Biostatistics (2000), **1**, 1, *pp*. 69–88
Printed in Great Britain

Assessing the effect of an influenza vaccine in an encouragement design

KEISUKE HIRANO AND GUIDO W. IMBENS
Department of Economics, University of California, Los Angeles, CA 90095, USA

DONALD B. RUBIN
Department of Statistics, Science Center 709, Harvard University, Cambridge, MA 02138, USA

XIAO-HUA ZHOU
Division of Biostatistics, Department of Medicine, Indiana University School of Medicine and Regenstrief Institute for Health Care, Indianapolis, IN 46202, USA

SUMMARY

Many randomized experiments suffer from noncompliance. Some of these experiments, so-called encouragement designs, can be expected to have especially large amounts of noncompliance, because encouragement to take the treatment rather than the treatment itself is randomly assigned to individuals. We present an extended framework for the analysis of data from such experiments with a binary treatment, binary encouragement, and background covariates. There are two key features of this framework: we use an instrumental variables approach to link intention-to-treat effects to treatment effects and we adopt a Bayesian approach for inference and sensitivity analysis. This framework is illustrated in a medical example concerning the effects of inoculation for influenza. In this example, the analyses suggest that positive estimates of the intention-to-treat effect need not be due to the treatment itself, but rather to the encouragement to take the treatment: the intention-to-treat effect for the subpopulation who would be inoculated whether or not encouraged is estimated to be approximately as large as the intention-to-treat effect for the subpopulation whose inoculation status would agree with their (randomized) encouragement status whether or not encouraged. Thus, our methods suggest that global intention-to-treat estimates, although often regarded as conservative, can be too coarse and even misleading when taken as summarizing the evidence in the data for the effects of treatments.

Keywords: Bayesian analysis; Causal inference; Instrumental variables; Noncompliance; Rubin Causal Model; Potential outcomes; Treatment effects; Sensitivity analysis.

1. INTRODUCTION

Many empirical studies in medicine and the social sciences seek to establish causal relations between treatments and outcomes, rather than mere associations. The only generally accepted approach for inferring causality requires that the receipt of the various treatments is randomized. In many cases, however, it is not possible to randomize the receipt of treatment. For example, even if the assignment to treatment is random, some units may opt not to comply with their assignment. The standard intention-to-treat (ITT) analysis focuses on the causal effect of *assignment* of treatment rather than the causal effect of *receipt* of treatment. This approach is valid for measuring the effect of the encouragement, but interest often centers

on the effect of the treatment itself, and the interpretation of an ITT analysis is sometimes based on an implicit assumption that the effect of assignment is indicative of the effect of the treatment.

In this paper we investigate the effect of the receipt of a treatment in a situation where incentives for the treatment are randomly assigned, but incentives have only a limited correlation with the actual treatment received. For such cases, Imbens and Angrist (1994) and Angrist *et al*. (1996), showed that econometric instrumental variables (IV) methods can be interpreted as estimating a well-defined causal effect under the potential outcomes approach to causal inference advocated by Rubin (1974, 1978, 1990b), often referred to as the Rubin Causal Model (Holland, 1986). Imbens and Rubin (1997a) developed likelihood-based, including fully Bayesian, methods that improve upon conventional econometric IV estimators. As in Little and Yau (1997), we extend the analysis of Imbens and Rubin (1997a) to allow for the presence of pretreatment variables (covariates). We consider the consequences of econometric 'exclusion' restrictions that disallow, for various subpopulations, direct links between assignment and outcome other than through the effect of assignment on the treatment received. Some combinations of such restrictions are similar to the absence of arrows between assignment and outcomes in graphical causal models (Pearl, 1995), but our approach has the benefit of allowing for the comparison of results based on different combinations of these assumptions, thereby assessing sensitivity to violations of the exclusion restrictions. We emphasize the use of 'weakly identified' models: 'identified' in the sense of having a proper posterior distribution, but 'weak' in the sense of not having unique maximum likelihood estimates. We use these sensitivity analyses to investigate violations of various exclusion restrictions. Because these potential violations render the model only weakly identified, the choice of the form of the likelihood function and its associated prior distribution are more important than usual, and we discuss their specification in detail below.

We apply these methods to a reanalysis of a data set on influenza vaccinations previously studied by McDonald *et al*. (1992). In this study, physicians were randomly selected to receive a letter encouraging them to inoculate patients at risk for flu. The treatment of interest is the actual flu shot, and the outcome is an indicator for flu-related hospital visits. A standard ITT analysis suggests a moderate effect of assignment. That is, the receipt of a letter to the physician encouraging the physician to consider influenza inoculation for patients appears to reduce flu-related hospitalizations. Our data set involved randomization of an encouragement by the doctor, and doctors had multiple patients. Although we do not have information on the clustering of patients by doctor, we do have some covariate information on patients. To avoid an overly cumbersome analysis involving unknown cluster indicators, we assume exchangeability of patients conditional on covariates. To the extent that outcomes and compliance behavior are still correlated with missing cluster indicators after conditioning on the covariates, our analysis may lead to underestimation of standard errors and posterior standard deviations.

Although imperfect, our analysis suggests that there is little evidence that this ITT effect is actually due to the taking of the vaccine. In fact, under a very plausible model, we find that the subpopulation of the patients who would receive the vaccine regardless of whether their physician received a letter, appear to benefit as much from the letter (i.e. from assignment) as the subpopulation of patients who would only receive the vaccine if their physician received the encouragement letter. Because these subpopulations are not directly observable (i.e. latent), the analysis is not immediate, and our approach involves both instrumental variables techniques and Bayesian modeling.

We also find that a directly observable subclass of patients, those who have chronic obstructive pulmonary disease (COPD), are more likely to receive the influenza vaccine than patients who do not have COPD, regardless of whether their physicians received letters about the upcoming flu season. This result suggests that the link between assignment and treatment is related to health status, thereby invalidating two naive alternatives to an intention-to-treat analysis: both an 'as treated' analysis, which directly compares recipients of the vaccine with nonrecipients, and a 'per protocol' analysis, which directly compares recipients who were encouraged with nonrecipients who were not encouraged.

2. INTENTION-TO-TREAT ANALYSES

Because of epidemiologic evidence of increased morbidity related to influenza (Housworth and Langmuir, 1974), experimental evidence of serologic efficacy of the influenza vaccine (Francis and Magill, 1937), and observational studies suggesting improved outcomes in vaccinated patients (Patriarca *et al*., 1985), health officials in most countries recommend annual influenza vaccination for elderly persons and other people at high risk of influenza. However, no controlled randomized trials of the effects of the influenza vaccination on pulmonary morbidity in high-risk adults have been published (McDonald *et al*., 1992). One reason for this is that demonstrated efficacy for some subpopulations raises ethical barriers against performing randomized controlled trials on other subpopulations, which would require withholding vaccination from some subjects. One way around this impasse is to perform a randomized trial of an intervention that increases the use of influenza vaccine in one group of patients without affecting the use of influenza vaccine in another group. McDonald *et al*. (1992) exploited this idea to study influenza vaccine efficacy in reducing morbidity in high-risk adults, using a computer-generated reminder for flu shots. The study was conducted over a 3-year period (1978–1980) in an academic primary-care practice affiliated with a large urban public teaching hospital. Physicians in the practice were randomly assigned to either an intervention or a control group at the beginning of the study. Since physicians at the clinic each cared for a fixed group of patients, their patients were similarly classified. During the study period, physicians in the intervention group received a computer-generated reminder when a patient with a scheduled appointment was eligible for the influenza vaccine under U.S. Public Health Service Criteria.*

We reanalyse this study using the 2893 individuals observed in 1980, a particularly severe flu epidemic season. For each person i we observe: a binary variable Z_i^{obs}, the 'assignment' or 'encouragement', equal to one if patient i's physician received a reminder letter indicating that the patient was eligible to receive the influenza vaccine under U.S. Public Health Service Criteria and zero otherwise; a binary variable D_i^{obs}, the 'treatment', equal to 1 if person i received the vaccine and 0 otherwise; a binary outcome Y_i^{obs}, equal to 1 if person i subsequently experienced a flu-related hospitalization during the winter, which we define as being hospitalized for respiratory problems, and 0 otherwise; and two covariates, X_{i1}^{obs}, age in years, and X_{i2}^{obs}, an indicator for chronic obstructive pulmonary disease. The vectors $\mathbf{Z}^{\text{obs}}$, $\mathbf{D}^{\text{obs}}$, and $\mathbf{Y}^{\text{obs}}$ are N-dimensional vectors with ith elements equal to Z_i^{obs}, D_i^{obs} and Y_i^{obs}, respectively. The $N \times Z$ matrix $\mathbf{X}^{\text{obs}}$ has ith row equal to $(X_{i1}^{\text{obs}}, X_{i2}^{\text{obs}})$. As stated in the introduction, for simplicity, we assume that each patient has a distinct doctor, so that i indexes distinct doctor–patient pairs. Table 2 presents some summary statistics for the sample, classified by assignment, Z_i^{obs}, and treatment status, D_i^{obs}.

As can be seen in Table 2, the randomization of the assignment leads to the pretreatment variables being closely balanced in the two subsamples defined by assignment. The randomization does not, however, imply that the pretreatment variables are balanced in the subsamples defined by the actual treatment status. In fact, both age and chronic obstructive pulmonary disease (COPD) rates are significantly different between patients with flu shots and patients without flu shots. This imbalance indicates that we cannot simply compare outcomes by treatment status to obtain credible estimates of the effect of receipt of flu shots.

The conventional ITT approach to estimation of treatment effects compares outcomes by assignment, that is, by the receipt of the letter by the patient's physician, ignoring the actual receipt of treatment, that is, ignoring the receipt of the influenza vaccine. In our case the 'assignment' is merely an encouragement to take the treatment, so that nonencouraged patients may end up receiving the treatment, but this does not compromise the validity of standard methods for estimating ITT effects, which rest on the randomization of encouragement. The third row of Table 2 in the second block of columns provides a simple ITT analysis

*Patients over 65 years of age or with chronic lung disease, asthma, diabetes mellitus, congestive heart failure, or severe renal or hepatic failure were eligible.

Table 1. *Summary statistics, flu data (sample size 2893)*

		Means			Means		
	Grand mean	No letter $Z_i^{\text{obs}}=0$	Letter $Z_i^{\text{obs}}=1$	t-stat.	No flu shot $D_i^{\text{obs}}=0$	Flu shot $D_i^{\text{obs}}=1$	t-stat.
Letter (Z_i^{obs})	0.514	0	1	—	0.475	0.631	−7.5
Flu Shot (D_i^{obs})	0.250	0.190	0.307	−7.3	0	1	—
Hospitalization (Y_i^{obs})	0.085	0.092	0.078	1.4	0.085	0.084	0.1
Age (X_{i1}^{obs})	65.2	65.0	65.4	−0.8	64.7	66.8	−4.1
COPD (X_{i2}^{obs})	0.283	0.290	0.277	0.8	0.264	0.343	−4.0

of the data, which indicates a 15% ($= (0.092 - 0.078)/0.092 \times 100\%$) reduction in hospitalization rates due to encouragement to get flu shots.

More formal ITT analysis are summarized in Table 2. Since the relevant outcome is a binary hospitalization indicator, we estimate the logistic regression model:

$$\Pr\left[Y_i^{\text{obs}}=1|Z_i^{\text{obs}},X_{i1}^{\text{obs}},X_{i2}^{\text{obs}}\right]=\frac{\exp(\beta_0+\beta_1 Z_i^{\text{obs}}+\beta_2 X_{i1}^{\text{obs}}+\beta_3 X_{i2}^{\text{obs}})}{1+\exp(\beta_0+\beta_1 Z_i^{\text{obs}}+\beta_2 X_{i1}^{\text{obs}}+\beta_3 X_{i2}^{\text{obs}})}.$$

The first column of Table 2 shows posterior means (based on uniform prior distributions) in a model with no covariates (i.e. $\beta_2=\beta_3=0$), whereas the second column reports estimates for the full model.[†] The last row gives estimates of the ITT effect, defined as follows. Let $Y_i(1)$ denote the potential outcome for unit i if $Z_i=1$, and let $Y_i(0)$ denote the potential outcome if $Z_i=0$. We assume that $Y_i^{\text{obs}}=Y_i(Z_i^{\text{obs}})$. The ITT effect is defined as

$$ITT=\frac{1}{N}\sum_{i=1}^{N}(Y_i(1)-Y_i(0)).$$

For any individual, only one of the two potential outcomes is observed, but knowledge of β_0, β_1, and β_2 implies a distribution for the other potential outcome. We assume that the unobserved potential outcome is independent of the observed potential outcome conditional on the covariates and parameters.[‡] This in turn defines a distribution for the ITT effect conditional on the data Z^{obs}, Y^{obs}, and X^{obs}. We simulate the posterior distribution of ITT by taking draws for β_0, β_1, and β_2 from their posterior distribution, and imputing the missing potential outcomes for each set of parameter draws. This gives a set of draws from the posterior distribution of ITT. The last row of Table 2 reports the posterior means and standard deviations for ITT in the two models.

In both cases, the estimate of the average ITT effect is approximately equal to 1.3%, with a standard deviation of 0.7. The posterior probability that the ITT effect is negative, that is, that the receipt of the letter decreases average morbidity, is approximately 97%. Thus, there appears to be some evidence that

[†]We use the uniform prior for convenience, but the results appear to be insensitive to the choice of the prior.

[‡]Because $Y_i(0)$ and $Y_i(1)$ are never jointly observed, we cannot expect to learn anything about the partial correlation between the potential outcomes from the data given the covariates, although we can learn about the simple correlation (Rubin and Thayer, 1978). If we regard the N subjects in the study as a random sample from a much larger population, and the estimand is the corresponding average difference in this population, the independence assumption has little inferential effect for average treatment effects (Rubin, 1978; Imbens and Rubin, 1997a).

Table 2. *Intention-to-treat analysis using logistic models: summaries of posterior distributions (sample size 2893)*

	No covariates		Covariates	
	Mean	S.D.	Mean	S.D.
Intercept	−2.298	(0.094)	−1.998	(0.348)
Letter	−0.176	(0.133)	−0.179	(0.125)
Age	—	—	−0.007	(0.005)
COPD	—	—	0.373	(0.133)
ITT effect	−0.013	(0.008)	−0.014	(0.007)

the influenza vaccine reduces morbidity, although a standard two-sided t-test suggests this is not quite significant at the 5% level.

It is tempting to conclude from this analysis that the influenza vaccine is likely to have a direct effect in reducing morbidity. In their analysis, McDonald *et al*. (1992, p. 304), using the larger sample period, find a larger and statistically significant effect, and conclude, in a way that is typical of the ITT interpretation of randomized trials with noncompliance, that 'the most likely explanation for the difference [by assignment] is the greater use of influenza vaccine in the intervention group'.

Here we address the question of interpreting the results as estimating causal effects of the influenza vaccine on morbidity, by making explicit the assumptions underlying McDonald *et al.*'s (1992) claims. We then discuss some possible violations of the key exclusion restrictions necessary to identify causal effects of the influenza vaccine, provide a 'weakly identified' approach to estimation of more general models, and argue that in fact the evidence for the efficacy of the influenza vaccine from these data is extremely weak.

3. Modeling compliance behavior

We have already introduced a potential outcome notation to define a causal effect of the randomized encouragement. In this section we focus on defining the causal effect of interest, the effect of the influenza vaccine on flu-related hospitalizations. To do so we will use the extension to the standard potential outcomes model introduced by Angrist *et al*. (1996) and Imbens and Rubin (1997a), which we call the Causal Instrumental Variables Model. Throughout this analysis we will make the stability assumption (Rubin, 1978, 1980) that there is interference between neither units (Cox, 1958) nor different versions of the treatment.[§]

Let $D_i(z)$ be an indicator for the receipt of flu shot given assignment z; $D_i(0)$ is equal to 1 if patient i would receive a flu shot if i's physician did not receive a letter ($Z_i = 0$) and 0 if patient i would not, and

[§]In the context of an infectious disease with individuals as the units, the stability assumption is undesirably strong, although it could be more plausible if we could treat the family as the unit. Unfortunately, because so little is known about identification of causal effects without the stability assumption, the assumption is implicit in most approaches to causal inference. If stability does not hold, the precision of our estimates could be overstated, and in addition there could be bias of unknown direction.

$D_i(1)$ is equal to 1 if patient i would receive a flu shot if i's physician did receive a letter ($Z_i = 1$) and 0 if patient i would not; $\mathbf{D}$ is the $N \times 2$ matrix with ith row equal to $(D_i(0), D_i(1))$.

We partition the population of patients by 'compliance' behavior, where compliance is taken to mean that the treatment is the same as the encouragement. The combination of responses to the two assignments defines the compliance behavior of unit i, which we denote by C_i:

$$C_i = \begin{cases} c & \text{(i.e. unit } i \text{ is a complier)} & \text{if } D_i(z) = z, \text{ for } z = 0, 1, \\ n & \text{(i.e. unit } i \text{ is a never-taker)} & \text{if } D_i(z) = 0, \text{ for } z = 0, 1, \\ a & \text{(i.e. unit } i \text{ is an always-taker)} & \text{if } D_i(z) = 1, \text{ for } z = 0, 1, \\ d & \text{(i.e. unit } i \text{ is a defier)} & \text{if } D_i(z) = 1 - z, \text{ for } z = 0, 1. \end{cases}$$

We observe the compliance behavior only partially, through the response to the actual assignment, $D_i^{\text{obs}} = D_i(Z_i^{\text{obs}})$. We do not observe the response to the alternative assignment, $D_i^{\text{mis}} = D_i(1 - Z_i^{\text{obs}})$. 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 cannot know a unit's type, merely that the unit belongs to the subset of types consistent with its observed compliance behavior. Let $\mathcal{C}(t) = \{i | C_i = t\}$ for $t \in \{c, n, a, d\}$; $\mathbf{C}$ is the N component vector with ith element C_i, and N_t is the number of units of type t.

In addition, we define, for $z = 0, 1$, the potential outcomes $Y_i(z, D_i(z))$: $Y_i(z, D_i(z))$ is equal to 1 if, given assignment z and given receipt of treatment $D_i(z)$, unit i is hospitalized, and 0 otherwise; $\mathbf{Y}$ is the $N \times 2$ matrix with ith row equal to $(Y_i(0, D_i(0)), Y_i(1, D_i(1)))$. Using this notation, the ITT effect of assignment on the outcome can be defined as the weighted average

$$\text{ITT} = \sum_{t \in \{c,n,a,d\}} N_t \cdot \text{ITT}_t / N,$$

where, for $t \in \{c, n, a, d\}$,

$$\text{ITT}_t = \sum_{i \in \mathcal{C}(t)} [Y_i(1, D_i(1)) - Y_i(0, D_i(0))]/N_t,$$

is the average ITT effect of Z on Y for each of the four subpopulations defined by compliance behavior, and N_t/N is the weight assigned to ITT_t.

We observe for each unit i the actual assignment Z_i^{obs}, the actual treatment $D_i^{\text{obs}} = D_i(Z_i^{\text{obs}})$, the actual outcome $Y_i^{\text{obs}} = Y_i(Z_i^{\text{obs}}, D_i(Z_i^{\text{obs}}))$, and the pretreatment variables X_{i1}^{obs} and X_{i2}^{obs}.

Random assignment of the letter to doctors implies

$$\Pr\left(Z_i \middle| D_i(0), D_i(1), Y_i(0, 0), Y_i(0, 1), Y_i(1, 0), Y_i(1, 1), X_{i1}, X_{i2}\right) = \Pr(Z_i).$$

Although in our application the randomization was performed without taking into account the values of the pretreatment variables, one can allow conditioning on pretreatment variables with no change in our Bayesian analysis because assignment remains ignorable (Rubin, 1978). In general, we therefore only require:

ASSUMPTION 1. (IGNORABILITY OF TREATMENT ASSIGNMENT)

$$\Pr\left(Z_i \middle| D_i(0), D_i(1), Y_i(0, 0), Y_i(0, 1), Y_i(1, 0), Y_i(1, 1), X_{i1}^{\text{obs}}, X_{i2}^{\text{obs}}\right) = \Pr(Z_i | X_{i1}^{\text{obs}}, X_{i2}^{\text{obs}}).$$

We make one additional assumption at this point:

ASSUMPTION 2. (MONOTONICITY OF COMPLIANCE)
For all i,

$$D_i(1) \geq D_i(0).$$

This assumption rules out the existence of defiers, patients who would receive the vaccine if their physician did not receive the letter, but would not receive the vaccine if their physician did receive the letter. Underlying this assumption is the notion that although physicians need not give every patient a flu shot upon receipt of the letter, they are unlikely to decide after receiving the letter not to give flu shots to patients to whom they would have given a flu shot in the absence of the letter—encouragement makes it more likely for everybody that the treatment was in fact received. This assumption appears very plausible in our application, and in many other applications of encouragement designs, and we therefore make it throughout this discussion.

4. EXCLUSION RESTRICTIONS

In this section we consider, but do not necessarily impose, two additional assumptions which rule out direct effects of the letter on hospitalizations. The concepts underlying these assumptions have a long tradition in the econometric instrumental variables literature (Reiersol, 1941; Haavelmo, 1943) and are widely used in economics (e.g. Heckman and Robb, 1985; Angrist, 1990; Manski, 1990; Angrist and Krueger, 1991). Similar ideas have been considered in other fields by, among others, Zelen (1979, 1990), Hearst *et al.* (1986), Holland (1988), Permutt and Hebel (1989), Robins (1989), Efron and Feldman (1991), Sommer and Zeger (1991), Baker and Lindeman (1994), McClellan and Newhouse (1994), Pearl (1995), and Little and Yau (1998). For a discussion of the specific form of the assumptions we employ here and further references see Angrist *et al.* (1996). These assumptions formalize McDonald *et al.*'s (1992) argument that the most likely explanation for the ITT effects is the effect of the influenza vaccine.

In contrast to the previous literature, we distinguish two components of this assumption, one for never-takers and one for always-takers. In the first component we assume that within subpopulations of never-takers with the same values of the covariates, the distributions of the two potential outcomes $Y_i(0, D_i(0))$ and $Y_i(1, D_i(1))$ are the same:

ASSUMPTION 3. (STOCHASTIC EXCLUSION RESTRICTION FOR NEVER-TAKERS)

$$\Pr(Y_i(1, D_i(1)) = 1 | X_{i1}, X_{i2}, C_i = n) = \Pr(Y_i(0, D_i(0)) = 1 | X_{i1}, X_{i2}, C_i = n).$$

In the second component of the exclusion restriction we assume that within subpopulations of always-takers with the same values of the covariates, the distributions of the two potential outcomes $Y_i(0, D_i(0))$ and $Y_i(1, D_i(1))$ are the same:

ASSUMPTION 4. (STOCHASTIC EXCLUSION RESTRICTION FOR ALWAYS-TAKERS)

$$\Pr(Y_i(1, D_i(1)) = 1 | X_{i1}, X_{i2}, C_i = a) = \Pr(Y_i(0, D_i(0)) = 1 | X_{i1}, X_{i2}, C_i = a).$$

These two assumptions rule out, for the two types of units for whom there is no effect of assignment on receipt of treatment, any systematic effect of assignment on the outcome, by asserting that the two distributions of potential outcomes indexed by assignment do not vary with assignment within subpopulations indexed by covariates and compliance type. It formalizes the notion that any ITT effect of assignment on

the outcome should be mediated by an effect of assignment on the treatment received. We regard these two assumptions as possibly controversial, and we will investigate their consequences in some detail. It should be noted, however, that even under the exclusion restrictions for both never-takers and always-takers, the attribution of the complier population ITT effect, ITT_c, to the change in *treatment* for compliers is an assumption. The desire to make this attribution more plausible underlies the widespread practice of blinding and double blinding in medical evaluations of treatments, typically impossible in encouragement designs. In the remainder of the paper, we do not require this assumption. We focus on inference for the complier ITT effect and leave its interpretation open.

The two exclusion restrictions suffice to identify the ITT effect for compliers without any further parametric assumptions (Imbens and Angrist, 1994; Angrist *et al.*, 1996). Some testable restrictions are implied by the two restrictions (Balke and Pearl, 1993; Imbens and Rubin, 1997b), but in order to relax fully one or both exclusion restrictions, it is useful to make auxiliary assumptions. In the next section, we do this by imposing a parametric form on the likelihood function and using a relatively diffuse but proper prior distribution. Because of the reliance of the weakly identified analyses on these auxiliary assumptions, their interpretation will require care; nevertheless, we will argue below that they can play an important role in assessing sensitivity of the inference to the exclusion restrictions. In particular, because it is possible that doctors took actions other than administering the treatment of interest in response to the letter, we believe that the weakly identified models can yield more relevant answers than models with assumptions chosen primarily for purposes of identification rather than for scientific reasons.

5. PARAMETRIC MODELS

Following Imbens and Rubin (1997a), we model the conditional distribution of the compliance type C_i given pretreatment variables, and the conditional distribution of potential outcomes given pretreatment variables and compliance type, rather than the joint distribution of the observed variables D_i^{obs}, Y_i^{obs} and Z_i^{obs} given the pretreatment variables. Both distributions are parametrized so that conditional on a general parameter, denoted by π, the model has an independent and identical distribution (i.i.d.) structure. Incorporating the compliance type into the parametric model has two key advantages. First, it simplifies the process of imposing the substantive restrictions (the monotonicity condition and the exclusion restrictions) discussed in the previous sections. Second, it allows us to examine directly average treatment effects for subpopulations, such as the subpopulation of compliers. The unknown C_i values will be treated as missing data in the analyses.

In the general model, which does not impose the monotonicity assumption or either of the two exclusion restrictions, there are eight outcome distributions: one given receipt of letter and one given no receipt of letter, for each of the four types of units, never-takers, always-takers, compliers and defiers. The monotonicity assumption eliminates the two outcome distributions for defiers. Because our outcome is dichotomous, we assume that the remaining six outcome distributions take the form of logistic regressions:

$$\Pr(Y_i(Z_i, D_i(Z_i)) = 1 | C_i = t, Z_i = z, X_{i1} = x_1, X_{i2} = x_2, \pi) = \Lambda(x_1, x_2, \beta_{tz}),$$

where $\beta_{tz} = (\beta_{tz0}, \beta_{tz1}, \beta_{tz2})'$, and

$$\Lambda(x_1, x_2, \beta_{tz}) = \frac{\exp(\beta_{tz0} + \beta_{tz1} \cdot x_1 + \beta_{tz2} \cdot x_2)}{1 + \exp(\beta_{tz0} + \beta_{tz1} \cdot x_1 + \beta_{tz2} \cdot x_2)},$$

for all $t \in \{c, n, a\}$ and $z = 0, 1$. We assume that conditional on X_i and π, the two outcomes $Y_i(0, D_i(0))$ and $Y_i(1, D_i(1))$ are independent. This assumption can easily be relaxed, but since the data are not informative about this partial association structure, there is typically little gain in doing so. (See footnote ‡.)

For the distribution of types we use a multinomial logit model:

$$\Pr(C_i = c|X_{i1} = x_1, X_{i2} = x_2, \pi) = \Psi(c, x_1, x_2, \psi_c, \psi_n, \psi_a),$$

$$\Pr(C_i = n|X_{i1} = x_1, X_{i2} = x_2, \pi) = \Psi(n, x_1, x_2, \psi_c, \psi_n, \psi_a),$$

and

$$\Pr(C_i = a|X_{i1} = x_1, X_{i2} = x_2, \pi) = \Psi(a, x_1, x_2, \psi_c, \psi_n, \psi_a),$$

where, for $t \in \{c, n, a\}$, we have:

$$\Psi(t, x_1, x_2, \psi_c, \psi_n, \psi_a) = \frac{\exp(\psi_{t0} + \psi_{t1}x_1 + \psi_{t2}x_2)}{\sum_{v\in\{c,n,a\}} \exp(\psi_{v0} + \psi_{v1}x_1 + \psi_{v2}x_2)}.$$

We normalize these probabilities by setting ψ_n equal to the three-dimensional vector of zeros. The full parameter vector is $\pi = (\beta_c, \beta_n, \beta_a, \psi_c, \psi_a)$, where $\beta_c = (\beta_{c0}, \beta_{c1})$, $\beta_n = (\beta_{n0}, \beta_{n1})$, and $\beta_a = (\beta_{a0}, \beta_{a1})$, for a total of 26 parameters.

Consider the complete-data likelihood function, based on observing $\mathbf{Z}^{\text{obs}}$, $\mathbf{D}^{\text{obs}}$, $\mathbf{Y}^{\text{obs}}$, $\mathbf{X}^{\text{obs}}$, as well as the vector of compliance type indicators $\mathbf{C}$:

$$\mathcal{L}_{\text{comp}}(\pi|\mathbf{Z}^{\text{obs}}, \mathbf{D}^{\text{obs}}, \mathbf{Y}^{\text{obs}}, \mathbf{X}^{\text{obs}}, \mathbf{C}) =$$

$$\prod_{i\in\mathcal{C}(c)} \Psi(c, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a) \cdot \Lambda(X_{i1}, X_{i2}, \beta_{cZ_i})^{Y_i} \Big(1 - \Lambda(X_{i1}, X_{i2}, \beta_{cZ_i})\Big)^{1-Y_i}$$

$$\prod_{i\in\mathcal{C}(n)} \Psi(n, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a) \cdot \Lambda(X_{i1}, X_{i2}, \beta_{nZ_i})^{Y_i} \Big(1 - \Lambda(X_{i1}, X_{i2}, \beta_{nZ_i})\Big)^{1-Y_i}$$

$$\prod_{i\in\mathcal{C}(a)} \Psi(a, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a) \cdot \Lambda(X_{i1}, X_{i2}, \beta_{aZ_i})^{Y_i} \Big(1 - \Lambda(X_{i1}, X_{i2}, \beta_{aZ_i})\Big)^{1-Y_i}.$$

The complete-data likelihood function has a simple form with nine factors, one for each of the six outcome distributions and three involving the parameters of the type distribution. For inference based on the observed data, we cannot work directly with this complete-data likelihood function, because we do not observe the type C_i of each unit. However, we can exploit the complete-data likelihood function by using missing data methods such as the EM algorithm (Dempster *et al.*, 1977), and the Data Augmentation (DA) algorithm (Tanner and Wong, 1987). In the Appendix, we describe the numerical methods used to generate the inferences reported below.

There are four possible patterns of missing and observed data in $(\underline{D_i}, \underline{Y_i})$ corresponding to the four possible values for $(Z_{\text{obs},i}, D_{\text{obs},i})$: (0,0), (0,1), (1,0), (1,1). Indicate the subsets of units exhibiting each pattern by $\mathcal{S}(0, 0)$, $\mathcal{S}(0, 1)$, $\mathcal{S}(1, 0)$, and $\mathcal{S}(1, 1)$. We can then write the actual (i.e. observed) likelihood function in terms of the observed data as

$$\mathcal{L}_{\text{obs}}(\pi|\mathbf{Z}^{\text{obs}}, \mathbf{D}^{\text{obs}}, \mathbf{Y}^{\text{obs}}, \mathbf{X}^{\text{obs}})$$

$$= \prod_{i\in\mathcal{S}(0,0)} \Big[\Psi(c, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a) \cdot \Lambda(X_{i1}, X_{i2}, \beta_{c0})^{Y_i} \Big(1 - \Lambda(X_{i1}, X_{i2}, \beta_{c0})\Big)^{1-Y_i}$$

$$+ \Psi(n, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a) \cdot \Lambda(X_{i1}, X_{i2}, \beta_{n0})^{Y_i} \Big(1 - \Lambda(X_{i1}, X_{i2}, \beta_{n0})\Big)^{1-Y_i} \Big]$$

$$\times \prod_{i\in\mathcal{S}(1,0)} \Psi(n, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a) \cdot \Lambda(X_{i1}, X_{i2}, \beta_{n1})^{Y_i} \Big(1 - \Lambda(X_{i1}, X_{i2}, \beta_{n1})\Big)^{1-Y_i}$$

$$\times \prod_{i\in\mathcal{S}(1,0)} \Psi(a, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a) \cdot \Lambda(X_{i1}, X_{i2}, \beta_{a0})^{Y_i} \Big(1 - \Lambda(X_{i1}, X_{i2}, \beta_{a0})\Big)^{1-Y_i}$$

$$\times \prod_{i\in\mathcal{S}(1,1)} \Big[\Psi(c, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a) \cdot \Lambda(X_{i1}, X_{i2}, \beta_{c1})^{Y_i} \Big(1 - \Lambda(X_{i1}, X_{i2}, \beta_{c1})\Big)^{1-Y_i}$$

$$+\Psi(a, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a) \cdot \Lambda(X_{i1}, X_{i2}, \beta_{a1})^{Y_i} \Big(1 - \Lambda(X_{i1}, X_{i2}, \beta_{a1})\Big)^{1-Y_i}\Big].$$

The posterior distribution can be sensitive to the choice of prior distribution, because the observed-data likelihood has a mixture structure over a large amount of missing data. For example, standard diffuse, improper prior distributions can lead to improper posterior distributions. We therefore use a proper prior distribution with a simple conjugate form. Our prior distribution corresponds to adding to the likelihood function 30 extra observations: there are 10 additional observations for each type (complier, never-taker, and always-taker); for each type the 10 additional observations are split into 2.5 for each of the four combinations of the binary variables (Z_i, Y_i), further split equally into $2.5/N$ artificial observations for each of the N observed pairs of values of the pretreatment variables, X_{i1} and X_{i2}. More formally, the prior distribution is proportional to

$$p(\pi) \propto \prod_{i=1}^{N} \times \prod_{t\in\{c,n,a\}} \times \prod_{z=0,1} \prod_{y=0,1}$$

$$\Big[\Psi(t, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a) \cdot \Lambda(X_{i1}, X_{i2}, \beta_{tz})^{y} (1 - \Lambda(X_{i1}, X_{i2}, \beta_{tz}))^{(1-y)}\Big]^{2.5/N}.$$

In the application in this paper, we impose prior equality of the slope coefficients in the outcome regressions: $\beta_{c01} = \beta_{c11} = \beta_{n01} = \beta_{n11} = \beta_{a01} = \beta_{a11} \equiv \beta_{0.1}$ and $\beta_{c02} = \beta_{c12} = \beta_{n02} = \beta_{n12} = \beta_{a02} = \beta_{a12} \equiv \beta_{0.2}$, reducing the number of parameters to 14. Relaxing these restrictions would not complicate the computational methodology greatly, but given the relatively small sample size, would lead to imprecise estimates.

To demonstrate that this proper prior distribution does not lead to a highly informative prior distribution for the estimands of interest, Table 3 presents summary statistics, obtained by the methods described in the Appendix, of the marginal prior distributions of the ITT effects for the three subpopulations and of the overall ITT effect, given each of the four combinations of exclusion restrictions. The joint distributions of the ITT effects were obtained using the same computational techniques used to obtain the actual posterior distribution with the data. The comparison of the standard deviations in Table 3 for the ITT effects with the corresponding values in Table 4 below, indicates that the prior distribution is relatively uninformative about quantities of interest.

6. CAUSAL INFERENCE UNDER EXCLUSION RESTRICTIONS

In Table 4, summary statistics of the posterior distributions of the estimands of interest are presented under the four combinations of the exclusion restrictions. Figures 1–4 show simulation scatterplots for two-way joint distributions of subpopulation ITT effects. Figure 1 shows the joint distribution of ITT_a and ITT_c in the model that relaxes the exclusion restriction for always-takers only; Figure 2 shows the joint distribution of ITT_n and ITT_c in the model that relaxes the exclusion restriction for never-takers only; and

Table 3. *Summary statistics: prior distributions*

Excl. Res. Never-takers ⟶	Yes		Yes		No		No	
Excl. Res. Always-takers ⟶	Yes		No		Yes		No	
Estimand	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
ITT_c	0.005	(0.278)	−0.002	(0.282)	0.003	(0.285)	−0.013	(0.280)
ITT_n	0	0	0	0	0.009	(0.282)	0.001	(0.287)
ITT_a	0	0	−0.005	(0.279)	0	0	−0.001	(0.283)
ITT	0.002	(0.095)	−0.002	(0.135)	0.004	(0.139)	−0.004	(0.169)

Table 4. *Summary statistics: posterior distributions*

Excl. Res. Never-takers ⟶	Yes		Yes		No		No	
Excl. Res. Always-takers ⟶	Yes		No		Yes		No	
Estimand	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
ITT_c	−0.082	(0.068)	−0.037	(0.078)	−0.196	(0.147)	−0.168	(0.161)
ITT_n	0	0	0	0	0.022	(0.026)	0.025	(0.027)
ITT_a	0	0	−0.053	(0.032)	0	0	−0.058	(0.033)
ITT	−0.010	(0.008)	−0.014	(0.008)	−0.009	(0.007)	−0.013	(0.008)
$E[Y_i(0, D_i(0))\|C_i = c]$	0.121	(0.063)	0.124	(0.063)	0.236	(0.145)	0.263	(0.160)
$E[Y_i(1, D_i(1))\|C_i = c]$	0.039	(0.026)	0.087	(0.047)	0.040	(0.026)	0.095	(0.049)
$E[Y_i(0, D_i(0))\|C_i = n]$	0.082	(0.005)	0.082	(0.005)	0.062	(0.025)	0.058	(0.026)
$E[Y_i(1, D_i(1))\|C_i = n]$	0.082	(0.005)	0.082	(0.005)	0.083	(0.006)	0.083	(0.006)
$E[Y_i(0, D_i(0))\|C_i = a]$	0.100	(0.008)	0.114	(0.014)	0.100	(0.008)	0.114	(0.014)
$E[Y_i(1, D_i(1))\|C_i = a]$	0.100	(0.008)	0.061	(0.029)	0.100	(0.008)	0.056	(0.029)
$\Pr(C_i = c)$	0.119	(0.014)	0.117	(0.014)	0.121	(0.014)	0.117	(0.014)
$\Pr(C_i = n)$	0.692	(0.008)	0.693	(0.008)	0.692	(0.008)	0.693	(0.008)
$\Pr(C_i = a)$	0.189	(0.007)	0.190	(0.007)	0.188	(0.007)	0.190	(0.007)

Figures 3 and 4 show joint distributions for the model with no exclusion restrictions. In addition, Table 5 summarizes posterior distributions of the parameter vector π in the different models.

First, consider the last block of columns in Table 4, presenting results for the case with no exclusion restrictions. The standard ITT estimand is still precisely estimated, with essentially the same posterior mean and standard deviation as in the original logistic ITT analysis. The subpopulation ITT effects, however, are estimated very imprecisely.

In the first block of columns we impose both exclusion restrictions. Now we estimate the complier ITT effect, the only subpopulation ITT effect, which is not assumed to equal zero, fairly precisely. These

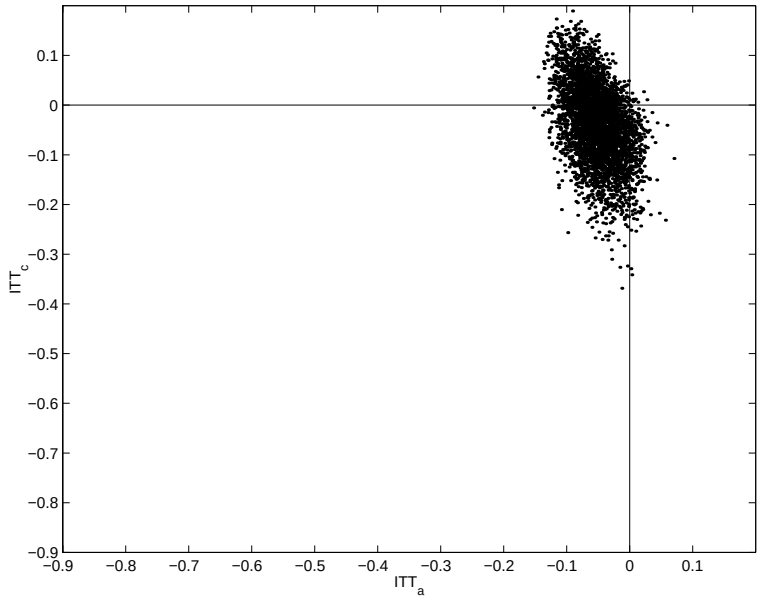

Fig. 1. Simulation scatterplot of the joint posterior distribution of ITT_a and ITT_c, in the model with exclusion restriction only for never-takers.

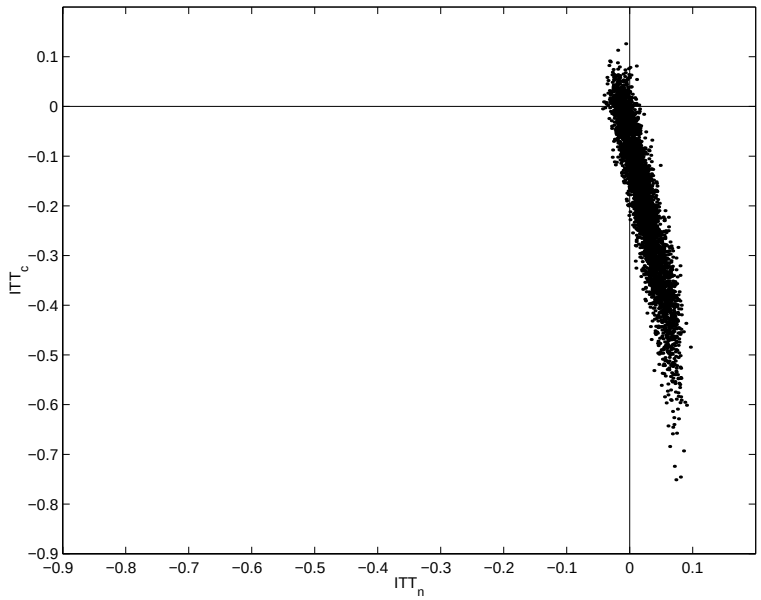

Fig. 2. Simulation scatterplot of the joint posterior distribution of ITT_n and ITT_c, in the model with exclusion restriction only for always-takers.

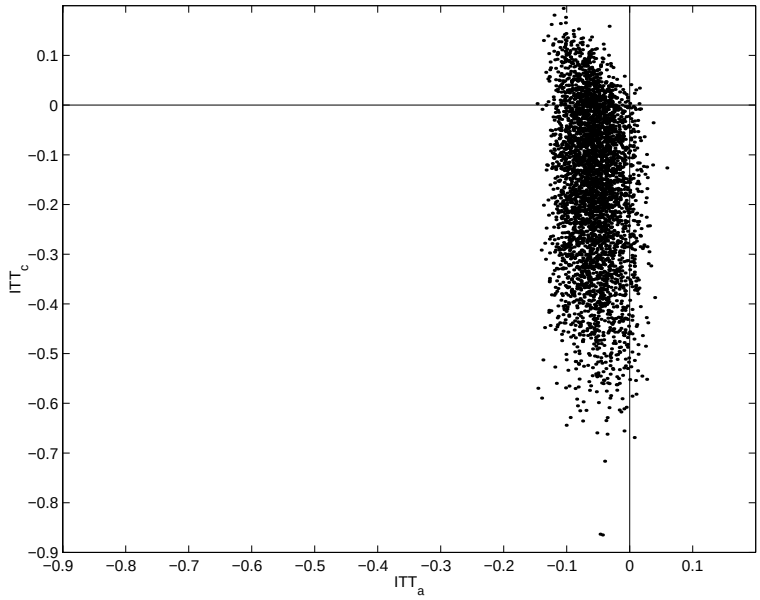

Fig. 3. Simulation scatterplot of the joint posterior distribution of ITT_a and ITT_c, in the model with no exclusion restrictions.

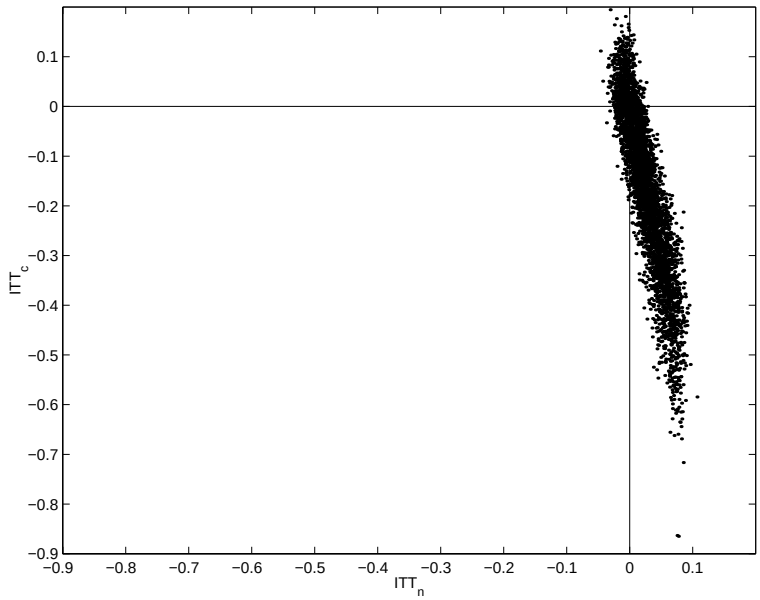

Fig. 4. Simulation scatterplot of the joint posterior distribution of ITT_n and ITT_c, in the model with no exclusion restrictions.

restrictions correspond to the standard IV analysis, as well as to the notion articulated by McDonald *et al*. (1992) that the ITT effect can be largely attributed to the effect of the vaccine on hospitalization, rather than the effect of the letter on hospitalization. The estimated ITT effect for compliers is a reduction of flu-related hospitalizations of 8.2%, from 12.1% without a flu shot to 3.9% with the flu shot. Note that this estimated effect is much larger than the ITT effect, 1.0%, because only 12.1% of the population is estimated to be compliers when both exclusion restrictions are in force ($1.0\% / 12.1\% \approx 8.2\%$).

The middle two blocks of columns in Table 4 represent a key benefit of our framework. Rather than having to impose the exclusion restriction for all types of noncompliers, as with conventional econometric IV methods, we can impose it for any combination of the subpopulations of always-takers and never-takers. In this application, as in many others (e.g. the military service example discussed in Hearst *et al*. 1986; Angrist, 1990; and Angrist *et al*. 1996), the two exclusion restrictions have very different interpretations, and their plausibility rests on very different arguments.

Consider first the exclusion restriction for always-takers. The always-takers are patients who would receive the influenza vaccine irrespective of the receipt of the encouragement letter by their physician. Such patients are predominantly at higher risk for the flu; in our analysis this is revealed by the positive multinomial logit coefficients on age and COPD (see Table 5), which imply that always-takers tend to be older and more likely to have COPD. How could the exclusion restriction be violated for such patients? That is, why would such patients be affected by a letter warning their physicians about the upcoming flu season when they will be inoculated irrespective of this warning? One reason might be that the letter prompts the physician to take other measures beyond the influenza vaccine, such as advising the patient about ways to avoid exposure or providing other medical treatment, or perhaps giving earlier administration of the vaccine. If these other measures or early administration affect health outcomes, the exclusion restriction would be violated.

Reasons for believing the exclusion restriction for never-takers are quite different, and appear less tenuous, than for always-takers. These patients would not receive the vaccine in any case. If these patients and their physicians did not regard the risk of flu as high enough to warrant inoculation, they might not be subject to other medical actions either, and so it might be reasonable to assume that these patients were completely unaffected by their physicians' receipt of the letter, implying that the exclusion restriction would be satisfied for the never-takers.

Given the possibility that physicians took actions other than administering the vaccine in response to the encouragement, we find it more plausible to impose the exclusion restriction for never-takers than for always-takers. Therefore, we focus on the second block of columns in Table 4. The marginal distributions of the subpopulation ITT effects suggest that the effects for compliers and always-takers are of roughly the same size. Examining their joint distribution in Figure 1, we see that the effects are somewhat negatively correlated; nevertheless, the ITT effect for always-takers appears likely to be sizable at any plausible value of the complier ITT effect. Although this result necessarily relies more heavily on the specific form of the likelihood function and prior distribution, it casts considerable doubt on scientific validity of the practical inference that would be drawn from the 'strongly identified' analysis, which imposes both exclusion restrictions, namely, that the receipt of the influenza vaccine is quite effective at reducing flu-related hospitalizations.

A similar, but weaker, conclusion can be drawn from the model with no exclusion restrictions at all. Figure 3 gives the joint distribution of the ITT effects for always-takers and compliers with no exclusion restrictions. This joint distribution is less correlated than Figure 1, and there is still considerable posterior weight given to negative values of the ITT effect for always-takers, regardless of the complier ITT effect, which is, however, likely to be more negative than the always-taker ITT effect. This result occurs to some extent because of an estimated positive ITT effect for never-takers, as shown in Figure 4, which must be regarded as implausible if the ITT effect is negative for always-takers. Certainly, our data provide little evidence that the overall ITT effect arises entirely or even largely from the effect of the vaccine on

Table 5. *Posterior distributions for parameters*

Excl. Res. Never-takers ⟶	Yes		Yes		No		No	
Excl. Res. Always-takers ⟶	Yes		No		Yes		No	
Estimand	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
β_{n00}	−2.42	(0.14)	−2.44	(0.14)	−2.85	(0.60)	−2.98	(0.74)
β_{n10}	−2.42	(0.14)	−2.44	(0.14)	−2.42	(0.15)	−2.43	(0.15)
β_{a00}	−2.22	(0.21)	−2.11	(0.22)	−2.20	(0.22)	−2.11	(0.23)
β_{a10}	−2.22	(0.21)	−2.94	(0.75)	−2.20	(0.22)	−3.08	(0.92)
β_{c00}	−2.14	(0.78)	−2.10	(0.74)	−1.43	(1.15)	−1.25	(1.08)
β_{c10}	−3.45	(1.01)	−2.55	(0.87)	−3.36	(0.88)	−2.42	(0.78)
$\beta_{0.1}$	−0.31	(0.26)	−0.29	(0.26)	−0.35	(0.28)	−0.33	(0.28)
$\beta_{0.2}$	0.34	(0.15)	0.37	(0.15)	0.36	(0.15)	0.39	(0.16)
ψ_{c0}	−1.86	(0.23)	−1.83	(0.22)	−1.83	(0.20)	−1.84	(0.21)
ψ_{c1}	0.41	(0.51)	0.32	(0.50)	0.34	(0.46)	0.29	(0.43)
ψ_{c2}	−0.17	(0.41)	−0.29	(0.64)	−0.12	(0.36)	−0.16	(0.39)
ψ_{a0}	−1.85	(0.12)	−1.86	(0.12)	−1.86	(0.12)	−1.85	(0.12)
ψ_{a1}	1.09	(0.25)	1.13	(0.24)	1.12	(0.25)	1.11	(0.24)
ψ_{a2}	0.65	(0.13)	0.66	(0.12)	0.65	(0.12)	0.64	(0.12)

hospitalizations. A conventional ITT analysis could therefore *overstate* the efficacy of the receipt of the flu shot, and clearly does not provide a fair summary of the evidence in the data for the efficacy of the flu shot itself.

In our application, the substantive interpretations of the potential effect of the letter on never-takers and always-takers are very different, and as argued above, it may be more reasonable to exclude an ITT effect for never-takers than to exclude an ITT effect for always-takers. In other applications, it may be desirable to consider the assumption that the ITT effects are the same for all types of noncompliers. Such an assumption would lead to the restriction $\beta_n = \beta_a$, which is easy to impose and implement computationally in our framework. It would also be possible to modify our simulation methods to incorporate other prior restrictions, for example, parameter constraints requiring that the ITT effect for all subpopulations to be of the same sign, or requiring the ITT effects for never-takers and always-takers to be of the same sign.

7. Conclusion

We have set out a framework for the analysis of a randomized experiment in which, instead of randomizing the treatment of interest (in our case an influenza vaccine), the researchers randomly assigned an

encouragement to give the treatment. A standard intention-to-treat analysis demonstrates that the encouragement decreases hospitalization rates. It is tempting, and rather standard applied practice, to interpret such a result as indicating a beneficial effect of the receipt of treatment, rather than just the effect of the encouragement to receive treatment; moreover, such interpretations of ITT analyses are often regarded as conservative, in the sense that the data would only support that conclusion when there really is a positive effect of the treatment. Our framework allows researchers to go beyond such an analysis to allow for different assumptions concerning the effect of the assignment for various subpopulations defined by compliance behavior. In particular, our approach of relaxing exclusion restrictions selectively by compliance type generalizes previous work on causal instrumental variables methods and facilitates a comparison of the effect of receipt of treatment under these alternative assumptions. The plausibility of these assumptions should be assessed, as in our discussion, by the underlying science of the application.

In our application we find little evidence that the flu shot had any beneficial effects. The strongest evidence is that the encouragement appears to have a similar beneficial effect on people who would have received a flu shot regardless of the encouragement, the always-takers, and on those who would only receive the flu shot when encouraged, the compliers. We interpret this result as evidence that physicians may have been inclined to provide always-takers and compliers with more or earlier preventative measures after receiving the encouragement, and that these other measures or their timing might have had a beneficial effect on reducing flu-related hospitalization. Thus our analysis illustrates the difficulty with automatically interpreting intention-to-treat effects as indicative of the effect of receipt of treatment, and provides a framework for discussing the assumptions under which such an interpretation may or may not be plausible.

APPENDIX

Details of calculations

Our approach to inference treats the latent compliance types $\mathbf{C} = (C_1, \ldots, C_n)$ as missing data and applies modern missing data technology for Bayesian models.

MARKOV CHAIN MONTE CARLO

We construct a general state space Markov chain that has the joint distribution of the model parameters π and the missing type vector $\mathbf{C}$ as its unique invariant equilibrium distribution. The Markov chain algorithm is a variant of the Metropolis–Hastings algorithm (Metropolis *et al*., 1953; Hastings, 1970; see also Tierney, 1994), which uses the Data Augmentation (DA) method of Tanner and Wong (1987). The algorithm can be described as follows. Let $(\mathbf{C}^{(j)}, \pi^{(j)})$ denote the state of the chain at time j. The state of the chain at time $j+1$ follows from applying the following steps.

First, we draw $\mathbf{C}^{(j+1)}$ according to $P(C|\pi^{(j)}, W)$, where we use $W = (\mathbf{Z}^{\text{obs}}, \mathbf{D}^{\text{obs}}, \mathbf{Y}^{\text{obs}}, \mathbf{X}^{\text{obs}})$ to simplify the notation. This conditional distribution has a simple form. Conditional on π and W, the C_i are independent of $C_j, Z_j^{\text{obs}}, D_j^{\text{obs}}, Y_j^{\text{obs}}, X_j^{\text{obs}}$ for all $j \neq i$. Then, by the monotonicity assumption,

$$\Pr(C_i = n | Z_i^{\text{obs}} = 1, D_i^{\text{obs}} = 0, Y_i^{\text{obs}}, X_i^{\text{obs}}) = 1;$$

$$\Pr(C_i = a | Z_i^{\text{obs}} = 0, D_i^{\text{obs}} = 1, Y_i^{\text{obs}}, X_i^{\text{obs}}) = 1.$$

It remains to consider the cases $(Z_i^{\text{obs}} = 1, D_i^{\text{obs}} = 1)$ and $(Z_i^{\text{obs}} = 0, D_i^{\text{obs}} = 0)$. For observations with $Z_i^{\text{obs}} = 1, D_i^{\text{obs}} = 1$,

$$\Pr(C_i = c | Z_i^{\text{obs}} = 1, D_i^{\text{obs}} = 1, Y_i^{\text{obs}}, X_i^{\text{obs}}) \propto$$

$$\Psi(c, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a)\Lambda(X_{i1}, X_{i2}, \beta_{cZ_i})^{Y_i}\left(1 - \Lambda(X_{i1}, X_{i2}, \beta_{cZ_i})\right)^{1-Y_i};$$

$$\Pr(C_i = n | Z_i^{\text{obs}} = 1, D_i^{\text{obs}} = 1, Y_i^{\text{obs}}, X_i^{\text{obs}}) = 0;$$

$$\Pr(C_i = a | Z_i^{\text{obs}} = 1, D_i^{\text{obs}} = 1, Y_i^{\text{obs}}, X_i^{\text{obs}}) \propto$$

$$\Psi(a, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a)\Lambda(X_{i1}, X_{i2}, \beta_{aZ_i})^{Y_i}\left(1 - \Lambda(X_{i1}, X_{i2}, \beta_{aZ_i})\right)^{1-Y_i}.$$

Analogous results hold for observations with $Z_i^{\text{obs}} = 0$, $D_i^{\text{obs}} = 0$:

$$\Pr(C_i = c | Z_i^{\text{obs}} = 0, D_i^{\text{obs}} = 0, Y_i^{\text{obs}}, X_i^{\text{obs}}) \propto$$

$$\Psi(c, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a)\Lambda(X_{i1}, X_{i2}, \beta_{cZ_i})^{Y_i}\left(1 - \Lambda(X_{i1}, X_{i2}, \beta_{cZ_i})\right)^{1-Y_i};$$

$$\Pr(C_i = n | Z_i^{\text{obs}} = 0, D_i^{\text{obs}} = 0, Y_i^{\text{obs}}, X_i^{\text{obs}}) \propto$$

$$\Psi(n, X_{i1}, X_{i2}, \psi_c, \psi_n, \psi_a)\Lambda(X_{i1}, X_{i2}, \beta_{nZ_i})^{Y_i}\left(1 - \Lambda(X_{i1}, X_{i2}, \beta_{nZ_i})\right)^{1-Y_i};$$

$$\Pr(C_i = a | Z_i^{\text{obs}} = 0, D_i^{\text{obs}} = 0, Y_i^{\text{obs}}, X_i^{\text{obs}}) = 0.$$

This exhausts the possible cases for $(Z_i^{\text{obs}}, D_i^{\text{obs}})$.

We then draw for the following subvectors of π in sequence, conditional on all others: $\{\psi_c, \psi_n, \psi_a\}$; $\{\beta_{c00}\}$; $\{\beta_{c10}\}$; $\{\beta_{n00}\}$; $\{\beta_{n10}\}$; $\{\beta_{a00}\}$; $\{\beta_{a10}\}$; $\{\beta_{0.1}, \beta_{0.2}\}$,
where we assume equality of the slope coefficients β_{tz1} and β_{tz2} and some other components of β as implied by the exclusion restrictions.

If we could draw directly from the appropriate conditional distributions, this would define a Gibbs sampler (see Geman and Geman, 1984; and Gelfand and Smith, 1990), which in our specification is rather difficult to do; however, it is straightforward to calculate the (complete-data) posterior density up to a normalizing constant at any parameter value, so we can use Metropolis–Hastings steps. To draw $\psi = (\psi_c, \psi_n, \psi_a)$, we draw *candidate* values ψ^{cand} from a density $g(\psi|\pi^{(j)})$. The candidate draw is accepted with probability

$$\alpha = \min\left\{\frac{p(\beta^{(j)}, \psi^{\text{cand}}|W, \mathbf{C})}{p(\beta^{(j)}, \psi^{(j)}|W, \mathbf{C})} \cdot \frac{g(\psi^{(j)}|\beta^{(j)}, \psi^{\text{cand}})}{g(\psi^{\text{cand}}|\beta^{(j)}, \psi^{(j)})}, 1\right\},$$

where p is the posterior density, up to a normalizing constant, of the parameter vector. For the candidate density g, we use a vector of scaled t random variables with five degrees of freedom, centered at $\psi^{(j)}$. This has the convenient property that

$$g(\psi^{\text{cand}}|\beta, \psi^{(j)}) = g(\psi^{(j)}|\beta, \psi^{\text{cand}}),$$

simplifying the expression for α slightly.

The scaling factors were chosen based on preliminary runs of the chain. It is desirable to strike a balance between rejecting too often and rejecting too infrequently, so that the resulting chain will cover the support of the target distribution relatively efficiently—not staying at the same point too much but also not taking steps that are too small.

To assess convergence of interative simulation methods, we use, following Gelman and Rubin (1992), multiple chains from some overdispersed initial distribution and compare their realizations. As the initial distribution, we take a multivariate normal approximation derived from a simulation based on a single chain, and inflate the variance matrix. The chains for the various models appear to converge in only a few hundred to one or two thousand iterations, even when the Metropolis candidate distributions are not tuned too carefully. Thus we discarded the first 2000 iterations of every chain used in the analysis. For the posterior distributions, the chains were run for 98 000 iterations after the burn-in stage, saving every 25th iteration. For the prior distributions, the chains were run for 48 000 iterations after burn-in, saving every 10th iteration.

REFERENCES

ANGRIST, J. D. (1990). Lifetime earnings and the Vietnam era draft lottery: evidence from Social Security Administrative Records. *American Economic Review* **80**, 313–335.

ANGRIST, J. D. AND KRUEGER, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* **106**, 979–1014.

ANGRIST, J. D., IMBENS, G. W. AND RUBIN, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association* **91**, 444–455.

BAKER, S. G. AND LINDEMAN, K. (1994). The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statistics in Medicine* **13**, 2269–2278.

BALKE, A. AND PEARL, J. (1993). Nonparametric bounds on causal effects from partial compliance data. Technical Report R-199, Computer Science Department, University of California, Los Angeles.

COX, D. R. (1958),. *The Planning of Experiments*. New York: John Wiley.

DEMPSTER, A. P., LAIRD, N. AND RUBIN, D. B. (1977). Maximum likelihood estimation from incomplete data using the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, Series B **39**, 1–38.

EFRON, B. AND FELDMAN, D. (1991). Compliance as an explanatory variable in clinical trials (with discussion). *Journal of the American Statistical Association* **86**, 9–26.

FRANCIS, T. AND MAGILL, T.P. (1937). The antibody response of human subjects vaccinated with the virus of influenza. *Journal of Experimental Medicine* **65**, 251.

GELFAND, A. E. AND SMITH, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

GELMAN, A. AND RUBIN, D. B. (1992). Inference from iterative simulations using multiple sequences. *Statistical Science* **7**, 457–511.

GEMAN, S. AND GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

HAAVELMO, T. (1943). Statistical implications of a system of simultaneous equations. *Econometrica* **11**, 1–12.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

HEARST, N., NEWMAN, T. AND HULLEY, S. (1986). Delayed effects of the military draft on mortality: a randomized natural experiment. *New England Journal of Medicine* **314**, 620–624.

HECKMAN, J., AND ROBB, R. (1985). Alternative methods for evaluating the impact of interventions. In *Longitudinal Analysis of Labor Market Data*, Eds J. Heckman and B. Singer. New York: Cambridge University Press.

HOLLAND, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81**, 945–970.

HOLLAND, P. (1988). Causal inference, path analysis, and recursive structural equations models. In *Sociological Methodology*, Chapter 13. Washington: American Sociological Association.

HOUSWORTH, J. AND LANGMUIR, A. D. (1974). Excess mortality from epidemic influenza, 1957–1966. *American Journal of Epidemiology* **100**, 40–48.

IMBENS, G. W. AND ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62**, 467–476.

IMBENS, G. W. AND RUBIN, D. B. (1997a). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* **25**, 305–327.

IMBENS, G. W. AND RUBIN, D. B. (1997b). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies* **64**, 555–574.

LITTLE, R. AND YAU, L. (1998). Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model. *Psychological Methods*, **3**, 147–159.

MANSKI, C. F. (1990). Non-parametric bounds on treatment effects. *American Economic Review, Papers and Proceedings* **80**, 319–323.

MCCLELLAN, M. AND NEWHOUSE, J. P. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality. *Journal of the American Medical Association* **272**, 859–866.

MCDONALD, C., HIU, S., AND TIERNEY, W. (1992). Effects of computer reminders for influenza vaccination on morbidity during influenza epidemics. *MD Computing* **9**, 304–312.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. AND TELLER, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1091.

PATRIARCA, P. A., WEBER, J. A., PARKER, R. A., HALL, W. N., KENDAL, A. P., BERGMAN, D. J. AND SCHONBERGER, L. B. (1985). Efficacy of influenza vaccine in nursing homes: reduction in illness and complications during an influenza A (H3/N2) epidemic. *Journal of the American Medical Association* **253**, 1136–1139.

PEARL, J. (1995). Causal diagrams for empirical research. *Biometrika* **82**, 669–688.

PERMUTT, T. AND HEBEL, J. (1989). Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight. *Biometrics* **45**, 619–622.

REIERSOL, O. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica* **9**, 1–24.

ROBINS, J. M. (1989). The analysis of randomized and nonrandomized aids treatment trials using a new approach to causal inference in longitudinal studies. *Health Service Research Methodology: A Focus on AIDS*, Eds L. Sechrest, H. Freeman, and A. Bailey. Rockville, MD: NCHSR, U.S. Public Health Service.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.

RUBIN, D. B. (1977). Assignment to a treatment group on the basis of a covariate. *Journal of Education Statistics* **2**, 1–26.

RUBIN, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *Annals of Statistics* **6**, 34–58.

RUBIN, D. B. (1980). Discussion of 'randomization analysis of experimental data in the Fisher randomization test', by Basu. *Journal of the American Statistical Association* **75**, 591–593.

RUBIN, D. B. (1990a). Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science* **5**, 472–480.

RUBIN, D. B. (1990b). Formal modes of statistical inference for causal effects. *Journal of Statistical Planning and Inference* **25**, 279–292.

RUBIN, D. B. AND THAYER (1978). Relating tests given to different samples. *Psychometrika* **43**, 3–10.

SOMMER, A., AND ZEGER, S. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine* **10**, 45–52.

TANNER, M. AND WONG, W. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* **82**, 528–550.

TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics* **22**, 1701–1762.

ZELEN, M. (1979). A new design for randomized clinical trials. *New England Journal of Medicine* **300**, 1242–1245.

ZELEN, M. (1990). Randomized consent designs for clinical trials: an update. *Statistics in Medicine* **9**, 645–656.

[*Received July 2, 1999. Revised September 8, 1999*]