Bayesian inference for causal effects in randomized experiments with noncompliance: The role of multivariate outcomes

Fan Li, Alessandra Mattei and Fabrizia Mealli

Abstract Principal Stratification (PS) is a principled framework for addressing noncompliance issues. Due to the latent nature of principal strata, model-based PS analysis usually involves weakly identified models and identification of causal effects relies on untestable structural assumptions, such as exclusion restriction. This article develops a Bayesian approach to exploit multivariate outcomes to sharpen inferences for weakly identified models within PS. Simulation studies are performed to illustrate the potential gains in identifiability of jointly modelling more than one outcome. The method is applied to evaluate the causal effect of a job search program on depression.

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

1 Introduction

Many randomized experiments suffer from noncompliance, which *breaks* randomization, implying that assignment to the treatment rather than the treatment itself is randomly administered to individuals. In the presence of noncompliance, the treatment actually received is a post-treatment intermediate variable, which is potentially affected by the assignment and also may affect the response. A standard intention-to-treat analysis gives a valid inference of the effect of assignment on outcome,

Fan Li
Department of Statistical Science, Duke University, e-mail: fli@stat.duke.edu

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

but usually the goal is to study the effect of receiving the treatment rather than the assignment.

A principled framework to noncompliance is principal stratification (PS) (Frangakis and Rubin, 2002), a generalization of the instrumental variable approach to noncompliance by Angrist et al.(1996) and Imbens and Rubin (1997). While PS is applicable to a wide range of situations involving intermediate variables, such as truncation by death, mediation, this paper focuses on the special case of noncompliance. A PS with respect to the intermediate variable "receipt of the treatment" is a cross-classification of units into latent classes defined by the joint potential compliance statuses under both treatment and control. Principal causal effects (PCE), that is, comparisons of potential outcomes under different treatment levels within compliance principal strata, are in general the causal estimands of primary interest in a PS analysis.

Since at most one potential outcome is observed for any unit, compliance principal strata are generally latent and the key of PS analysis is to address the identifiability issue of PCEs. There are two streams of work in the existing literature regarding this: (1) deriving nonparametric bounds for the causal effects under minimal structural assumptions (e.g., Manski, 1990); (2) specifying additional structural or modelling assumptions, such as exclusion restriction, to identify PCEs and conducting sensitivity analysis to check the consequences of violations to such assumptions (e.g., Schwartz et al., 2012).

Using auxiliary information from covariates to identify causal effects has been also discussed (e.g., Jo, 2002). However, the importance of exploiting multiple outcomes is less acknowledged. In fact, information on multiple outcomes is routinely collected in randomized experiments and observational studies, but it is rarely used in analysis unless the goal is to study the relationships between outcomes. Exceptions include Jo and Muthen (2001), Mattei et al. (2012) and Mealli and Pacini (2012). In this article we further investigate the role of multivariate outcomes to sharpen inferences for weakly identified models within PS, proceeding from a parametric perspective, particularly under the Bayesian paradigm.

The article is organized as follows. Section 2 introduces the PS framework and Section 3 proposes a Bayesian approach to exploit multivariate outcomes to sharpen inferences for weakly identified models within PS. In Section 4, we perform simulation studies to examine the benefit from using multivariate outcomes under various scenarios. In Section 5, we re-analyze the Job Search Intervention Study (JOBS II) using the proposed bivariate approach. Section 6 concludes with a discussion.

2 The principal stratification approach to noncompliance

Discussion of causal inference in this article is carried out under the potential outcome framework, also known as the Rubin Causal Model (RCM) (Rubin, 1978). Consider a large population of units, each of which can potentially be assigned a treatment indicated by z, with $z = 1$ for treatment and $z = 0$ for control. A ran-

dom sample of n units from this population comprises the participants in a study, designed to evaluate the effect of the treatment on all or a subset of M outcomes $\mathbf{Y} = (Y_1, \ldots, Y_M)'$.

Assuming the standard Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980), for each outcome Y_m, we can define two potential outcomes for each unit, $Y_{im}(0)$ and $Y_{im}(1)$, corresponding to each of the two possible treatment level. For each unit i, let $\mathbf{Y}_i(z) = (Y_{i1}(z), \ldots, Y_{im}(z))'$ be the vector of the potential outcomes given assignment z.

In the presence of noncompliance, the actual taking of the treatment is beyond the control of the researcher, therefore there are also two potential treatment receipt indicators for each unit: $D_i(0)$ and $D_i(1)$. Let $S_i = (D_i(0), D_i(1))$ be the joint potential treatment outcomes. Applying the idea of principal stratification, units can be classified into four principal strata according to their compliance behaviour, defined by S_i: compliers ($S_i = (0,1) = c$); never takers ($S_i = (0,0) = n$); always takers ($S_i = (1,1) = a$); and defiers ($S_i = (1,0) = d$). By definition the principal stratum membership S_i is not affected by treatment assignment. Therefore, comparisons of summaries of $Y(1)$ and $Y(0)$ within a principal stratum, the so-called principal causal effects (PCEs), have a causal interpretation because they compare quantities defined on a common set of units. The causal estimands of interest in this article are the population principal average causal effects for the *first* outcome:

$$\tau_s = E(Y_{i1}(1) - Y_{i1}(0) | S_i = s), \tag{1}$$

for $s = c, a, n$, where τ_c is the well-known complier average causal effect (CACE).

Since $D_i(0)$ and $D_i(1)$ are never jointly observed, principal stratum S_i is latent. Specifically, for each unit i and for each post-treatment variable, only one potential outcome is observed. Let Z_i for $i = 1, ..., n$ be the binary variable indicating whether unit i is assigned to the treatment ($Z_i = 1$) or to the control ($Z_i = 0$). Then, the observed potential outcomes are: $D_i^{obs} = D_i(Z_i)$ and $\mathbf{Y}_i^{obs} = \mathbf{Y}_i(Z_i)$. Let $\mathbf{Z}, \mathbf{D}^{obs}, \mathbf{Y}^{obs}$ denote column vectors/matrices of the corresponding unit-level observed variables. The other potential outcomes $D_i^{mis} = D_i(1 - Z_i)$ and $\mathbf{Y}_i^{mis} = \mathbf{Y}_i(1 - Z_i)$, are missing. Henceforth, the bold denotes column vectors/matrices of the corresponding unit-level variables. Without loss of generality, we concentrate on the case of two outcomes ($M = 2$). Since we focus on randomized experiments, the following assumption holds by design:

Assumption 1. *Randomization of treatment assignment*

$$Pr(Z_i | D_i(0), D_i(1), \mathbf{Y}_i(0), \mathbf{Y}_i(1)) = Pr(Z_i).$$

3 Multivariate Bayesian principal stratification analysis

Following Imbens and Rubin (1997), we model the conditional distribution of the compliance type: $\pi_s = Pr(S_i = s)$, $s = a, c, d, n$; and the conditional distribution

of potential outcomes given compliance type: $f^i_{sz} = Pr(\mathbf{Y}^{obs}_i | S_i = s, Z_i = z; \boldsymbol{\theta}_{s,z})$, $z = 0,1$. Let $\boldsymbol{\theta} = (\pi_a, \pi_c, \pi_d, \pi_n, \{\boldsymbol{\theta}_{s,z}\}_{s=a,c,d,n;z=0,1})$ be the parameter vector and let $p(\boldsymbol{\theta})$ denote its prior distribution. Then, the posterior distribution of $\boldsymbol{\theta}$ can be written as

$$Pr(\boldsymbol{\theta}|\mathbf{Z}, \mathbf{D}^{obs}, \mathbf{Y}^{obs}) \propto p(\boldsymbol{\theta}) \prod_{i:Z_i=1, D^{obs}_i=1} \left[\pi_c f^i_{c1} + \pi_a f^i_{a1}\right] \prod_{i:Z_i=1, D^{obs}_i=0} \left[\pi_n f^i_{n1} + \pi_d f^i_{d1}\right]$$
$$\prod_{i:Z_i=0, D^{obs}_i=1} \left[\pi_a f^i_{a0} + \pi_d f^i_{d0}\right] \prod_{i:Z_i=0, D^{obs}_i=0} \left[\pi_n f^i_{n0} + \pi_c f^i_{c0}\right]$$

Without additional assumptions, inference on PCEs, τ_s, though possible and relatively straightforward from a Bayesian perspective, can be very imprecise, even in large samples, because models are only weakly identified. Jointly modelling multiple outcomes may help to reduce uncertainty about the treatment effects on the primary outcomes. Specifically, though additional outcomes do not play extra role in the compliance model, they can improve the prediction of principal strata membership through the outcome model. In addition, some key substantive identifying assumptions, such as *exclusion restriction* (ER), may be more plausible for secondary outcomes than the primary one. This condition is referred to as "partial exclusion restriction (PER)" in Mealli and Pacini (2012):

Assumption 2. *Stochastic Partial Exclusion Restriction*

$$\Pr(Y_{i2}(0)|S = s) = \Pr(Y_{i2}(1)|S = s) \quad for \quad s = a, n.$$

Restrictions on secondary outcomes reduce the parameter space of the joint distribution of all outcomes and in turn the marginal distribution of the primary one.

In our setting a strong monotonicity assumption holds by design, $D_i(1) \geq D_i(0)$ and $D_i(0) = 0$ for all i, implying that $\pi_d = 0$ and $\pi_a = 0$, so that the population is only composed of compliers and never-takers. Therefore a simple Bernoulli model is used for the compliance principal strata membership: $Pr(S_i = c) = \pi_c$, $\pi_c \in (0,1)$.

The outcome variables we focus on consists of either two continuous variables or a continuous variable and a binary indicator. For two continuous outcomes, conditional on the principal stratum, we assume a bivariate normal distribution: $\mathbf{Y}_i(z)|S_i = s, \sim N_2\left(\boldsymbol{\mu}^{s,z}, \boldsymbol{\Sigma}^{s,z}\right)$, where $\boldsymbol{\mu}^{s,z} = \begin{pmatrix} \mu^{s,z}_1 \\ \mu^{s,z}_2 \end{pmatrix}$ and $\boldsymbol{\Sigma}^{s,z} = \begin{pmatrix} \sigma^{s,z}_{11} & \sigma^{s,z}_{12} \\ \sigma^{s,z}_{12} & \sigma^{s,z}_{22} \end{pmatrix}$, $s = c,n$; $z = 0,1$. In the model for a continuous outcome Y_1 and a binary outcome Y_2, we replace Y_{i2} in the previous normal model by a latent variable Y^*_{i2} and assume in addition that $Y_{i2}(z) = I(Y^*_{i2}(z) > 0)$ with $\sigma^{s,z}_2 = 1$. This is equivalent to assuming a generalized linear model with probit link for Y_2: $Pr(Y_{i2}(z) = 1|S_i = s) = \Phi(\mu^{s,z}_2)$. The full set of parameters is $\boldsymbol{\theta} = \{\pi_c, \boldsymbol{\mu}^{s,z}, \boldsymbol{\Sigma}^{s,z}\}$. We assume that parameters are a priori independent. A conjugate prior Beta distribution is used for the compliance principal strata: $\pi_c \sim Beta(\alpha_0, \beta_0)$. Conjugate prior distributions are also used for the parameters of bivariate continuous outcome models: $\boldsymbol{\Sigma}^{s,z} \sim Inv - Wishart_{\nu_0}((\Lambda^{s,z}_0)^{-1})$; and $\boldsymbol{\mu}^{s,z}|\boldsymbol{\Sigma}^{s,z} \sim N_2(\boldsymbol{\mu}^{s,z}_0, \boldsymbol{\Sigma}^{s,z}/k^{s,z}_0)$. For continuous-binary outcomes, we use semi-conjugate diffused normal prior distri-

Table 1 True values of parameters of the six simulation scenarios.

	$\boldsymbol{\mu}^{c,0}$	$\boldsymbol{\mu}^{c,1}$	$\boldsymbol{\mu}^{n,0}$	$\boldsymbol{\mu}^{n,1}$	$\boldsymbol{\Sigma}^{c,0}$	$\boldsymbol{\Sigma}^{c,1}$	$\boldsymbol{\Sigma}^{n,0}$	$\boldsymbol{\Sigma}^{n,1}$
I	$\begin{bmatrix}2.5\\8\end{bmatrix}$	$\begin{bmatrix}0.5\\6.5\end{bmatrix}$	$\begin{bmatrix}2.75\\12\end{bmatrix}$	$\begin{bmatrix}4.25\\13\end{bmatrix}$	$\begin{bmatrix}0.09 & 0.24\\0.24 & 1\end{bmatrix}$	$\begin{bmatrix}0.01 & 0.08\\0.08 & 1\end{bmatrix}$	$\begin{bmatrix}0.16 & 0.16\\0.16 & 4\end{bmatrix}$	$\begin{bmatrix}0.04 & 0.08\\0.082 & 4\end{bmatrix}$
II	$\begin{bmatrix}2.5\\8\end{bmatrix}$	$\begin{bmatrix}0.5\\6.5\end{bmatrix}$	$\begin{bmatrix}2.75\\12\end{bmatrix}$	$\begin{bmatrix}4.25\\24\end{bmatrix}$	$\begin{bmatrix}0.09 & 0.24\\0.24 & 1\end{bmatrix}$	$\begin{bmatrix}0.01 & 0.08\\0.08 & 1\end{bmatrix}$	$\begin{bmatrix}0.16 & 0.16\\0.16 & 4\end{bmatrix}$	$\begin{bmatrix}0.04 & 0.12\\0.12 & 9\end{bmatrix}$
III	$\begin{bmatrix}2.5\\8\end{bmatrix}$	$\begin{bmatrix}0.5\\6.5\end{bmatrix}$	$\begin{bmatrix}2.75\\24\end{bmatrix}$	$\begin{bmatrix}4.25\\36\end{bmatrix}$	$\begin{bmatrix}0.09 & 0.24\\0.24 & 1\end{bmatrix}$	$\begin{bmatrix}0.01 & 0.08\\0.08 & 1\end{bmatrix}$	$\begin{bmatrix}0.16 & 0.96\\0.96 & 9\end{bmatrix}$	$\begin{bmatrix}0.04 & 0.8\\0.8 & 25\end{bmatrix}$

butions for the mean parameters, $\boldsymbol{\mu}^{s,z} \sim N_2(\boldsymbol{\mu}_0^{s,z}, \boldsymbol{\Sigma}_0^{s,z})$. For the covariance matrices $\boldsymbol{\Sigma}^{s,z}$, there is no conjugate prior, due to the constraint of $\sigma_{22} = 1$. As in Chib and Hamilton (2000), we assume a flexible truncated bivariate normal prior for the covariance parameters $\boldsymbol{\sigma}^{s,z} = (\sigma_{11}^{s,z}, \sigma_{12}^{s,z})$: $\boldsymbol{\sigma}^{s,z} \sim N_2(\boldsymbol{\sigma}_0^{s,z}, V_0^{s,z}) I_{\mathscr{A}}(\boldsymbol{\sigma}^{s,z})$ where $\boldsymbol{\sigma}_0^{s,z}$ and $V_0^{s,z}$ are hyperparameters, $\mathscr{A} = \{\boldsymbol{\sigma}^{s,z} :\in \Re^2 : \sigma_{11}^{s,z} > (\sigma_{12}^{s,z})^2\}$ is the region where $\boldsymbol{\Sigma}^{s,z}$ is a positive definite matrix, and $I_{\mathscr{A}}$ is the indicator function taking the value one if $\boldsymbol{\sigma}^{s,z}$ is in $\mathscr{A}$ and the value zero otherwise.

The joint posterior distribution, $\Pr(\boldsymbol{\theta}, \mathbf{D}^{mis} | \mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{Z})$, is obtained from a Markov Chain algorithm that uses the Data Augmentation method (Tanner and Wong 1987) to impute at each step the missing indicators and to exploit complete compliance data posterior distributions to update the parameter distribution. .

4 Simulations

To assess the improvement in the estimation of the PCEs by exploiting multivariate outcomes, we conduct simulation studies to compare the posterior inferences obtained by jointly modelling two outcomes with those by only one outcome. Consistently with the JOBS II data, the primary outcome is considered to be "depression score", measured with a 5-point rating scale (1 = not at all distressed to 5 = extremely distressed). To simplify the computation, we focus on two continuous outcomes, using alcohol use (in percent) as auxiliary outcome in the simulation study.

Here we present simulation results under three different scenarios, accounting for different correlation structures between the outcomes for compliers and never-takers and various deviations from the ER for the secondary outcome. The true simulation parameters are shown in Table 1 and all simulated data sets have $N = 600$ sample units, generated using principal strata probabilities of 0.7 for compliers and 0.3 for never-takers. The simulated samples are randomly divided in two groups, half assigned to the treatment and half to the control.

Figure 1 shows the histograms and 95% posterior intervals of the PCEs for compliers and never-takers on the primary outcome, in both the univariate and bivariate cases. The results clearly demonstrate that simultaneous modelling of both out-

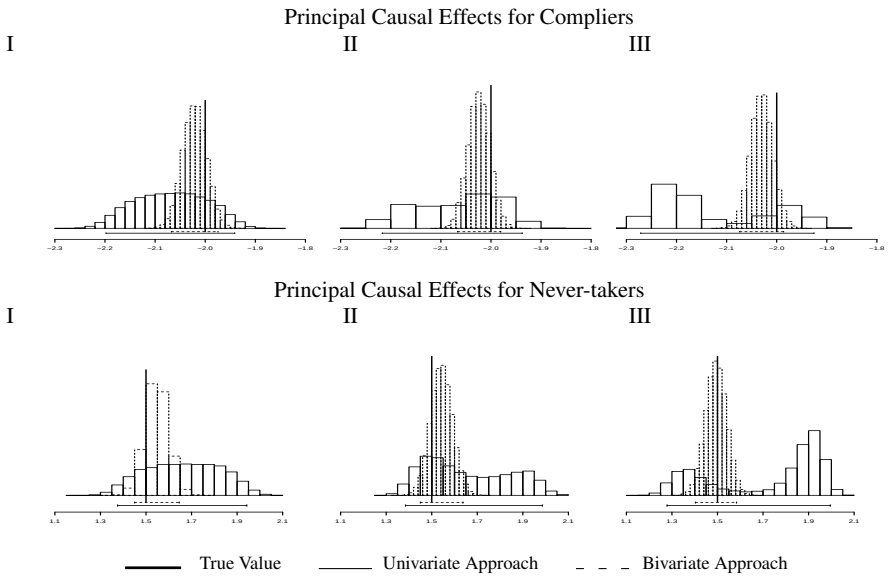

Fig. 1 Simulation Results: Histograms and 95% Posterior Intervals of PCEs for Compliers and Never-Takers.

comes significantly reduces posterior uncertainty for the causal estimates, providing considerably more precise estimates of the PCEs for compliers and never-takers.

In addition, the histograms in the upper and lower panels of Figure 1 suggest that the posterior distributions of the PCEs are much more informative in the bivariate case. Specifically, histogram (I)s show that the posterior distributions of the PCEs for compliers and never-takers are somewhat flat in the univariate approach, but become much tighter in the bivariate case. The improvement is even more dramatic in scenario (II) and (III), where the histograms show that posterior distributions of the PCEs for compliers and never-takers are bimodal in the univariate case, but both become unimodal in the bivariate case. Biases and MSEs (based on posterior mean) were also calculated (not shown here) and suggest that jointly modelling the two outcomes reduces the average biases by more than 64% and the MSEs by more than 79% in these scenarios. Several other scenarios with additional structural assumptions were also examined: magnitude of the improvement varies, and the pattern is consistent with what is described here.

5 Application to the JOBS II study

The Job Search Intervention Study (JOBS II) (Vinokur et al., 1995) is a randomized field experiment intended to prevent poor mental health and to promote high-quality reemployment among unemployed workers. The intervention consisted of 5 half-

Table 2 Posterior Distributions of PCEs on Depression for Compliers and Never-takers.

	Bivariate Approach				Univariate Approach		
	Without PER		With PER		Without ER		With ER
	τ_c	τ_n	τ_c	τ_n	τ_c	τ_n	τ_c
Mean	−0.135	−0.192	−0.211	−0.110	−0.207	−0.097	−0.269
SD	0.157	0.176	0.196	0.229	0.178	0.207	0.170
2.5%	−0.486	−0.526	−0.620	−0.587	−0.573	-0.532	−0.621
50%	−0.122	−0.197	−0.200	−0.100	−0.201	-0.086	−0.262
97.5%	0.143	0.179	0.144	0.306	0.123	0.281	0.045
Width $PCI_{0.95}$	0.629	0.706	0.764	0.893	0.696	0.812	0.666

Note that PER in the bivariate model is for remployment, whereas ER in the univariate model is for depression.

day job-search skills seminar. The control condition consisted of a mailed booklet briefly describing job-search methods and tips. Our analysis focuses on a sample of 398 subjects who were at high-risk of depression.

Since the treatment condition is only available to the individuals assigned to the intervention in JOBS II, there is no defiers and always-takers. Noncompliance arises in JOBS II because a substantial proportion (45%) of individuals invited to participate in the job-search seminar did not show up to the intervention. Our focus is on estimating causal effects of the intervention on a depression score measured six months after the intervention, relaxing ER, but using reemployment status as secondary outcome. ER on depression may be controversial, because, for example, never-takers randomized to the intervention might feel more demoralized by inability to take advantage of the opportunity. We use reemployment status as secondary outcome.

Table 2 reports summaries of the posterior distributions of PCEs for compliers and never-takers on depression in the bivariate (columns 1 through 4) and univariate (columns 5 through 7) case. Although the benefits of the bivariate approach are not pronounced, jointly modelling the two outcomes improves inference: the bivariate approach (without PER) provides more precise estimates of the PCEs for compliers and never-takers, and tighter 95% posterior credible intervals. Also it is worth noting that the bivariate approach leads to posterior distributions of τ_c and τ_n cantered at different means and medians. In the light of the simulation results, which show that jointly modelling two outcomes generally reduces the average biases, these findings make the bivariate estimates more faithful, suggesting that the univariate estimates may be affected by larger biases.

6 Discussion

We develop a Bayesian parametric bivariate model that exploits multiple outcomes of different types to improve the estimation of weakly identified causal estimands.

Although we focus on randomized experiments with noncompliance, our approach is immediately applicable to casual inference problems with alternative confounded post-treatment variables, and also in observational studies, whenever the exclusion restriction assumptions for the instrument are often questionable.

Our approach has several benefits. First, the Bayesian approach provides a refined map of identifiability, clarifying what can be learned when causal estimands are intrinsically not fully identified, but only weakly identified. Second, in a Bayesian setting, the effect of relaxing or maintaining assumptions (regardless of structural or modelling) can be directly checked by examining how the posterior distributions for causal estimands change, therefore serving as a natural framework for sensitivity analysis. Third, the use of multiple outcomes improves model identifiability, leading to smaller posterior variance of the parameters. However the additional information provided by secondary outcomes is obtained at the cost of having to specify more complex multivariate models, which may increase the possibility of misspecification. Therefore, model checking procedures to ensure sensible model specification is a valuable topic for future research.

References

1. Angrist, J.D., Imbens, G. W., Rubin, D.B.: Identification of causal effects using instrumental variables. Journal of the American Statistical Association, **91**, 444–455 (1996)
2. Chib, S., Hamilton, B.H.: Bayesian analysis of cross-section and clustered data treatment models. Journal of Econometrics, **97**, 25–50 (2000)
3. Frangakis, C.E., Rubin, D.B.: Principal stratification in causal inference. Biometrics, **58**, 191–199 (2002)
4. Imbens, G. W., Rubin, D. B.: Bayesian inference for causal effects in randomized experiments with noncompliance. Annals of Statistics, **25**, 305–327 (1997).
5. Jo, B.: Estimation of intervention effects with noncompliance: Alternative model specifications. Journal of Educational and Behavioral Statistics. **27**, 385–420 (2002).
6. Jo, B., Muthen, B.: Modeling of intervention effects with noncompliance: a latent variable approach for randomized trials. In G. Marcoulides, R. Schumacker (Eds.), New developments and techniques in structrual equation modeling, 57–87. Lawrence Erlbaum Associates, Publishers. Mahwah, New Jersey (2001).
7. Manski, C.: Nonparametric bounds on treatment effects. The American Economic Review, **80**, 319–323 (1990).
8. Mattei A., Mealli F., Pacini B.: Exploiting Multivariate Outcomes in Bayesian Inference for Causal Effects with Noncompliance. In Studies in Theoretical and Applied Statistics (SIS2010 Scientific Meeting). Forthcoming (2012).
9. Mealli, F., Pacini, B.: Using secondary outcomes and covariates to sharpen inference in randomized experiments with noncompliance. Technical report, Department of Statistics, University of Florence (2012).
10. Rubin, D.B: Comment on 'Randomization analysis of experimental data: The Fisher randomization test' by D. Basu. Journal of the American Statistical Association, **75**, 591–593 (1980).
11. Schwartz, S., Li, F., Reiter J.: Sensitivity analysis for unmeasured con- founding in principal stratification. Statistics in Medicine. In press (2012).
12. Tanner, M., Wong, W.: The calculation of posterior distributions by data augmentation (with discussion). Journal of the American Statistical Association, **82**, 528–550 (1987).
13. Vinokur, A., Price, R., Schul, Y.: Impact of the jobs intervention on unemployed workers varying in risk for depression. Journal of American Community Psychology, **23**, 39–74 (1995).