursue a Bayesian approach and analyze the role of jointly modelling more than one outcome; the improved classification of units eases the identification of the mixture components, thereby reducing the variability of the posterior distribution of causal estimands.

It is also worth noting that, in a Bayesian setting, the effect of relaxing or maintaining assumptions can be directly addressed by examining how the posterior distributions for causal estimands change, therefore serving as a sort of sensitivity analysis.

In what follows we explain the structure of Bayesian inference in the presence of noncompliance and briefly describe methods for posterior inference based on data augmentation and MCMC algorithm (Section 2). Then we present the artificial data set we use and show some simulation results (Section 3); a brief discussion is provided in Section 4.

2 The structure of Bayesian inference for causal estimands in the presence of noncompliance

As a motivating example, suppose to design a social experiment to assess the efficacy of a language program for immigrants. The program is randomly offered to a sample of eligible secondary school students. Compliance is not perfect because only a subset of those students assigned to treatment (i.e., offered to participate in the program) attend the classes; access to the program is denied to those assigned to control. The primary outcome of interest is language ability measured with the score

achieved in a written test; in addition to the test result, also the time to complete the test is recorded and can be considered as a secondary outcome.

Let now introduce the potential outcome notation. Z_i is a binary treatment assignment for unit i ($Z_i = 0$ indicates the unit is assigned to the control group, whereas $Z_i = 1$ indicates the unit is assigned to the treatment group). $D_i(z)$ is the indicator of the treatment unit i would receive if assigned to treatment z ($z = 0, 1$), so that $D_i(Z_i)$ is the actual treatment received. Throughout this analysis we will make the stability assumption that there is neither interference between units nor different versions of the treatment (SUTVA; Rubin, 1978). $\mathbf{Y}_i(z)$ represents a bivariate (continuous) outcome for unit i if assigned to treatment z .

G_i is an indicator (stratum membership) which partitions units in the target population into 4 types based on their compliance behavior: c (complier if $D_i(z) = z$ for $z = 0, 1$), n (never-takers if $D_i(z) = 0$ for $z = 0, 1$), a (always-takers if $D_i(z) = 1$ for $z = 0, 1$), d (defiers if $D_i(z) = 1 - z$ for $z = 0, 1$). In our setting a strong monotonicity of compliance assumption holds by design ($D_i(0) = 0$ for all i), which rules out the presence of defiers and always takers. This implies that the population is only composed of compliers (c) and never-takers (n).

The observed data are: Z_i^{obs} , $D_i^{obs} = D_i(Z_i^{obs})$, $\mathbf{Y}_i^{obs} = \mathbf{Y}_i(Z_i^{obs})$ whereas the following quantities are missing: $D_i(1 - Z_i^{obs})$, $\mathbf{Y}_i(1 - Z_i^{obs})$. We observe the compliance behavior only partially, through the response to the actual assignment, D_i^{obs} . Because the type of a unit is a function of both compliance under assignment to the treatment and compliance under assignment to control, which we can never jointly observe, we generally do not know the compliance type.

Following Imbens and Rubin (1997b), we model the distribution of the stratum membership G_i and the conditional distribution of potential outcomes given the compliance type, rather than the joint distribution of the observed variables D_i^{obs} , $\mathbf{Y}_i^{obs}$ and Z_i^{obs} . Both distributions are parameterized so that, conditional on a general parameter θ , the model has an independent and identical distribution (i.i.d.) structure.

For simplicity and specificity, here we focus on commonly used outcome distributions and use conventional prior distributions, although our approach is widely applicable, also using more complex models. Specifically, we assume that math scores and time to complete the test follow a bivariate normal distribution conditional on the stratum membership and treatment assignment:

$$f(\mathbf{Y}_i(Z_i)|G_i = g, Z_i = z) = \phi[\mu_1^{g,z}, \mu_2^{g,z}, \Sigma^{g,z}],$$

where $g = c, n$ and $Z = 0, 1$. We use a Bernoulli distribution for the strata membership:

$$P[G_i = c] = \pi_c, \quad P[G_i = n] = (1 - \pi_c).$$

We have the following full parameter vector θ , containing a total of 21 parameters:

$$\theta = \{\mu^{c,0}, \mu^{c,1}, \mu^{n,0}, \mu^{n,1}, \Sigma^{c,0}, \Sigma^{c,1}, \Sigma^{n,0}, \Sigma^{n,1}; \pi_c\},$$

which characterizes the complete Likelihood function:

$$\begin{aligned}
L_{comp}(\theta|Z^{obs}, D^{obs}, \mathbf{Y}^{obs}, G) &= \\
&= \prod_{i:G_i=c} \pi_c \cdot \Phi[\boldsymbol{\mu}^{c,Z_i}, \boldsymbol{\Sigma}^{c,Z_i}] \cdot \prod_{i:G_i=n} (1 - \pi_c) \cdot \Phi[\boldsymbol{\mu}^{n,Z_i}, \boldsymbol{\Sigma}^{n,Z_i}].
\end{aligned} \tag{1}$$

We assume that parameters are a priori independent and use conjugate prior distributions:

$$\begin{aligned}
\pi_c &\sim \text{Beta}(\alpha_0, \beta_0), \\
\boldsymbol{\Sigma}^{g,z} &\sim \text{Inv-Wishart}_{\mathbf{V}_0^{g,z}}((\boldsymbol{\Lambda}_0^{g,z})^{-1}), \\
\boldsymbol{\mu}^{g,z} | \boldsymbol{\Sigma}^{g,z} &\sim N(\boldsymbol{\mu}_0^{g,z}, \boldsymbol{\Sigma}^{g,z}/k_0^{g,z}).
\end{aligned}$$

We construct a general state-space Markov Chain that has the joint distribution of the model parameters θ and the missing membership indicators G as its unique invariant equilibrium distribution. The Markov Chain algorithm that we adopt uses the data augmentation (DA) method of Tanner and Wong (1987) to impute at each step the missing indicators and to exploit the complete likelihood function to update the parameter distribution. At step $j+1$ we first draw G according to $P(G|\theta^{(j)}; Z^{obs}, D^{obs}, \mathbf{Y}^{obs})$. Given our prior assumptions these conditional distributions have a simple form. Conditional on θ , Z^{obs} , D^{obs} and $\mathbf{Y}^{obs}$, the G_i are independent of G_j , Z_j^{obs} , D_j^{obs} and $\mathbf{Y}_j^{obs}$ for all $j \neq i$. By the strong monotonicity assumption

$$P(G_i = n | Z_i^{obs} = 1, D_i^{obs} = 0, \mathbf{Y}_i^{obs}) = 1 \text{ and } P(G_i = c | Z_i^{obs} = 1, D_i^{obs} = 1, \mathbf{Y}_i^{obs}) = 1.$$

For observations with $Z_i = 0$ the following holds:

$$\begin{aligned}
P(G_i = c | Z_i^{obs} = 0, D_i^{obs} = 0, \mathbf{Y}_i^{obs}) &= \\
&= \frac{\pi_c^{(j)} \cdot \phi(\mathbf{Y}_i | \boldsymbol{\mu}^{c0(j)}, \boldsymbol{\Sigma}^{c0(j)})}{\pi_c^{(j)} \cdot \phi(\mathbf{Y}_i | \boldsymbol{\mu}^{c0(j)}, \boldsymbol{\Sigma}^{c0(j)}) + (1 - \pi_c^{(j)}) \cdot \phi(\mathbf{Y}_i | \boldsymbol{\mu}^{n0(j)}, \boldsymbol{\Sigma}^{n0(j)})},
\end{aligned}$$

$$P(G_i = n | Z_i^{obs} = 0, D_i^{obs} = 0, \mathbf{Y}_i^{obs}) = 1 - P(G_i = c | Z_i^{obs} = 0, D_i^{obs} = 0, \mathbf{Y}_i^{obs}).$$

Once the missing G_i are drawn, the prior distribution of the parameters is updated using the complete likelihood function in (1). Specifically, due to the assumptions of a priori independence of the parameters and conjugacy of priors, the complete data posterior distribution of θ is

$$\pi_c | G, Z^{obs}, D^{obs}, \mathbf{Y}^{obs} \sim \text{Beta}(\alpha_0 + \sum I(G_i = c); \beta_0 + \sum I(G_i = n)),$$

$$\boldsymbol{\Sigma}^{g,z} | G, Z^{obs}, D^{obs}, \mathbf{Y}^{obs} \sim \text{Inv-Wishart}_{\mathbf{V}_0^{g,z}}((\boldsymbol{\Lambda}^{g,z})^{-1}),$$

where

$$\begin{aligned}
\mathbf{V}_0^{g,z} &= \mathbf{V}_0^{g,z} + N^{g,z}, \text{ where } N^{g,z} = \sum I(G_i = n)I(Z_i = z), \\
\boldsymbol{\Lambda}^{g,z} &= \boldsymbol{\Lambda}_0^{g,z} + S^{g,z} + \frac{k_0^{g,z} N^{g,z}}{k_0^{g,z} + N^{g,z}} (\bar{\mathbf{Y}}^{g,z} - \boldsymbol{\mu}_0^{g,z})(\bar{\mathbf{Y}}^{g,z} - \boldsymbol{\mu}_0^{g,z})^T,
\end{aligned}$$

with $S^{g,z} = \sum_{i:G_i=c, Z_i=z} (\mathbf{Y}_i^{g,z} - \bar{\mathbf{Y}}^{g,z})(\mathbf{Y}_i^{g,z} - \bar{\mathbf{Y}}^{g,z})^T$, and

$$\boldsymbol{\mu}^{g,z} | \boldsymbol{\Sigma}^{g,z}; G, Z^{obs}, D^{obs}, \mathbf{Y}^{obs} \sim N(\boldsymbol{\mu}^{g,z}, \boldsymbol{\Sigma}^{g,z} / k^{g,z}),$$

where $k^{g,z} = k_0^{g,z} + N^{g,z}$ and $\boldsymbol{\mu}^{g,z} = \frac{k_0}{k_0 + N^{g,z}} \boldsymbol{\mu}_0^{g,z} + \frac{N^{g,z}}{k_0 + N^{g,z}} \bar{\mathbf{Y}}^{g,z}$. At the j th step of the MCMC algorithm the parameters vector $\boldsymbol{\theta}^{(j)}$ are easily drawn from these complete data posteriors and then used for the DA step.

3 An application to artificial data

Simulated data sets have been used to compare posterior inferences obtained by jointly modelling the two outcomes with those obtained working with each outcome separately¹. Different scenarios have been considered, to take account of alternative correlation structures between outcomes for compliers and never-takers and alternative deviations from the exclusion restriction.

We present results for six data sets each containing 600 observations, generated using the parameters' values reported in Table 1.

We specify relatively flat prior distributions by setting the hyperparameters equal to:

$$\alpha_0 = 1, \beta_0 = 1, v_0^{g,z} = 5, \boldsymbol{\Lambda}_0^{g,z} = \begin{bmatrix} 30 & 0 \\ 0 & 15 \end{bmatrix}, \boldsymbol{\mu}_0^{g,z} = \begin{bmatrix} 50 \\ 30 \end{bmatrix}, k_0^{g,z} = 1, g = c, n; z = 0, 1.$$

In order to assess the convergence of the Markov chain Monte Carlo algorithm, for each model we ran three chains from different starting values. Each chain was run for 210000 iterations. Results suggest good mixing of the chains and provide no evidence against convergence. Inference is based on the last 200000 iterations.

Results generally confirm our prior intuition that the simultaneous modelling of more than one outcome may improve estimation by reducing posterior uncertainty for causal estimands. Tables 2 and 3 report summaries of the posterior distributions of *ITT* effect for compliers and never-takers on the first outcome, in the univariate and bivariate case respectively.

The bivariate approach outperforms the univariate one in any scenario we consider, by providing considerably more precise estimates of the *ITT* effects for compliers and never-takers. The benefits of the bivariate approach especially arise when the mixture components of the observed likelihood are weakly overlapping, irrespective of the correlation structures between outcomes for compliers and never-takers (scenarios I and II). In such a case, jointly modelling the two outcomes reduces the width of the 95% posterior credible intervals more than 75%. In the overlapping scenarios, noticeable efficiency gains over the univariate models are observed when deviations from the exclusion restriction assumption are smaller (scenarios V and VI), and when compliers and never-takers are characterized by different correlation structures (scenario III).

¹ Consistently with our motivating example, the two outcomes are supposed to be language ability measured by the score achieved in a written test (from 0 to 100) and the time to complete the test (in minutes).

	$\mu^{c,0}$	$\mu^{c,1}$	$\mu^{n,0}$	$\mu^{n,1}$	$\Sigma^{c,0}$	$\Sigma^{c,1}$	$\Sigma^{n,0}$	$\Sigma^{n,1}$
I	$\begin{bmatrix} 50 \\ 30 \end{bmatrix}$	$\begin{bmatrix} 70 \\ 20 \end{bmatrix}$	$\begin{bmatrix} 40 \\ 50 \end{bmatrix}$	$\begin{bmatrix} 30 \\ 60 \end{bmatrix}$	$\begin{bmatrix} 25 & -12 \\ -12 & 9 \end{bmatrix}$	$\begin{bmatrix} 16 & -7 \\ -7 & 4 \end{bmatrix}$	$\begin{bmatrix} 36.00 & -6.75 \\ -6.75 & 20.25 \end{bmatrix}$	$\begin{bmatrix} 42.25 & -7.15 \\ -7.15 & 30.25 \end{bmatrix}$
II	$\begin{bmatrix} 50 \\ 30 \end{bmatrix}$	$\begin{bmatrix} 70 \\ 20 \end{bmatrix}$	$\begin{bmatrix} 40 \\ 50 \end{bmatrix}$	$\begin{bmatrix} 30 \\ 60 \end{bmatrix}$	$\begin{bmatrix} 25 & -12 \\ -12 & 9 \end{bmatrix}$	$\begin{bmatrix} 16 & -7 \\ -7 & 4 \end{bmatrix}$	$\begin{bmatrix} 36.00 & -21.60 \\ -21.60 & 20.25 \end{bmatrix}$	$\begin{bmatrix} 42.25 & -28.60 \\ -28.60 & 30.25 \end{bmatrix}$
III	$\begin{bmatrix} 50 \\ 30 \end{bmatrix}$	$\begin{bmatrix} 70 \\ 20 \end{bmatrix}$	$\begin{bmatrix} 48 \\ 50 \end{bmatrix}$	$\begin{bmatrix} 30 \\ 60 \end{bmatrix}$	$\begin{bmatrix} 25 & -12 \\ -12 & 9 \end{bmatrix}$	$\begin{bmatrix} 16 & -7 \\ -7 & 4 \end{bmatrix}$	$\begin{bmatrix} 36.00 & -6.75 \\ -6.75 & 20.25 \end{bmatrix}$	$\begin{bmatrix} 42.25 & -7.15 \\ -7.15 & 30.25 \end{bmatrix}$
IV	$\begin{bmatrix} 50 \\ 30 \end{bmatrix}$	$\begin{bmatrix} 70 \\ 20 \end{bmatrix}$	$\begin{bmatrix} 48 \\ 50 \end{bmatrix}$	$\begin{bmatrix} 30 \\ 60 \end{bmatrix}$	$\begin{bmatrix} 25 & -12 \\ -12 & 9 \end{bmatrix}$	$\begin{bmatrix} 16 & -7 \\ -7 & 4 \end{bmatrix}$	$\begin{bmatrix} 36.00 & -21.60 \\ -21.60 & 20.25 \end{bmatrix}$	$\begin{bmatrix} 42.25 & -28.60 \\ -28.60 & 30.25 \end{bmatrix}$
V	$\begin{bmatrix} 50 \\ 30 \end{bmatrix}$	$\begin{bmatrix} 70 \\ 20 \end{bmatrix}$	$\begin{bmatrix} 48 \\ 50 \end{bmatrix}$	$\begin{bmatrix} 45 \\ 60 \end{bmatrix}$	$\begin{bmatrix} 25 & -12 \\ -12 & 9 \end{bmatrix}$	$\begin{bmatrix} 16 & -7 \\ -7 & 4 \end{bmatrix}$	$\begin{bmatrix} 36.00 & -6.75 \\ -6.75 & 20.25 \end{bmatrix}$	$\begin{bmatrix} 42.25 & -7.15 \\ -7.15 & 30.25 \end{bmatrix}$
VI	$\begin{bmatrix} 50 \\ 30 \end{bmatrix}$	$\begin{bmatrix} 70 \\ 20 \end{bmatrix}$	$\begin{bmatrix} 48 \\ 50 \end{bmatrix}$	$\begin{bmatrix} 45 \\ 60 \end{bmatrix}$	$\begin{bmatrix} 25 & -12 \\ -12 & 9 \end{bmatrix}$	$\begin{bmatrix} 16 & -7 \\ -7 & 4 \end{bmatrix}$	$\begin{bmatrix} 36.00 & -21.60 \\ -21.60 & 20.25 \end{bmatrix}$	$\begin{bmatrix} 42.25 & -28.60 \\ -28.60 & 30.25 \end{bmatrix}$

Table 1 Parameters' values used to generate the six data sets.

	Mean	SD	2.5%	25%	50%	75%	97.5%
I <i>c</i>	23.452	3.230	19.259	20.440	21.839	26.718	28.239
I <i>n</i>	-15.241	4.752	-21.921	-20.095	-13.074	-10.805	-8.801
II <i>c</i>	22.912	3.379	18.472	19.615	21.960	26.203	27.760
II <i>n</i>	-15.580	5.015	-22.303	-20.497	-14.521	-10.698	-8.576
III <i>c</i>	19.987	1.767	16.830	18.509	20.114	21.426	22.986
III <i>n</i>	-19.114	2.618	-23.409	-21.284	-19.341	-16.925	-14.370
IV <i>c</i>	21.046	1.179	18.736	20.249	21.042	21.840	23.374
IV <i>n</i>	-18.398	1.754	-21.829	-19.580	-18.413	-17.220	-14.928
V <i>c</i>	20.879	1.533	18.048	19.705	20.905	22.068	23.627
V <i>n</i>	-4.124	2.293	-8.180	-5.916	-4.181	-2.359	0.125
VI <i>c</i>	20.269	1.396	17.661	19.221	20.292	21.323	22.818
VI <i>n</i>	-4.149	2.089	-7.913	-5.736	-4.197	-2.582	-0.220

Table 2 Summaries of the posterior distributions of *ITT* effect for *c* and *n* in the univariate case.

4 Discussion

Although the illustration provided is necessarily limited, the approach we propose appears to be widely applicable, also in observational studies, whenever the exclusion restriction assumption for the instrument may not hold. The additional information provided by the second outcome is obtained at the cost of having to specify more complex models for the multivariate outcome, which may increase the chances of misspecification. Therefore, directions for future research include the development of specific tools for assessing model fit, such as posterior predictive checks (Mattei and Mealli, 2007) and model comparisons by marginal likelihoods

		Mean	SD	2.5%	25%	50%	75%	97.5%
I	<i>c</i>	19.747	0.490	18.787	19.417	19.747	20.076	20.709
	<i>n</i>	-9.772	0.825	-11.388	-10.327	-9.772	-9.217	-8.155
II	<i>c</i>	19.167	0.502	18.182	18.828	19.166	19.505	20.150
	<i>n</i>	-10.002	0.847	-11.662	-10.572	-10.003	-9.433	-8.338
III	<i>c</i>	19.380	0.517	18.366	19.031	19.380	19.728	20.397
	<i>n</i>	-18.206	0.792	-19.762	-18.739	-18.206	-17.672	-16.652
IV	<i>c</i>	19.894	0.494	18.928	19.562	19.893	20.227	20.863
	<i>n</i>	-16.679	0.801	-18.259	-17.217	-16.679	-16.140	-15.112
V	<i>c</i>	20.195	0.487	19.238	19.866	20.195	20.523	21.150
	<i>n</i>	-3.100	0.798	-4.667	-3.638	-3.101	-2.563	-1.540
VI	<i>c</i>	19.908	0.471	18.983	19.592	19.908	20.226	20.832
	<i>n</i>	-3.610	0.779	-5.144	-4.134	-3.611	-3.086	-2.083

Table 3 Summaries of the posterior distributions of *ITT* effect for *c* and *n* in the bivariate case.

and Bayes factors (Chib and Jacobi, 2008), as well as the specification of Bayesian semiparametric models using, e.g., Dirichlet processes.

Analogously to covariates, assumptions on which may replace exclusion restriction assumptions to enable identification, our current effort is directed to showing how imposing specific dependence structures between outcomes may allow us to achieve nonparametric identification of causal effects.

References

1. Angrist, J.D., Imbens, G. W., Rubin, D.B.: Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association*, **91**, 444–472 (1996)
2. Chib, S., Jacobi, L.: Causal effects from panel data in randomized experiments with partial compliance. In: Chib, S., Griffiths, W.E., Koop, G. (eds.) *Advances in Econometrics*, **23**, pp 183–215. Jai Press, Bingley (2008)
3. Frangakis, C.E.: Comment on Causal effects in the presence of non compliance: a latent variable interpretation, by A. Forcina. *Metron*, LXIV, 8–14 (2006)
4. Frangakis, C.E., Rubin, D.B.: Principal stratification in causal inference. *Biometrics*, **58**, 191–199 (2002)
5. Hirano, K., Imbens, G.W., Rubin, D.B., Zhou, X-H.: Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, **1**, 1, 69–88 (2000)
6. Imbens, G. W., Rubin, D. B.: Estimating Outcome Distributions for Compliers in Instrumental Variables Models. *Review of Economic Studies*, **64**, 555–574 (1997a)
7. Imbens, G. W., Rubin, D. B.: Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*, **25**, 305–327 (1997b)
8. Jo, B.: Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics*. **27**, 385–420 (2002)
9. Mattei, A., Mealli, F.: Application of the Principal Stratification Approach to the Faenza Randomized Experiment on Breast Self-Examination. *Biometrics*, **63**, 437–446 (2007)
10. Rubin, D.B.: Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, **6**, 34–58 (1978)
11. Tanner, M., Wong, W.: The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, **82**, 528–550 (1987)