Inference for the Complier-Average Causal Effect From Longitudinal Data Subject to Noncompliance and Missing Data, With Application to a Job Training Assessment for the Unemployed

Linda H.Y. YAU and Roderick J. LITTLE

Longitudinal studies involving human participants are often complicated by subjects who do not comply with their treatment assignment or do not provide complete data. A treatment effect of interest in the presence of noncompliance is the complier-average causal effect (CACE; Imbens and Rubin 1997a), which is the treatment effect for subjects who would comply regardless of the assigned treatment. Imbens and Rubin (1997a,b) proposed maximum likelihood and Bayesian inferential methods for CACE, which make explicit assumptions for causal inference in the presence of noncompliance and are more efficient than standard instrumental variable methods. A model for inference about the CACE based on this approach is developed which allows for the inclusion of baseline covariates and handles missing data in the repeated outcome measures. Our methods are applied to a randomized trial of a job training intervention for unemployed workers. Results suggest that the intervention trial significantly reduced depression for high-risk compliers up to six months postintervention but not for low-risk compliers.

KEY WORDS: Attrition; Bayesian inferential method; Causal inference; EM algorithm; Gibbs sampler; Instrumental variable.

1. INTRODUCTION

Noncompliance and attrition often complicate the statistical analysis of longitudinal studies. Noncompliance arises when subjects fail to comply with their assigned treatment. Attrition arises when subjects prematurely drop out from the study due to side effects, loss to follow-up, or subject withdrawal. We address these issues in the context of the JOBS II intervention trial (Vinokur, Price, and Schul 1995), a study that tested the effectiveness of a job training intervention for preventing deterioration in mental health as a result of job loss, and facilitating high quality reemployment. After administration of a questionnaire to collect baseline information and determine their eligibility, 1801 subjects were randomly assigned to either an experimental treatment group or to a control group. The experimental treatment consisted of a five half-day job-search seminar designed to help the participants enhance their job search strategies. The control treatment was an instruction booklet on job search. This booklet was mailed to respondents not invited to the seminar and was also mailed to respondents invited to the seminar. Follow-up questionnaires were mailed to these respondents two months (T2), six months (T3), and two years (T4) after the week of the intervention seminar. Demographic variables and measures of depression, financial strain, assertiveness, risk, distress, role and emotional functioning, job-search efficacy, self-esteem, mastery, and reemployment were obtained or constructed from the questionnaires. Noncompliance arises in that 46% of subjects who were randomized to the seminar did not show up, and in effect received the control treatment. Subjects in this group continued to be followed and have outcome data collected. Attrition refers to losses to follow-up from T2 to T4 among subjects in both the experimental and the control groups. Nonmonotone patterns of missing data where subjects were missing at some follow-up times and reentered the study are also handled by our methods.

Two kinds of analyses can be distinguished in randomized studies such as JOBS II. An intent-to-treat (IT) analysis compares the outcome of subjects by randomization groups (an "as-randomized" analysis), ignoring compliance information. The IT-effect measures the effect of treatment randomization rather than the effect of treatment for those who actually received it. The IT estimator is protected from the selection bias by randomized treatment allocation, but is distorted as a measure of the effect of treatment by noncompliance. Alternatively, subjects can be classified according to the treatments actually received (an "as-treated" analysis). Then the problem is that randomization is violated and confounding factors associated with switching potentially corrupt the causal interpretation of treatment effects. Because both of these approaches to analysis have problems, in practice both "as-randomized" and "as-treated" analyses are often carried out when assessing the effects of treatments. Attempts have also been made to provide adjustments or bounds to the treatment effect in the presence of noncompliance (Balke and Pearl 1997; Baker 1998; Robins and Finkelstein 2000). Recently, Rotnitzky, Scharfstein, Su, and Robins (2001) proposed a semiparametric cause-specific selection model that can be used to assess sensitivity of the treatment effect in the presence of attrition.

1.1 The CACE

Our focus here is on a particular form of "as-treated" analysis where inference concerns the complier-average causal effect (CACE) (Imbens and Rubin 1997a), the average effect of treatment in the subpopulation of compliant subjects. The

Linda H.Y. Yau is Senior Statistician, GlaxoSmithKline, Research Triangle Park, NC 27709 (E-mail: *ly20993@gsk.com*). Roderick J. Little is Professor, Department of Biostatistics, School of Public Health, University of Michigan. Ann Arbor, MI 48109 (E-mail: *rlittle@umich.edu*). This work was supported by NIMH grant P30 MH38330-12 and was completed in part while the first author was graduate student, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48019. The authors thank Rick Price, Amiram Vinokur for their valuable support and suggestions, and Don Rubin for his helpful advice.

CACE is commonly known as the local average treatment effect (LATE) in Economics literature (e.g., Imbens and Angrist 1994; Imbens and Rubin 1997b). In a previous article (Little and Yau 1998), we applied maximum-likelihood (ML)-based methods of inference for the CACE proposed by Angrist, Imbens, and Rubin (1996, henceforth AIR) and Imbens and Rubin (1997a,b) to a scalar depression outcome from the JOBS II study, ignoring issues of attrition. This article generalizes that work by formulating and fitting a model for longitudinal JOBS II data subject to attrition and noncompliance, using ML and Bayesian inferential methods.

To define the CACE more precisely in our setting, consider a longitudinal study involving initial allocation to an active or an experimental treatment ($R = 1$) and a control treatment ($R = 0$). Let Y be a $\tau \times 1$ vector of repeated measures on the outcome of interest, and initially assume that this vector is fully recorded for all subjects. Let $T_i(R_i)$ be the treatment received for respondent i randomized to treatment R_i. Let μ_1 and μ_0 be the $\tau \times 1$ population mean potential outcomes of respondents assigned to experimental and control treatments, respectively, perhaps in subpopulations defined by the values of between-subject covariates. Then the average IT effect of treatment, which AIR call the population average causal effect of treatment assignment R on outcome Y, is defined as $\delta = \mu_1 - \mu_0$. If $\bar{y}_1$ and $\bar{y}_0$ are the sample mean vectors of outcomes for subjects randomized to the experimental and control groups respectively, then $\bar{y}_1 - \bar{y}_0$ is an unbiased estimate of δ. The average treatment effect for compliers, or the CACE, is defined as:

$$\delta_c = \mu_{c1} - \mu_{c0}, \tag{1}$$

where μ_{c1} and μ_{c0} are the mean outcomes of the population of compliers when assigned to experimental and control treatments, respectively. The CACE is the average causal effect of treatment received T on outcome Y restricted to compliers. It is arguably a more relevant effect of interest than the IT effect, because only those who took the treatment would be able to benefit from the treatment if there was in fact a treatment effect. If there is a perfect compliance in all the randomized groups, that is, $R_i = T_i(R_i)$ for all participants, then the CACE is the same as the average IT effect of R on Y. If the compliance status of all the respondents in the study is known, then δ_c can be estimated simply by the difference between the treatment and the control means in the subsample of compliers. However, in the JOBS II study, the compliance status of subjects in the control group is unknown because we do not know if they would have complied with the experimental treatment if they had been assigned to it. This complicates the estimation of the CACE.

1.2 Estimation of the CACE

An instrumental variable (IV) estimator of the CACE in the presence of noncompliance is given by:

$$d_c = (\bar{y}_1 - \bar{y}_0)/p_c, \tag{2}$$

(e.g., Bloom 1984) where $\bar{y}_1$ and $\bar{y}_0$ are defined above and p_c is the sample proportion of compliers in the experimental group, and estimates the population proportion of compliers π_c.

Imbens and Rubin (1997a, b) proposed a causal framework for inference about the CACE that yields an alternative to (2). The extension of their framework to a vector outcome, ignoring attrition, is immediate. Imbens and Rubin (1997a, b) classify participants into four types: compliers, defiers, never-takers, and always-takers. Compliers are people who would adopt the treatment assigned, defiers are those who would adopt the opposite treatment to their assignment, never-takers are people who always take the control treatment regardless of what they are assigned, and always-takers are people who always take the active treatment regardless of what they are assigned. So, for compliers, $T_i(1) = 1$ and $T_i(0) = 0$; for defiers, $T_i(1) = 0$ and $T_i(0) = 1$; for never-takers, $T_i(1) = T_i(0) = 0$; and for always-takers, $T_i(1) = T_i(0) = 1$. Let C_i denote the compliance status of the participant, where $C_i = 1$ for compliers and $C_i = 0$ for noncompliers, that is, defiers, never-takers, and always-takers. In practice, knowledge of the compliance status C_i of subjects is very often incomplete. Imbens and Rubin (1997a) treat compliance as a missing-data problem. Specifically, if a subject is assigned to the new treatment and complies, then $T_i(1) = 1$, and that subject may be a complier (if $T_i(0) = 0$) or an always-taker (if $T_i(0) = 1$). If a subject is assigned to the new treatment and fails to comply, then $T_i(1) = 0$ and that subject may be a never-taker (if $T_i(0) = 0$) or a defier (if $T_i(0) = 1$). If a subject is assigned to the control treatment and complies, then $T_i(0) = 0$, and that subject may be a complier (if $T_i(1) = 1$) or an never-taker (if $T_i(1) = 0$). If a subject is assigned to the control treatment and obtains the new treatment, then $T_i(0) = 1$, and that subject may be an always-taker (if $T_i(1) = 1$) or a defier (if $T_i(1) = 0$).

Let $y_i(R_i, T_i)$ be the potential $\tau \times 1$ vector of outcomes for respondent i, who was randomized to group R_i and received treatment T_i. Implicit in this notation is the Stable Unit Treatment Value Assumption (SUTVA) (Lewis 1963; Rubin 1978; Imbens and Rubin 1997a, b), defined as:

Assumption (1) (SUTVA). The potential outcome for each individual i does depend on the treatment status of other individuals in the sample.

With Assumption (1), the individual level causal effect can be defined. The unit level causal effect of R on T for respondent i is defined by $T_i(1) - T_i(0)$, and the unit level causal effect of R on Y is defined by $y_i(1, T_i(1)) - y_i(0, T_i(0))$. Note that the causal effect of R on Y is generally not observable, because respondents are assigned to either the experimental or the control treatment but not to both. However, the average causal effects can be estimated by the average causal effects over groups of respondents in a randomized trial. In addition, the following assumption allows us to identify the causal effect of T on Y meaningfully:

Assumption (2) (Exclusion Restriction). The treatment assignment R is unrelated to the potential outcomes given the treatment received T—*Exclusion Restriction of Treatment Assignment given Treatment Received.*

The Exclusion Restriction implies that $y_i(1, T_i) = y_i(0, T_i)$ for $T_i = 0, 1$, so that the potential outcomes can now be defined as a function of T alone, namely, $y_i(T_i)$. Also, the unit level causal effect of T on Y for respondent i is then defined by

$y_i(1) - y_i(0)$. Under Assumptions (1), (2), and the following additional assumptions, the CACE, δ_c, given by (1) (or referred to as the LATE in AIR) can be estimated from the data (Imbens and Rubin 1997a, b):

Assumption (3) (Monotonicity). There are no defiers—*Monotonicity of Treatment Assignment and Treatment Received*;

Assumption (4) (Nonzero Denominator). For the population of compliers, the proportion of compliers taking the active treatment is nonzero; and

Assumption (5) (Randomization). Treatment assignment is random.

Under these assumptions, the CACE (δ_c) is the same as the instrumental variable estimand (IVE) ($\delta_c^* = \delta/\pi_c$), that is, $\delta_c = \delta_c^*$, hence the CACE or the average of the unit level causal effects $y_i(1) - y_i(0)$ for the subpopulation of compliers characterized by $T_i(1) = 1$ and $T_i(0) = 0$ can be estimated by taking the ratio of the average of the unit level causal effect of R on Y and the average of the unit level causal effect of R on T. This formulation makes explicit assumptions that were hidden in previous derivations. Bloom (1984) argued that the IV estimator (2) is valid if (a) there are no missing data and (b) the randomization assumption (5) holds. Under these assumptions, d_c is an unbiased estimator of the IVE δ_c^*, but this does not imply that $\delta_c = \delta_c^*$, which is required for Bloom's estimator to be a valid estimator of the CACE. The additional assumptions (1)–(4) are needed to assure that $\delta_c = \delta_c^*$. Given assumptions (1), (3)–(5) (5), the population mean potential outcomes for respondents randomized to the experimental and control groups, namely, μ_1 and μ_0, can be expressed as the weighted proportions of the mean potential outcomes of the population of compliers and noncompliers. Specifically, $\mu_1 = \pi_c\mu_{c1} + (1-\pi_c)\mu_{nc1}$ and $\mu_0 = \pi_c\mu_{c0} + (1-\pi_c)\mu_{nc0}$, where μ_{nc1} and μ_{nc0} are the mean outcomes of the population of noncompliers when assigned to the experimental and control treatments, respectively. In particular, Assumption (2), the exclusion restriction, plays a key role, because by virtue of the assumption, $\mu_{nc1} = \mu_{nc0}$, which implies that $\mu_1 - \mu_0 = \pi_c(\mu_{c1} - \mu_{c0})$, or equivalently, $\delta_c^* = \delta_c$. Another advantage of the Imbens and Rubin (1997a) formulation is that it suggests more efficient estimators of the CACE than the IV estimator (2).

In this article, we extend Imbens and Rubin's methods to the estimation of CACE for longitudinal data subject to noncompliance and attrition, and apply the methodology to the JOBS II trial. Section 2 provides more information on the design of JOBS II and the problems of noncompliance and attrition in that study. Section 3 presents the longitudinal model used for the JOBS II trial, and discusses ML and Bayesian inference. Results of the JOBS II analysis appear in Section 4, and Section 5 presents conclusions and further discussion.

2. JOBS II INTERVENTION PROJECT

2.1 Design of JOBS II

Past research has shown that job loss has adverse effects on workers' social and psychological functioning, physical health, and on the family (Catalano and Dooley 1977; Cobb and Kasl 1977; Justice and Duncan 1977; Dew, Bromet, and Schulberg 1987; Vinokur, Caplan, and Williams 1987; Kessler, Turner, and House 1988, 1989; Catalano 1991). Although it is perceived that the problems that result from unemployment have to be addressed by national and state social and economic policies (Blinder 1987), efforts can be undertaken to reduce the social and psychological impact of unemployment at the local level. The JOBS Intervention Project developed at the University of Michigan was designed to test a preventive intervention for unemployed workers. The goals of the intervention were to prevent deterioration in mental health of unemployed workers as a result of job loss and to promote high quality reemployment. Analyses of the data provided evidence that the intervention had accomplished its goals (Caplan, Vinokur, Price, and van Ryn 1989). In addition, it was demonstrated that the intervention provided beneficial mental health effects primarily to a subset of respondents who were identified at a high risk of experiencing clinical significant setback in mental health such as experiencing a depression episode (Price, van Ryn, and Vinokur 1992). High-risk respondents were those identified with higher combined scores on depressive symptoms, financial strain, and low assertiveness at pretest. These variables are prominent risk factors for poor mental health and continued unemployment of persons who lose their job (Kessler, Turner, and House 1988). In this article, we focus on the analysis of the JOBS II study, which was an extension of the original JOBS project that included advance screening and oversampling of high-risk job losers who might benefit most from the intervention. Our primary variable of interest is depression, because depression symptoms have been shown to reduce the chances of employment due to their effects on the motivation for, and effectiveness of, job-search behavior (Hamilton, Hoffman, Broman, and Rauma 1993). It is anticipated that if the JOBS intervention is more effective in reducing the depression symptomatology among high-risk than low-risk respondents, it would also be more effective in improving reemployment of the high-risk than low-risk respondents.

JOBS II compared an experimental treatment consisting of a five half-day job-search seminar, which helped the unemployed to enhance their job-search strategies, with a control treatment consisting of an instruction booklet on job search. Respondents were recruited at the four offices of the Michigan Employment Security Commission in southeastern Michigan. Among the 31,560 people contacted, over 23,000 were ineligible for participation because they were new entrants to the labor market, already employed, or were just accompanying others in line. Only 3402 respondents who filled out a screening questionnaire (T0), met all the screening criteria (reporting to be looking for a job, not being on strike or expecting to be recalled, or planning to retire in the next 2 years, within 13 weeks of losing job, no preference between control and experimental programs, absence of very high depression score). To prevent selection and attrition bias due to a strong preference for the experimental or the control treatment, respondents who indicated a preference for one of the treatments or who refused both treatments were excluded from the sample. There were 2,445 low-risk and 957 high-risk respon-

dents among the 3,402 who met the criteria. The oversampling of the high-risk respondents in the sample was achieved by randomly selecting 1,507 low-risk respondents from the 2,445 respondents, so that the proportion of high-risk respondents in the sample was increased from 28% to 39%. As a result, a total of 2,464 respondents were invited and randomized to the field study. All respondents subject to randomization were mailed a pretest (T1) questionnaire. A total of 1,801 respondents who returned their T1 questionnaire were enrolled to the study, with 552 and 1,249 respondents in the control and experimental groups, respectively. After the intervention seminar, respondents randomized to the experimental group were also mailed the same instructional booklet sent to the control group. Follow-up questionnaires were mailed to all respondents two months (T2), six months (T3), and two years (T4) after the week of the intervention seminar. Demographic variables such as age, sex, education, marital status, occupation, family income, and race were collected from the survey questionnaires. In addition, measures for depression, financial strain, assertiveness, risk, distress, role and emotional functioning, job-search efficacy, self-esteem, internal control orientation, and mastery were constructed from the questionnaires, based on responses from a 5-point rating scale. Specifically, depression was measured from the responses to a 11-item list based on the Hopkins Symptom Checklist (Derogatis, Lipman, Rickles, Uhlenuth, and Covi 1974); financial strain was measured from responses to a 3-item index; assertiveness was based on responses from a 4-item index; risk score was computed by a formula developed by Price et al. (1992) using the depression, financial strain, and assertiveness scores; distress was measured from an 18-item index; role and emotional functioning from a 15-item index (Caplan et al. 1984); job-search efficacy from a 6-item list; self-esteem from an 8-item list from Rosenberg's (1965) self-esteem scale; internal control orientation from an 10-item list based on the original Rotter I–E scale (Gurin, Gurin, and Morrison 1978); and mastery was computed by the mean scores of job-search efficacy, self-esteem, and internal control orientation.

2.2 Noncompliance and Attrition in JOBS II

About 46% of the subjects who were randomized to the seminar did not show up, and are treated as noncompliers. Subjects in this group continued to be followed and have outcome data collected. Missing values in the longitudinal data at times T2–T4 occur when follow-up questionnaires are not returned from respondents who were randomized to the experimental and the control groups. Response rates at T2, T3, and T4 were 80%, 87%, and 79%, respectively. We assume that if compliance status was known, the missing-data mechanism would be ignorable for likelihood inferences (Rubin 1976; Little and Rubin 1987; Little 1995), so that a model is not needed for the attrition mechanism. This is the "ignorability" assumption for the missing-data mechanism. It implies, in particular, that the missing data are missing at random conditional on compliance status, in that missingness does not depend on the values of missing outcomes, after conditioning on treatment group, compliance status, baseline covariates, and the values of observed outcomes. Information to assess this assumption is not available, but it is rendered more plausible by the considerable amount of information available to characterize incomplete cases.

In the JOBS II setting, the intervention seminar was only offered to those randomized to the experimental group. Those randomized to the control group had no way of receiving the intervention, so in terms of the AIR terminology, there are no defiers or always-takers in the study, only compliers and never-takers. Let R_i be the randomization indicator, taking the value 1 if the ith respondent was randomized to the experimental group, and otherwise 0; let T_i be the indicator for the treatment actually received by the ith respondent, taking the value 1 if the ith respondent received the experimental treatment and otherwise 0; and let $C_i = 1$ if the ith respondent is a complier and $C_i = 0$ if the ith respondent is a never-taker. For compliers, $T_i(R_i) = R_i$ for $R_i = 0, 1$; for never-takers, $T_i(R_i) = 0$ for $R_i = 0, 1$. Table 1 summarizes the missing data pattern of C_i in JOBS II. There are no missing data for C_i in the experimental group, the shows (N_{11}) in the intervention would solely be compliers, whereas the no-shows (N_{10}) would all be never-takers. There are both compliers and never-takers in the control group (N_{00}), and the compliance status of respondents in that group is unknown, because we do not know if they would comply if assigned to the treatment. The estimation of the CACE, δ_c, in this application is complicated by the fact that compliance status is unknown for respondents in the control group.

We now assess the assumptions concerning noncompliance underlying the estimation of CACE in the JOBS II setting. Assumptions (3) and (4) are clearly satisfied in JOBS II. The combination of these two assumptions gives the *strict monotonicity of treatment assignment on treatment received*, which implies the absence of defiers and the presence of at least one complier. This condition is satisfied because participants randomized to the control group had no way of attending the intervention, so the presence of defiers is impossible in the JOBS II design. Also, there is more than one complier because more than one subject participated in the intervention.

Assumptions (1) and (2) are more questionable in this application. The SUTVA (1) implies that the potential outcome of a participant does not depend on the treatment status of other participants. This assumption might not hold if a participant actively contributes to the discussions in the seminar and hence affects the outcomes of other participants in the seminar. The potential outcome of participants in the treated group then would be dependent on whether or not this influential participant is selected to be in the experimental group. Although the SUTVA is a nontrivial assumption for interventions like JOBS II that involve group interactions, it appears difficult to correct for violations of this assumption, at least without

Table 1. Missing Data Pattern for Compliance Status C_i

		C_i		
R_i	T_i	0	1	Total
0	0	?	?	N_{00}
1	0	N_{10}	0	N_{10}
1	1	0	N_{11}	N_{11}
Total		N_n	N_c	N

further information on how the participants interacted during the seminar. Assumption (2), the Exclusion Restriction of Treatment Assignment given Treatment Received, states that the outcome is independent of the treatment assignment given the actual treatment received. This assumption together with SUTVA and treatment randomization are the key assumptions for the treatment assignment indicator R to be an instrumental variable. It would be violated if the effect of no treatment for those randomized to the experimental group is not the same as the effect of no treatment for those in the control group; this might happen if a subject was randomized to the experimental group and did not comply, was demoralized by the inability to take advantage of the opportunity, whereas the same person randomized to control treatment would be less demoralized because the intervention was never offered.

The missing-data patterns of the change in depression scores from baseline T0 for the high- and low-risk groups from T2 to T4 are shown in Table 2. The 3-digit indicator in the tables represents the missing-data pattern for times T2, T3, and T4, with a one indicating missing outcome at that time, and a zero indicating observed outcome at that time. For example, a 010 pattern, represents observed outcomes at T2, T4, and a missing outcome at T3. About 64% of the high-risk group and 67% of the low-risk group had observed outcomes at all the time points, while others have one or more missing outcomes from T2 to T4. The attrition pattern for the same participant type are similar among the high-risk and low-risk groups. The proportion of experimental compliers who had complete data at all the follow-up times is greater than that of the control group or the experimental noncompliers, which is reflective of the complying behavior and also suggestive of possible positive effects of the experimental treatment. Moreover, the proportion of experimental noncompliers who did not have any follow-up data is about twice of that in the control group (8.7% in experimental noncompliers versus 4.7% in control for high-risk respondents; 9.7% in experimental noncompliers versus 4.7% in control for low-risk respondents). This finding might be seen as indirect evidence against the exclusion restriction, because subjects demoralized by failure to take advantage of the intervention might be expected to drop out of the study to a greater extent than the controls. However, the experimental noncompliers are a self-selected subgroup whereas the controls are not, and differences in attrition between these groups could reflect their lack of comparability rather than evidence that the exclusion restriction is violated.

3. MODEL

Our proposed model is an extension of the general location model (Olkin and Tate 1961) for mixed continuous and categorical data with missing data (Little and Schluchter 1985) to include fixed covariates. For subject i ($i = 1, \ldots, N$), let y_i denote the $\tau \times 1$ vector of outcomes, X_{Y_i} be a set of fixed covariates predictive of y_i, R_i be the randomization group indicator (1 for treatment, 0 for control), and C_i the compliance indicator (1 for complier, 0 for never-taker). In our full model,

Table 2. Missing Data Pattern for Difference in Depression From T2 to T4

Missing data pattern	(a) High-Risk Group Freq.	%	Cum. freq.	Cum. %	Missing data pattern	(b) Low-Risk Group Freq.	%	Cum. freq.	Cum. %
			Control					*Control*	
000	126	66.3	126	66.3	000	181	71.5	181	71.5
001	18	9.5	144	75.8	001	23	9.1	204	80.6
010	4	2.1	148	77.9	010	4	1.6	208	82.2
011	6	3.2	154	81.1	011	6	2.4	214	84.6
100	16	8.4	170	89.5	100	15	5.9	229	90.5
101	7	3.7	177	93.2	101	6	2.4	235	92.9
110	4	2.1	181	95.3	110	6	2.4	241	95.3
111	9	4.7	190	100.0	111	12	4.7	253	100.0
			Experimental Noncompliers					*Experimental Noncompliers*	
000	99	53.8	99	53.8	000	149	53.6	149	53.6
001	20	10.9	119	64.7	001	30	10.8	179	64.4
010	4	2.2	123	66.8	010	5	1.8	184	66.2
011	6	3.3	129	70.1	011	6	2.2	190	68.3
100	24	13.0	153	83.2	100	31	11.2	221	79.5
101	9	4.9	162	88.0	101	16	5.8	237	85.3
110	6	3.3	168	91.3	110	14	5.0	251	90.3
111	16	8.7	184	100.0	111	27	9.7	278	100.0
			Experimental Compliers					*Experimental Compliers*	
000	152	71.4	152	71.4	000	240	74.1	240	74.1
001	16	7.5	168	78.9	001	19	5.9	259	79.9
010	16	7.5	184	86.4	010	9	2.8	268	82.7
011	3	1.4	187	87.8	011	11	3.4	279	86.1
100	10	4.7	197	92.5	100	22	6.8	301	92.9
101	5	2.3	202	94.8	101	12	3.7	313	96.6
110	5	2.3	207	97.2	110	3	0.9	316	97.5
111	6	2.8	213	100.0	111	8	2.5	324	100.0

NOTE: Missing data pattern: 0—Observed, 1—Missing, Freq.—Frequency, Cum.—Cumulative.

the outcomes y_i given C_i, R_i, and covariates are assumed to follow a multivariate normal model with means

$$E(y_{it}) = \beta_{0,t} + \beta_{C,t}C_i + \beta_{CR,t}C_iR_i + \beta_{X,t}^{\mathrm{T}}X_{Y_i} + \beta_{CX,t}^{\mathrm{T}}C_iX_{Y_i} + \beta_{CRX,t}^{\mathrm{T}}C_iR_iX_{Y_i}, \quad (3)$$

for $t = 1, \ldots, \tau$, and unstructured covariance matrix Σ. This model implies that for a subject with covariates X_{Y_i}, the mean outcome at time t is:

$$\mu_{n,it} = \beta_{0,t} + \beta_{X,t}^{\mathrm{T}}X_{Y_i} \text{ for never-takers,}$$
$$\mu_{c0,it} = \beta_{0,t} + \beta_{C,t} + (\beta_{X,t} + \beta_{CX,t})^{\mathrm{T}}X_{Y_i} \text{ for control compliers,}$$
$$\mu_{c1,it} = \beta_{0,t} + \beta_{C,t} + \beta_{CR,t} + (\beta_{X,t} + \beta_{CX,t} + \beta_{CRX,t})^{\mathrm{T}}X_{Y_i} \text{ for experimental compliers.}$$

Thus the effect of compliance for controls at time t is $\beta_{C,t} + \beta_{CX,t}^{\mathrm{T}}X_{Y_i}$, and the CACE is $\beta_{CR,t} + \beta_{CRX,t}^{\mathrm{T}}X_{Y_i}$. We also fit reduced models where $\beta_{CX,t} = \beta_{CRX,t} = 0$, in which case the CACE at time t is assumed constant for all values of the covariates. Note that the main effect for R and interactions of R with X_Y are absent from the mean structure in (3). This is a consequence of the Exclusion Restriction Assumption (2), which asserts that the mean outcomes for never-takers with covariates X_Y are the same in both the treatment groups, namely $\mu_{n,it}$.

The distribution of C_i given R_i and covariates is assumed to have a Bernoulli distribution with probability of compliance π_{ci}, where

$$\text{logit}\,(\pi_{ci}) = \alpha_0 + \alpha_X^{\mathrm{T}}X_{C_i}, \quad (4)$$

and X_{C_i} is the value of a vector of covariates X_C assumed to predict compliance for subject i. X_Y and X_C can have variables in common. Let $\theta = (\alpha, \beta, \Sigma)$ denote the complete set of parameters in this model. The incomplete-data log-likelihood is given by

$$\begin{aligned} l(\theta|R, T, Y_{\text{obs}}, X_Y, X_C) &= -\frac{1}{2}N\{\tau\ln(2\pi) + \ln|\Sigma|\} \\ &+ \sum_{i\in\{R_i=1\}}[(1-C_i)\ln(1-\pi_{ci}) + C_i\ln(\pi_{ci})] \\ &- \sum_{i\in\{R_i=1,T_i=0\}}\frac{(y_{\text{obs},i}-\mu_{n,\text{obs},i})^{\mathrm{T}}\Sigma_{\text{obs},i}^{-1}(y_{\text{obs},i}-\mu_{n,\text{obs},i})}{2} \\ &- \sum_{i\in\{R_i=1,T_i=1\}}\frac{(y_{\text{obs},i}-\mu_{c1,\text{obs},i})^{\mathrm{T}}\Sigma_{\text{obs},i}^{-1}(y_{\text{obs},i}-\mu_{c1,\text{obs},i})}{2} \\ &+ \sum_{i\in\{R_i=0,T_i=0\}}\ln\Bigg\{(1-\pi_{ci}) \\ &\times\exp\left[-\frac{(y_{\text{obs},i}-\mu_{n,\text{obs},i})^{\mathrm{T}}\Sigma_{\text{obs},i}^{-1}(y_{\text{obs},i}-\mu_{n,\text{obs},i})}{2}\right] \\ &+\pi_{ci}\exp\left[-\frac{(y_{\text{obs},i}-\mu_{c0,\text{obs},i})^{\mathrm{T}}\Sigma_{\text{obs},i}^{-1}(y_{\text{obs},i}-\mu_{c0,\text{obs},i})}{2}\right]\Bigg\}, \end{aligned}$$

where $y_{\text{obs},i}$ are the observed components of y_i for subject i, with mean $\mu_{n,\text{obs},i}$ if subject i is a never-taker, $\mu_{c0,\text{obs},i}$ if subject i is a control complier and $\mu_{c1,\text{obs},i}$ if subject i is an experimental complier, $\Sigma_{\text{obs},i}$ is the covariance matrix of $y_{\text{obs},i}$. ML estimates of θ can be computed by the EM algorithm (Dempster, Laird, and Rubin 1977), treating the compliance indicators for subjects in the control group and the unobserved components of y_i as missing data. Likelihood ratio tests can be carried out to test the null hypotheses that particular components of θ are zero, and standard errors of the ML estimates can be computed using the SEM algorithm (Meng and Rubin 1991) or by numerical methods such as the bootstrap. Inference about the CACE and other regression parameters can then be based on the standard large-sample normal approximation.

An alternative approach to inference for the CACE is to assume a prior distribution for the parameters and simulate the posterior distribution using the Gibbs' sampler (Imbens and Rubin 1997a; Tanner 1996), again treating the unobserved compliance indicators and missing outcomes as missing data. This Bayesian approach draws θ from their complete-data posterior distribution and draws the missing outcomes $y_{\text{miss},i}$ from their predictive distribution conditional on $y_{\text{obs},i}$ and θ. Computational details for both of these approaches are provided in Appendixes A and B.

4. APPLICATION TO JOBS II

The model described in the previous section was fitted to the JOBS II data. For the purpose of comparison, an IT analysis, as well as an IV estimation were also performed. Due to the oversampling of the high-risk respondents and the prior study showing beneficial effects for the high-risk group but not for the low-risk group, analyses were performed separately for the low-risk and high-risk groups. The groups were defined by value of a risk score computed from financial strain score, assertiveness score, and depression score at screening (T0); subjects with a risk score greater or equal to 1.38 were assigned to the high-risk group, and were assigned to the low-risk group if it was otherwise. The outcome of interest is the difference in depression scores between Tt and baseline T0, for $t = 2, 3, 4$. A lower depression score indicates improvement, so a negative difference between depression scores at Tt and T0 is a positive outcome. Subjects with missing outcomes at all time points are not included in the analyses; this excludes about 5% and 6% of the high- and low-risk respondents, respectively from the analyses.

The model of (3) and (4) were fitted separately to high-risk and low-risk data sets, with $y_i = (y_{i2}, y_{i3}, y_{i4})^{\mathrm{T}}$, where y_{it} is the difference in depression between Tt and baseline T0, for $t = 2, 3, 4$. Covariates were chosen into model (3) if they were significant predictors for the IT model at the .05 significance level. Covariates measured at baseline that significantly predict compliance were selected based on a logistic regression model on the compliance information available for the experimental group. Age, education (highest school grade completed), motivation to attend, assertiveness at baseline, marital status (unmarried or married), economic hardship at baseline, and race (nonwhite or white) are significant in predicting compliance for the high-risk respondents. For low-risk respondents, age, education, motivation to attend, and income level are significant predictors of compliance. Reduced models with $\beta_{CX,t} = \beta_{CRX,t} = 0$ in (3) were also fitted to the data. The ML estimates and Bayesian posterior distribution with uniform priors on all the parameters were computed, using the

Table 3. High-Risk Group: Parameter Estimates (SEs) From Longitudinal Model

Parameter	Gibbs*	EM	IV model	IT model
$\beta_{0,T2}$	.86 (.25)	.87 (.29)	.85 (.28)	. 83 (.28)
$\beta_{0,T3}$	1.69 (.26)	1.71 (.27)	1.60 (.28)	1.58 (.28)
$\beta_{0,T4}$	1.36 (.28)	1.39 (.31)	1.43 (.31)	1.43 (.31)
$\beta_{C,T2}$	.10 (.15)	.02 (.07)	–	–
$\beta_{C,T3}$	.21 (.13)	.09 (.06)	–	–
$\beta_{C,T4}$	.21 (.12)	.09 (.09)	–	–
$\beta_{CR,T2}$	−.25 (.13)	−.19 (.08)	−.14 (.07)	–
$\beta_{CR,T3}$	−.34 (.11)	−.25 (.08)	−.18 (.08)	–
$\beta_{CR,T4}$	−.17 (.11)	−.08 (.11)	−.10 (.08)	–
$\beta_{R,T2}$	–	–	–	−.06 (.07)
$\beta_{R,T3}$	–	–	–	−.12 (.07)
$\beta_{R,T4}$	–	–	–	−.09 (.07)
$\beta_{NS,T2}$	–	–	.06 (.08)	–
$\beta_{NS,T3}$	–	–	−.04 (.08)	–
$\beta_{NS,T4}$	–	–	−.08 (.09)	–
$\beta_{X_1,T2}$ (depression at T0)	−1.06 (.16)	−1.04 (.18)	−.94 (.17)	−.93 (.18)
$\beta_{X_1,T3}$ (depression at T0)	−1.53 (.17)	−1.51 (.19)	−1.36 (.18)	−1.35 (.18)
$\beta_{X_1,T4}$ (depression at T0)	−1.15 (.17)	−1.14 (.19)	−1.05 (.19)	−1.05 (.19)
$\beta_{X_2,T2}$ (risk)	.84 (.25)	.82 (.26)	.68 (.25)	.68 (.26)
$\beta_{X_2,T3}$ (risk)	.97 (.25)	.95 (.26)	.82 (.26)	.82 (.26)
$\beta_{X_2,T4}$ (risk)	.56 (.24)	.54 (.29)	.45 (.28)	.45 (.28)
ρ_{12}	.36 (.04)	.36 (.04)	–	–
ρ_{13}	.21 (.05)	.21 (.03)	–	–
ρ_{23}	.24 (.05)	.25 (.04)	–	–
IVE (T2)	–	–	−.09 (.12)	–
IVE (T3)	–	–	−.21 (.13)	–
IVE (T4)	–	–	−.16 (.14)	–

*Gibbs estimates were based on two sequences with 600 iterations.

procedures detailed in the appendices. Standard errors of the ML estimates were computed based on 100 bootstrap samples drawn from the observed sample with replacement.

The likelihood ratio test indicates that there is no significant lack of fit of the reduced model (p-value of .99 for high-risk group, p-value of .95 for low-risk group); hence, only estimates from the reduced model ($\beta_{CX,t} = \beta_{CRX,t} = 0$ in (3)) are presented. Tables 3 and 4 give the EM and Gibbs parameter estimates from model (3) for the high-risk and low-risk groups, respectively; standard errors of the parameter estimates are shown in parenthesis. The estimates from EM and Gibbs are, in general, similar for both of the risk groups; however, estimates for β_C and β_{CR} from the two methods differ considerably. There are significant time effects ($\beta_{0,t}$) on the change in depression scores in both of the risk groups, with greatest magnitude of effect at T3, six months after intervention, a change of about 1.71 for the high-risk group and about .95 for the low-risk group, respectively. The effect of compliance ($\beta_{C,t}$) for controls, is not significant for both the risk groups, but the CACE ($\beta_{CR,t}$) is significant for the high-risk group at T2 (two months post-intervention) and T3 (six months post-intervention). CACE is greatest at T3 and smallest at T4 (two years after intervention) for both risk groups. Estimates of the correlations of the change in outcomes at the three time points, ρ_{ij}, are also given in Tables 3 and 4. For both of the risk groups, the correlation between T2 and T3 is greatest, whereas the correlation between T2 and T4 is smallest. It is to be expected that outcomes closer to each other in time would be more correlated than outcomes further apart in time. The results imply significant intervention effects for high-risk compliers at two months and six months after the intervention, with greater effects at T3; suggesting that there is intervention effect for high-risk compliers up to six months but negligible effects at two years after the intervention.

The EM and Gibbs parameter estimates for the logistic regression model (4) for compliance are given in Table 5. The model for the high-risk group predicts better compliance for respondents who are white, unmarried, older, more educated, less assertive, have less economic hardship, and more motivation to attend. For low-risk respondents, compliance is predicted to be higher for respondents who are older, more educated, at a higher income level, and have greater motivation to attend. These effects are generally in the expected directions.

These estimates can also be compared with those from IV estimation. An IV estimator in the longitudinal case is obtained by fitting a multivariate version of Bloom's model (1984) using SAS Proc Mixed (SAS 1992), with outcomes y_i and regression parameters now as vectors with three entries. Subjects with missing outcomes at two or less time points were included in the analysis by using the SAS Proc Mixed procedure. The model assumes that the outcomes follow a multivariate normal distribution with means

$$E(y_{it}) = \beta_{0,t} + \beta_{CR,t} C_i R_i + \beta_{NS,t}(1 - C_i) R_i + \beta_{X,t}^{\mathrm{T}} X_{Y_i}, \quad (5)$$

for $t = 2, 3, 4$, and an unstructured covariance matrix, where β_0 is the effect for controls, β_{CR} is the effect for experimental shows, β_{NS} is the effect for experimental no-shows, and β_X is the effect of covariates, $i = 1, \ldots, N$. Covariates included in this model are the same as those in the IT model as well as those in model (3). Likelihood ratio tests for model (5) versus a full version of (5) with interactions

Table 4. Low-Risk Group: Parameter Estimates (SE's) From Longitudinal Model

Parameter	Gibbs*	EM	IV model	IT model
$\beta_{0,T2}$	.83 (.09)	.83 (.10)	.92 (.09)	.92 (.09)
$\beta_{0,T3}$	.95 (.10)	.94 (.09)	.93 (.09)	.93 (.09)
$\beta_{0,T4}$	.92 (.11)	.92 (.10)	.96 (.10)	.96 (.10)
$\beta_{C,T2}$	.03 (.07)	.02 (.04)	–	–
$\beta_{C,T3}$	−.04 (.08)	−.02 (.04)	–	–
$\beta_{C,T4}$	−.03 (.08)	−.01 (.04)	–	–
$\beta_{CR,T2}$	−.05 (.06)	−.04 (.05)	−.06 (.05)	–
$\beta_{CR,T3}$	−.04 (.07)	−.05 (.05)	−.04 (.05)	–
$\beta_{CR,T4}$	.00 (.07)	−.02 (.05)	−.02 (.05)	–
$\beta_{R,T2}$	–	–	–	−.05 (.04)
$\beta_{R,T3}$	–	–	–	−.01 (.04)
$\beta_{R,T4}$	–	–	–	−.01 (.05)
$\beta_{NS,T2}$	–	–	−.04 (.05)	–
$\beta_{NS,T3}$	–	–	.02 (.05)	–
$\beta_{NS,T4}$	–	–	−.00 (.05)	–
$\beta_{X_1,T2}$ (depression at T0)	−.54 (.05)	−.54 (.06)	−.58 (.06)	−.58 (.06)
$\beta_{X_1,T3}$ (depression at T0)	−.63 (.06)	−.63 (.06)	−.63 (.06)	−.63 (.06)
$\beta_{X_1,T4}$ (depression at T0)	−.63 (.07)	−.64 (.07)	−.65 (.06)	−.65 (.06)
$\beta_{X_3,T2}$ (unmarried)	.06 (.04)	.06 (.04)	.06 (.04)	.06 (.04)
$\beta_{X_3,T3}$ (unmarried)	.12 (.04)	.12 (.04)	.11 (.04)	.11 (.04)
$\beta_{X_3,T4}$ (unmarried)	.17 (.05)	.17 (.04)	.15 (.04)	.15 (.04)
ρ_{12}	.35 (.03)	.35 (.05)	–	–
ρ_{13}	.23 (.04)	.23 (.03)	–	–
ρ_{23}	.28 (.04)	.28 (.04)	–	–
IVE (T2)	–	–	−.09 (.07)	–
IVE (T3)	–	–	−.02 (.08)	–
IVE (T4)	–	–	−.03 (.08)	–

*Gibbs estimates were based on two sequences with 300 iterations.

of show or no-show with covariates, indicate no significant lack of fit of model (5) ($p-$value $= .96$ for the high-risk model, $p-$value $= .78$ for the low-risk model). From (2), the IV estimator is $(E(y|T=1)-E(y|T=0))/p_c$, which when computed in terms of the model parameter estimates of (5) is $\hat{\beta}_{CR}+(1-\hat{p}_c)/\hat{p}_c\hat{\beta}_{NS}$, where $\hat{\beta}_{CR}$, $\hat{\beta}_{NS}$ are the estimates of the parameter vectors β_{CR}, β_{NS}, respectively, and $\hat{p}_c$ is an estimate of the average compliance rate of subjects over all the covariates X_{Y_i}. The estimated average compliance rate ($\hat{p}_c$) based on the observed sample is .55 for the high-risk group and .56 for the low-risk group. Standard errors for the IV estimates are again obtained by bootstrap. The parameters estimates of this model as well as the IVE are also shown in Tables 3 and 4.

Finally, a longitudinal IT model was fitted to the outcomes of interest $y_i = (y_{i2}, y_{i3}, y_{i4})$, where y_{it} is the difference in depression scores between Tt and baseline T0, for $t = 2, 3, 4$. The outcomes y_i are assumed to follow a multivariate normal model with means $E(y_{it}) = \beta_{0,t} + \beta_{R,t}R_i + \beta^{\mathrm{T}}_{X,t}X_{Y_i} + \beta^{\mathrm{T}}_{RX,t}R_iX_{Y_i}$, $t = 2, 3, 4$ and an unstructured covariance matrix. This model was fitted using SAS Proc Mixed (SAS 1992), where subjects with missing outcomes at two or less time points were included in the estimation. Covariates measured at baseline that significantly predict change in depression are included in the model. For the high-risk respondents, depression at baseline and the risk score significantly predict the change in depression, whereas depression at baseline and marital status (unmarried or married) significantly predict change in depression for the low-risk respondents. Note that in contrast to model (3), there are no effects of C and CR in the IT model, but instead the main effect of R is present. There is no significant lack of fit of the main effects model with $\beta^{\mathrm{T}}_{RX,t} = 0$ (the likelihood ratio test has a p-value of .98 for the high-risk model and .49 for the low-risk model), hence, only results from the main effects model are presented. The param-

Table 5. ML Estimates (SE's) From Logistic Regression for Compliance

	High-Risk Group		Low-Risk Group	
Parameter	Gibbs	EM	Gibbs	EM
Intercept	−8.76 (1.47)	−8.62 (1.58)	−5.24 (.81)	−5.08 (.97)
Age	.08 (.01)	.07 (.01)	.03 (.01)	.03 (.01)
School grade completed	.33 (.06)	.32 (.07)	.11 (.05)	.11 (.06)
Motivation to attend (2 item)	.67 (.15)	.65 (.16)	.51 (.11)	.49 (.11)
Assertiveness at T0	−.37 (.14)	−.35 (.14)	–	–
Unmarried	.56 (.25)	.59 (.28)	–	–
Economic hardship at T0	−.23 (.14)	−.22 (.14)	–	–
Nonwhite	−.54 (.29)	−.55 (.30)	–	–
Income	–	–	.06 (.02)	.05 (.02)
π_c	.55 (.01)	.55 (.03)	.55 (.01)	.55 (.02)

eter estimates and the corresponding standard errors from the IT main effects models for the high- and low-risk groups are shown in Tables 3 and 4. The IT effects are given by $\beta_{R,t}$, they are not significant at all the time points for both the high-risk and low-risk groups. This result is quite unexpected, because significant intervention effects were expected for high-risk respondents. However, for the high-risk group, there are incremental IT effects at T3 (six months postintervention) when compared to the other time points.

The estimates for the low-risk group are, in general, similar for all the methods, because the estimated effects are small and highly insignificant. For the high-risk group, the CACE estimates (β_{CR}) are significant at T2 and T3, with an increased intervention effect from T2 to T3 and a decrease in the effect at T4 are indicated. Similarly, both the IV estimator and the IT effect estimated an increase in effect from T2 to T3 and a reduction in effect at T4 for the high-risk group; but both the IV estimators and the IT effects are not significant at all the time points. Although there were beneficial effects of the intervention in reducing depression symptomatology for those who complied in the high-risk group, due to the considerable amount of high risk experimental noncompliers who were lost to follow-up at one or more time points (>40%, see Table 2), the magnitude of the estimated IT effect at each time point is generally less than the corresponding IV estimates. Moreover, the CACE estimates indicate greater compliance effects than the corresponding IV estimates for the high-risk group at T2 and T3. Also note that the CACE estimates based on the longitudinal model generally have smaller standard errors than the IV estimators, which is consistent with the theory suggesting that the CACE is more efficient than the IV estimator.

The CACE estimates from the three approaches look quite different for this application. The difference in the estimates of the CACE and IV estimand might be due to bias of the IV estimator caused by missing data due to attrition. Table 6 shows the CACE and IV estimates when univariate analyses on the difference in depression were performed at each of the time points separately. In accordance with the findings in Yau (1997) and Little and Yau (1998), the CACE estimates and the corresponding standard errors for the univarate analyses are, in general, similar between the two approaches. (The IV standard errors seem a bit smaller, but they are asymptotic and do not take into account estimation of the compliance rate). The gain in efficiency in the CACE estimated by EM is not evident in the univariate analysis (see Table 6). This is in agreement with the simulation study results in Yau (1997), which showed that for the univariate case, the CACE estimates and standard errors from EM and Gibbs, and the IV estimate from Bloom are in general quite similar. The univariate analyses were restricted to cases where that outcome was observed, and the difference in the cross-sectional and longitudinal CACE estimates might be in part due to the handling of attrition in the longitudinal case.

Table 6. CACE and IV Estimates Based on Univariate Analysis at Each Time Point

	High-Risk Group		Low-Risk Group	
Time	CACE: EM	IVE	CACE: EM	IVE
T2	−.18 (.15)	−.15 (.13)	−.08 (.09)	−.10 (.07)
T3	−.32 (.14)	−.27 (.12)	−.02 (.08)	−.04 (.08)
T4	−.18 (.18)	−.15 (.14)	.01 (.10)	−.03 (.09)

5. CONCLUSIONS AND DISCUSSION

The methods of Imbens and Rubin (1997a, b) for inference about CACE are extended to handle longitudinal data with noncompliance. When extended to the longitudinal case, in addition to incorporating the noncompliance of subjects in the experimental treatment, subjects with incomplete observations over time can also be included in the analysis, and the covariance structure of the outcomes can be modeled. The longitudinal model is applied to the JOBS II intervention project for unemployed workers, to assess the effect of intervention effect on change in depression scores between the follow-up times Tt and baseline T0, where $t = 2, 3, 4$. The CACE estimated from the model suggests significant intervention effects for the high-risk group at T2 and T3 but not at T4 or for the low-risk group. This result is consistent with the previous finding (Vinokur et al. 1995) that the intervention is beneficial to high-risk respondents but not to low-risk respondents. Our ML and Bayesian estimates are asymptotically optimal under the assumed model, and should yield similar results given large samples. In small or moderate samples, we believe that the Bayesian approach may provide better inferences than ML combined with the bootstrap, particularly when identification problems arise in fitting the model to bootstrapped samples; simulations to assess this issue would be useful. We have also not addressed the important question of robustness to model misspecification. Simulation studies that compared our proposed approach with the IV method on normal and nonnormal populations would be of interest. Our likelihood-based approach could be readily extended to handle nonnormal distributions, such as the multivariate t (Lange, Little, and Taylor 1989).

An attractive feature of the Imbens and Rubin's framework is the increase in efficiency of estimation over the instrumental variable approach. The gain arises from the fact that the ML method exploits the fact that the density functions of the distributions of compliers and noncompliers under the control treatment are positive. This gain is consistent with the finding in the JOBS II application that the standard errors of the EM and Gibbs estimates are generally lower than those of the IV estimators. Imbens and Rubin (1997b) presented a simulation study with low rates of compliance where gains from ML approach are important, using a nonparametric version of the model for scalar outcomes.

Another feature of the Imbens and Rubin approach is that it makes explicit the assumptions necessary for valid estimates of the CACE in the presence of noncompliance. These assumptions tend to be hidden in earlier formulations for causal inference. In particular, the assumptions of SUTVA and the exclusion restriction of treatment assigned given treatment received are usually not obviously satisfied in preventive trials like JOBS II. Methods for addressing and correcting violations of these assumptions are not well developed and provide a promising area for future research. Along these lines, Imbens and Rubin (1997a) proposed a Bayesian sensitivity analysis

to assess the impact of violations of the exclusion restriction. It is also possible to extend the current model to handle different levels of treatment, by assuming at the unit level that given the level of treatment received, the outcome of interest is unrelated to the treatment assignment.

Here, we assumed that the missing-data mechanism is ignorable, but it would be useful to determine sensitivity of results to alternative assumptions about the missing-data mechanism. In particular, Frangakis and Rubin (1999) consider an alternative assumption, called "latent ignorability," which assumes ignorability conditional on compliance status, but allows missingness to depend on the compliance status, which is missing in the control group. This assumption is weaker than ignorability condition, and the model is identified by strengthening the compound exclusion restriction to apply to both outcomes and drop-outs. Modifications of our methods to fit latent ignorable versions of our model are discussed in Peng (2001).

APPENDIX A: MAXIMUM LIKELIHOOD ESTIMATION FOR CACE—EM ALGORITHM

If the compliance indicators and missing outcomes were known for all subjects, then the complete-data log-likelihood for our model would be

$$
\begin{aligned}
&l(\theta|R, C, Y, X_Y, X_C)\\
&= \sum_{i=1}^{N} \ln f(\underline{y}_i|C_i, R_i, T_i, X_{Y_i}, \theta)\\
&\quad + \sum_{i=1}^{N}[(1-C_i)\ln(1-\pi_{ci}) + C_i \ln(\pi_{ci})]\\
&\propto -\frac{1}{2}N\ln|\Sigma| - \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\sum_{i=1}^{N} y_i y_i^{\mathrm{T}}\right)\\
&\quad + \mathrm{tr}\left(\Sigma^{-1}\sum_{i=1}^{N}[C_i R_i \mu_{c1,i} + C_i(1-R_i)\mu_{c0,i} + (1-C_i)\mu_{n,i}]y_i^{\mathrm{T}}\right)\\
&\quad - \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\sum_{i=1}^{N}[C_i(1-R_i)\mu_{c0,i}\mu_{c0,i}^{\mathrm{T}} + C_i R_i \mu_{c1,i}\mu_{c1,i}^{\mathrm{T}}\right.\\
&\quad \left. + (1-C_i)\mu_{n,i}\mu_{n,i}^{\mathrm{T}}]\right) + \sum_{i=1}^{N}[(1-C_i)\ln(1-\pi_{ci}) + C_i\ln(\pi_{ci})].
\end{aligned}
$$

This complete-data log-likelihood belongs to the exponential family, with complete-data sufficient statistics $y_i y_i^{\mathrm{T}}$, $C_i y_i^{\mathrm{T}}$, and C_i, for $i = 1, \ldots, N$. Hence, ML estimates for $\theta = (\alpha, \beta, \Sigma)$ can then be obtained via the EM algorithm for exponential families (Sundberg 1974; Little and Rubin 1987). The E-step computes expected values of the complete-data sufficient statistics given the current parameter estimates and the observed data. The M-step computes new parameter estimates by the complete-data ML estimates, replacing the complete-data sufficient statistics with estimates from the E-step. The E- and M-steps are then repeated until convergence. For this application, there are two levels of E and M steps embedded within each other: the E1 and M1 steps for the logistic regression model (4); and the E2 and M2 steps for the repeated measures model (3). The algorithm starts with the E1-step, iterates between the E2 and M2-steps, then followed by the M1-step. These steps are repeated until convergence. We now describe these steps in more detail.

E1-Step

The E1-step computes the expected value of C_i given observed values y_{obs}, R, T, and current parameter estimates $\theta^{(h)}$ for subjects in the control group:

$$
\begin{aligned}
&E(C_i|R_i = 0, T_i = 0, Y_{\mathrm{obs},i}, X_{C_i}, X_{Y_i}, \theta^{(h)})\\
&\quad = \Pr\{C_i = 1|R_i = 0, T_i = 0, Y_{\mathrm{obs},i}, X_{C_i}, X_{Y_i}, \theta^{(h)}\} = \omega_{xi},
\end{aligned}
$$

where

$$
\omega_{xi} = \frac{\pi_{ci} g_{c0}(Y_{\mathrm{obs},i}|X_{Y_i}, \theta^{(h)})}{\pi_{ci} g_{c0}(Y_{\mathrm{obs},i}|X_{Y_i}, \theta^{(h)}) + (1-\pi_{ci}) g_n(Y_{\mathrm{obs},i}|X_{Y_i}, \theta^{(h)})}, \quad \text{(A.1)}
$$

$$
g_{c0}(Y_{\mathrm{obs},i}|X_{Y_i}, \theta^{(h)}) = (y_{\mathrm{obs},i} - \mu_{c0,\mathrm{obs},i}^{(h)})^{\mathrm{T}} \Sigma_{\mathrm{obs},i}^{(h)^{-1}} (y_{\mathrm{obs},i} - \mu_{c0,\mathrm{obs},i}^{(h)}),
$$

and

$$
g_n(Y_{\mathrm{obs},i}|X_{Y_i}, \theta^{(h)}) = (y_{obs,i} - \mu_{n,\mathrm{obs},i}^{(h)})^{\mathrm{T}} \Sigma_{\mathrm{obs},i}^{(h)^{-1}} (y_{\mathrm{obs},i} - \mu_{n,\mathrm{obs},i}^{(h)}).
$$

E2-Step

The E2-step computes the expected values of the complete-data sufficient statistics $y_i y_i^{\mathrm{T}}$, $C_i y_i^{\mathrm{T}}$, given observed values, ω_{xi} in (A.1) computed in the E1-step, and current parameter estimates $\theta^{(h)}$:

$$
\begin{aligned}
&E(C_i y_{ij}|R_i, T_i, Y_{\mathrm{obs},i}, X_{C_i}, X_{Y_i}, \theta^{(h)})\\
&\quad = \begin{cases} y_{ij} & \text{if } y_{ij} \text{ is observed, } R_i = 1, \text{ and } T_i = 1,\\ \tilde{y}_{c1,ij}^{(h)} & \text{if } y_{ij} \text{ is missing, } R_i = 1, \text{ and } T_i = 1,\\ \omega_{xi}\tilde{y}_{c0,ij}^{(h)} & \text{if } y_{ij} \text{ is missing, } R_i = 0, \text{ and } T_i = 0, \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
&E(1-C_i)y_{ij}|R_i, T_i, Y_{\mathrm{obs},i}, X_{C_i}, X_{Y_i}, \theta^{(h)})\\
&\quad = \begin{cases} y_{ij} & \text{if } y_{ij} \text{ is observed, } R_i = 1, \text{ and } T_i = 0,\\ \tilde{y}_{n,ij}^{(h)} & \text{if } y_{ij} \text{ is missing, } R_i = 1, \text{ and } T_i = 0,\\ (1-\omega_{xi})\tilde{y}_{n,ij}^{(h)} & \text{if } y_{ij} \text{ is missing, } R_i = 0, \text{ and } T_i = 0, \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
&E(y_{ij}y_{ik}|R_i, T_i, Y_{\mathrm{obs},i}, X_{C_i}, X_{Y_i}, \theta^{(h)})\\
&\quad = (1-\omega_{xi})E(y_{ij}y_{ik}|C_i = 0, R_i, T_i, Y_{\mathrm{obs},i}, X_{C_i}, X_{Y_i}, \theta^{(h)})\\
&\qquad + \omega_{xi}E(y_{ij}y_{ik}|C_i = 1, R_i, T_i, Y_{\mathrm{obs},i}, X_{C_i}, X_{Y_i}, \theta^{(h)})\\
&\qquad + \mathrm{Cov}(y_{ij}, y_{ik}|R_i, T_i, Y_{\mathrm{obs},i}, X_{C_i}, X_{Y_i}, \theta^{(h)})\\
&\quad = y_{ij}y_{ik} \quad \text{if both } y_{ij} \text{ and } y_{ik} \text{ are observed.}
\end{aligned}
$$

If $R_i = 1, T_i = 1$ then

$$
= \begin{cases} \tilde{y}_{c1,ij}^{(h)} y_{ik} & \text{if } y_{ij} \text{ is missing, } y_{ik} \text{ is observed,}\\ \tilde{y}_{c1,ik}^{(h)} y_{ij} & \text{if } y_{ij} \text{ is observed, } y_{ik} \text{ is missing,}\\ \sigma_{jk\cdot\mathrm{obs},i} + \tilde{y}_{c1,ij}^{(h)}\tilde{y}_{c1,ik}^{(h)} & \text{if both } y_{ij} \text{ and } y_{ik} \text{ are missing.} \end{cases}
$$

If $R_i = 1, T_i = 0$ then

$$
= \begin{cases} \tilde{y}_{n,ij}^{(h)} y_{ik} & \text{if } y_{ij} \text{ is missing, } y_{ik} \text{ is observed,}\\ \tilde{y}_{n,ik}^{(h)} y_{ij} & \text{if } y_{ij} \text{ is observed, } y_{ik} \text{ is missing,}\\ \sigma_{jk\cdot\mathrm{obs},i} + \tilde{y}_{n,ij}^{(h)}\tilde{y}_{n,ik}^{(h)} & \text{if both } y_{ij} \text{ and } y_{ik} \text{ are missing.} \end{cases}
$$

If $R_i = 0, T_i = 0$ then

$$= \begin{cases} ((1-\omega_{xi})\tilde{y}^{(h)}_{n,ij} + \omega_{xi}\tilde{y}^{(h)}_{c0,ij})y_{ik} \\ \quad \text{if } y_{ij} \text{ is missing, } y_{ik} \text{ is observed,} \\ ((1-\omega_{xi})\tilde{y}^{(h)}_{n,ik} + \omega_{xi}\tilde{y}^{(h)}_{c0,ik})y_{ij} \\ \quad \text{if } y_{ij} \text{ is observed, } y_{ik} \text{ is missing,} \\ \sigma_{jk\cdot\text{obs},i} + ((1-\omega_{xi})\tilde{y}^{(h)}_{n,ij}\tilde{y}^{(h)}_{n,ik} + \omega_{xi}\tilde{y}^{(h)}_{c0,ij}\tilde{y}^{(h)}_{c0,ik}) \\ \quad \text{if both } y_{ij} \text{ and } y_{ik} \text{ are missing,} \end{cases}$$

where

$$\begin{aligned} \tilde{y}^{(h)}_{c1,ij} &= E(y_{ij}|C_i = 1, R_i = 1, T_i = 1, Y_{\text{obs},i}, X_{C_i}, X_{Y_i}, \theta^{(h)}) \\ &= \mu^{(h)}_{c1,ij} + \Sigma^{(h)}_{j\cdot\text{obs},i}\Sigma^{(h)^{-1}}_{\text{obs},i}(y_{\text{obs},i} - \mu^{(h)}_{c1\cdot\text{obs},i}), \\ \tilde{y}^{(h)}_{c0,ij} &= E(y_{ij}|C_i = 1, R_i = 0, T_i = 0, Y_{\text{obs},i}, X_{C_i}, X_{Y_i}, \theta^{(h)}) \\ &= \mu^{(h)}_{c0,ij} + \Sigma^{(h)}_{j\cdot\text{obs},i}\Sigma^{(h)^{-1}}_{\text{obs},i}(y_{\text{obs},i} - \mu^{(h)}_{c0\cdot\text{obs},i}), \\ \tilde{y}^{(h)}_{n,ij} &= E(y_{ij}|C_i = 0, T_i = 0, Y_{\text{obs},i}, X_{C_i}, X_{Y_i}, \theta^{(h)}) \\ &= \mu^{(h)}_{n,ij} + \Sigma^{(h)}_{j\cdot\text{obs},i}\Sigma^{(h)^{-1}}_{\text{obs},i}(y_{\text{obs},i} - \mu^{(h)}_{n\cdot\text{obs},i}) \end{aligned} \quad \text{(A.2)}$$

are the predicted values of y_{ij} for subject i from model (3) conditional on observed outcomes $y_{\text{obs},i}$ and current values of the parameters $\theta^{(h)}$; $\mu^{(h)}_{c1,ij}, \mu^{(h)}_{c0,ij}, \mu^{(h)}_{n,ij}$, are the j th entry of the estimated mean outcome vector of experimental compliers, control compliers, and never-takers, respectively for subject i based on current values of the parameters $\theta^{(h)}$; $\Sigma^{(h)}_{j\cdot\text{obs},i}$ is the submatrix obtained from the current estimate $\Sigma^{(h)}$, by taking the j th row and column corresponding to observed outcomes for subject i; $\Sigma^{(h)}_{\text{obs},i}$ is the submatrix obtained by taking the rows and columns of $\Sigma^{(h)}$ corresponding to the observed outcomes of subject i; and $\sigma_{jk\cdot\text{obs},i}$ is the jk th element of the conditional covariance of y_{ij} and y_{ik} given $y_{\text{obs},i}$; for $i = 1, \ldots, N$; $j, k = 1, \ldots, \tau$.

M2-Step

The M2-step computes parameter estimates for model (3). This step computes updated estimates of θ using the expected values of the complete-data sufficient statistics computed in the E1-step and the E2-step. Let $\hat{y}^{(h)}_{i|C_i,R_i}$ be the $(\tau \times 1)$ updated vector of outcomes for subject i given C_i and R_i, where

$$\hat{y}^{(h)}_{ij|C_i,R_i} = \begin{cases} y_{ij} & \text{if } y_{ij} \text{ is observed,} \\ \tilde{y}^{(h)}_{n,ij} & \text{if } y_{ij} \text{ is missing and } C_i = 0, \\ \tilde{y}^{(h)}_{c1,ij} & \text{if } y_{ij} \text{ is missing and } C_i = R_i = 1, \\ \tilde{y}^{(h)}_{c0,ij} & \text{if } y_{ij} \text{ is missing and } C_i = 1, R_i = 0. \end{cases}$$

Weights w_i are then assigned to all subjects, with $w_i = 1$ for shows in the experimental group, $w_i = 0$ for no-shows in the experimental group, and $w_i = \omega_{xi}$ for subjects in the control group. Then parameter estimates are given by:

$$\begin{aligned} \hat{\sigma}_{jk} = \frac{1}{N}\sum_{i=1}^{N}\Big[& R_i w_i(\hat{y}^{(h)}_{ij|C_i=1,R_i=1} - \mu^{(h)}_{c1,ij})(\hat{y}^{(h)}_{ik|C_i=1,R_i=1} - \mu^{(h)}_{c1,ik}) \\ & + (1-R_i)w_i(\hat{y}^{(h)}_{ij|C_i=1,R_i=0} - \mu^{(h)}_{c0,ij}) \\ & \times (\hat{y}^{(h)}_{ik|C_i=1,R_i=0} - \mu^{(h)}_{c0,ik}) + (1-w_i)(\hat{y}^{(h)}_{ij|C_i=0,R_i} - \mu^{(h)}_{n,ij}) \\ & \times (\hat{y}^{(h)}_{ik|C_i=0,R_i} - \mu^{(h)}_{n,ik}) + \sigma^{(h)}_{jk\cdot\text{obs},i}\Big] \end{aligned}$$

and

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i^{\mathrm{T}}\Sigma^{(h)^{-1}}X_i\right)^{-1} \times \left(\sum_{i=1}^{N} X_i^{\mathrm{T}}\Sigma^{(h)^{-1}}\left(w_i\hat{y}^{(h)}_{i|C_i=1,R_i} + (1-w_i)\hat{y}^{(h)}_{i|C_i=0,R_i}\right)\right), \quad \text{(A.3)}$$

where σ_{jk} is the jk th element of Σ and X_i is the design matrix of subject i for model (3).

The E2 and M2 steps are iterated until converged ML estimates $\beta^{(h+1)}$ and $\Sigma^{(h+1)}$ are obtained (Jennrich and Schluchter 1986; Little and Rubin 1987). Note that in these iterations, the estimated mean outcomes $\mu^{(h)}_n$, $\mu^{(h)}_{c0}$, $\mu^{(h)}_{c1}$ and the estimated covariance matrix $\Sigma^{(h)}$ in the E2-step are updated with the most current parameter values for β and Σ in the M2-step.

M1-Step

The M1-step computes updated parameter estimates $\alpha^{(h+1)}$ for the logistic regression model (4). Parameter estimates for the logistic model are obtained by weighted logistic regression, with subjects in the experimental group being assigned unit weights, and subjects in the control group being assigned weights equal to ω_{xi} for being a complier ($C_i = 1$) and weights equal to $1 - \omega_{xi}$ for being a never-taker ($C_i = 0$), where values of ω_{xi} are given by (A.1).

Once converged estimates of α, β, and Σ are obtained, the ML parameter estimates of θ are updated by $\theta^{(h+1)} = (\alpha^{(h+1)}, \beta^{(h+1)}, \Sigma^{(h+1)})$, where $\alpha^{(h+1)}$ is from the M1-step, $\beta^{(h+1)}$ and $\Sigma^{(h+1)}$ are from the M2-step. The algorithm then reiterates between the E1, E2, M2, and M1 steps until convergence.

APPENDIX B: BAYESIAN ESTIMATION FOR CACE—GIBBS SAMPLER

Step 1

The first step of the Bayesian approach draws values of θ from the complete-data posterior distribution given current $C_i^{(h)}$ values and observed values:

1. Draw the logistic regression parameters α from the complete-data posterior distribution (using method of Gilks and Wild 1992) given current values $C_i^{(h)}$.
2. Draw missing outcomes $y_{\text{miss},i}$ from its predictive distribution given current parameter values $\theta^{(h)}$, $C_i^{(h)}$, and observed values $y_{\text{obs},i}$, that is, from a multivariate normal distribution with mean given by (A.2) and covariance matrix $\Sigma^{(h)}_{\text{miss}\cdot\text{obs},i}$, obtained by sweeping the current estimate $\Sigma^{(h)}$ by columns corresponding to the observed outcomes for subject i.
3. Draw β and Σ from their corresponding posterior distributions given current $C_i^{(h)}$, drawn values of $\alpha^{(h+1)}$, $y^{(h+1)}_{\text{miss},i}$, and observed values $y_{\text{obs},i}$. With flat priors, this involves drawing $\beta^{(h+1)}$ from a multivariate normal distribution with mean $\tilde{\beta}(y_{\text{obs},i}, y^{(h+1)}_{\text{miss},i})$ and covariance matrix $(\sum_{i=1}^{N} X_i^{\mathrm{T}}\tilde{\Sigma}(y_{\text{obs},i}, y^{(h+1)}_{\text{miss},i})^{-1}X_i)^{-1}$, and $\Sigma^{(h+1)}$ from an inverse Wishart distribution (Odell and Feiveson 1966) with parameters $\tilde{\Sigma}(y_{\text{obs},i}, y^{(h+1)}_{\text{miss},i})$ and degrees of freedom $N - q$ (where q is the number of parameters in the longitudinal model (3)). Note that X_i is the design matrix for model (3), whereas $\tilde{\beta}(y_{\text{obs},i}, y^{(h+1)}_{\text{miss},i})$ and $\tilde{\Sigma}(y_{\text{obs},i}, y^{(h+1)}_{\text{miss},i})$ are the current estimates of the model given observed values $y_{\text{obs},i}$ and current drawn values $y^{(h+1)}_{\text{miss},i}$.

Step 2

The second step draws the unknown values of C_i from its conditional distribution given observed values and current drawn values of θ, this involves draws from a binomial distribution, such that $C_i = 1$ with probability ω_{xi} and $C_i = 0$ with probability $1 - \omega_{xi}$, where ω_{xi} is given by (A.1) evaluated at the current parameter values $\theta^{(h+1)}$.

Flat priors are used for all the parameters. In particular, uniform distributions from -200 to 200 are used for the logistic regression parameters and the longitudinal regression parameters. This range essentially covers all plausible values of the model parameters. For Σ, the prior distribution used is proportional to $|\Sigma|^{(\tau+1)/2}$, where τ is the length of the outcome vector of interest ($\tau = 3$ for the JOBS II application).

Two independent sequences of $2n$ sets of draws were generated for each model, with the first n iterations of each sequence discarded to eliminate the effect of starting values, and model inference was based on the remaining $2n$ iterations from the two sequences. Standard errors of parameter estimates are given by the standard deviations of the parameters in the combined sequence. The method of Gelman and Rubin (1992) was used to monitor convergence of the sequence.

[Received 16 January 1998. Revised 2 August 2001.]

REFERENCES

Angrist, J., Imbens, G. W., and Rubin, D. B. (1996), "Identification of Causal Effects using Instrumental Variables" (with Discussion), *Journal of the American Statistical Association*, 91, 444–472.

Baker, S. (1998), "Analysis of Survival Data From a Randomized Trial With All-or-None Compliance: Estimating the Cost-Effectiveness of a Cancer Screening Program," *Journal of the American Statistical Association*, 93, 929–934.

Balke, A. and Pearl, J. (1997), "Bounds on Treatment Effects From Studies With Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171–1176.

Blinder, A. S. (1987), *Head Heads and Soft Hearts: Tough Minded Economics for a Just Society*, Reading, MA: Addison-Wesley.

Bloom, H. S. (1984), "Accounting for No-Shows in Experimental Evaluation Designs," *Evaluation Review*, 8, 225–246.

Caplan, R. D., Abbey, A., Abramis, D. J., Andrews, F. M., Conway, T. L., and French, J. R. P. (1984), "Tranquilizer Use and Well-Being: A Longitudinal Study of Social and Psychological Effects," Technical Report Series, University of Michigan, Institute for Social Research.

Caplan, R. D., Vinokur, A. D., Price, R. H., and van Ryn, M. (1989), "Job Seeking, Reemployment, and Mental Health: A Randomized Field Experiment in Coping with Job Loss," *Journal of Applied Psychology*, 74, 759–769.

Catalano, R. (1991), "The Health Effects of Economic Insecurity," *American Journal of Public Health*, 81, 1148–1152.

Catalano, R., and Dooley, D. (1977), "Economic Predictor of Depressed Mood and Stressful Life Events in a Metropolitan Community," *Journal of Health and Social Behavior*, 18, 292–307.

Cobb, S., and Kasl, S. V. (1977), "Termination: The Consequences of Job Loss," NIOSH Research Report, DHEW Publication No. 77-224, Washington, DC: U.S. Government Printing Office.

Derogatis, L. R., Lipman, R. S., Rickles, K., Uhlenhuth, E. H., and Covi, L. (1974), "The Hopkins Symptom Checklist (HSCL)," in *Psychological Measurements in Psychopharmacology: Modern Problems in Pharmacopsychiatry* (Vol. 7), ed. P. Pichot, New York: Karger, pp. 79–110.

Dempster, A., 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*, Ser. B, 39, 1–38.

Dew, M. A., Bromet, E. J., and Schulberg, H. C. (1987), "A Comparative Analysis of Two Community Stressors Long Term, Mental Health Effects," *American Journal of Community Psychology*, 15, 167–184.

Frangakis, C. E., and Rubin, D. B. (1999), "Addressing Complications of Intention-to-Treat Analysis in the Combined Presence of All-or-None Treatment-Noncompliance and Subsequent Missing Outcomes," *Biometrika*, 86, 366–379.

Gelman, A., and Rubin, D. B. (1992), "Inference from Iterative Simulations Using Multiple Sequences," *Statistical Science*, 7, 457–511.

Gilks, W. R., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337–348.

Gurin, P., Gurin, G., and Morrison, B. M. (1978), "Personal and Ideological Aspects of Internal and External Control," *Social Psychology*, 41, 275–296.

Hamilton, V. L., Hoffman, W. S., Broman, C.L., and Rauma, D. (1993), "Unemployment, Distress, and Coping: A Panel Study of Autoworkers," *Journal of Personality and Social Psychology*, 65, 234–247.

Imbens, G. W., and Angrist, J. (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," *The Annals of Statistics*, 25, 305–327.

——— (1997b), "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *The Review of Economic Studies*, 64, 555–574.

Jennrich, R. I., and Schluchter, M. D. (1986), "Unbalanced Repeated-Measures Models with Structured Covariance Matrices," *Biometrics*, 42, 805–820.

Justice, B., and Duncan, D. F. (1977), "Child Abuse as a Work-Related Problem," *Corrective and Social Psychiatry, and Journal of Behavior Technology. Methods and Therapy*, 23, 53–55.

Kessler, R. C., Turner, J. B., and House, J. S. (1988), "The Effects of Unemployment on Health in a Community Survey: Main, Modifying, and Mediating Effects," *Journal of Social Issues*, 44, 69–85.

Kessler, R. C. Turner, J. B., and House, J. S. (1989), "Unemployment, Reemployment, and Emotional Functioning in a Community Sample," *American Sociological Review*, 54, 648–657.

Lange, K., Little, R. J. A., and Taylor, J. M. G. (1989), "Robust Statistical Inference Using the t Distribution," *Journal of the American Statistical Association*, 84, 881–896.

Lewis, H. G. (1963), *Unionism and Relative Wages in the United States; An Empirical Inquiry*, Chicago, IL: University of Chicago Press.

Little, R. J. A. (1995), "Modeling the Drop-out Mechanism in Repeated-Measures Studies," *Journal of the American Statistical Association*, 90, 1112–1121.

Little, R. J. A., and Rubin, D. B. (1987), *Statistical Analysis with Missing Data*, New York: Wiley

Little, R. J. A., and Schluchter, M. D. (1985), "Maximum Likelihood Estimation for Mixed Continuous and Categorical Data with Missing Values," *Biometrika*, 72, 497–512.

Little, R. J. A., and Yau, L. H. Y. (1998), "Statistical Techniques for Analyzing Data from Prevention Trials: Treatment of No-Shows Using Rubin's Causal Model," *Psychological Methods*, 3, 147–159.

Meng, X., and Rubin, D. B. (1991), "Using EM to Obtain Asymptotic Variance-Covariance Matrices: The SEM Algorithm," *Journal of American Statistical Association*, 86, 899–909.

Odell, P. L., and Feiveson, A. H. (1966), "A Numerical Procedure to Generate a Sample Covariance Matrix," *Journal of the American Statistical Association*, 61, 199–203.

Olkin, I., and Tate, R. F. (1961), "Multivariate Correlation Models with Mixed Discrete and Continuous Variables," *Annuals of Mathematical Statistics*, 32, 448–465.

Peng, Y. (2001), "Causal Inference for Data Subject to Noncompliance and Missing Values," Ph.D. dissertation, University of Michigan, School of Public Health, Dept. of Biostatistics.

Price, R. H., van Ryn, M., and Vinokur, A. D. (1992), "Impact of Preventive Job Search Intervention on the Likelihood of Depression Among the Unemployed," *Journal of Health and Social Behavior*, 33, 158–167.

Robins, J. M., and Finkelstein, D. M. (2000), "Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests," *Biometrics*, 56, 779–788.

Rotnitzky, A., Scharfstein, D., Su, T., and Robins, J. (2001), "Methods for Conducting Sensitivity Analysis of Trials with Potentially Nonignorable Competing Causes of Censoring," *Biometrics*, 57, 103–113.

Rosenberg, M. (1965), *Society and the Adolescent Self-Image*, Princeton, NJ: Princeton University Press.

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592.

——— (1978), "Bayesian Inference for Causal Effects: the Role of Randomization," *Annuals of Statistics*, 6, 34–58

SAS Institute, Inc. (1992), "The Mixed Procedure," in *SAS/STAT Software: Changes and Enhancements, Release 6.07*, chapter 16, Technical Report P-229, Cary, NC: Author.

Sundberg, R. (1974), "Maximum Likelihood Theory for Incomplete Data from an Exponential Family," *Scandinavian Journal of Statistics*, 1, 49–58.

Tanner, M. A. (1996), *Tools for Statistical Inference* (3rd ed.), New York: Springer-Verlag.

Vinokur, A. D., Caplan, R. D., and Williams, C. C. (1987), "Effects of Recent and Past Stress on Mental Health: Coping with Unemployment among Vietnam Veterans and Non-Veterans," *Journal of Applied Social Psychology*, 17, 708–728.

Vinokur, A. D., Price, R. H., and Schul, Y. (1995), "Impact of the JOBS Intervention on Unemployed Workers Varying in Risk for Depression," *American Journal of Community Psychology*, 23, 39–74.

Yau, H. Y. L. (1997), "Statistical Analysis of Longitudinal Data with Drop-Outs," Ph.D. dissertation, The University of Michigan, School of Public Health, Dept. of Biostatistics.