STATISTICS IN MEDICINE
Statist. Med. 2003; **22**:1285–1303 (DOI: 10.1002/sim.1380)

Weighting in instrumental variables and G-estimation

Marshall M. Joffe[1,*,†] and Colleen Brensinger[2]

[1]*Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Room 602 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, U.S.A.*
[2]*Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Room 618 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, U.S.A.*

SUMMARY

We propose here a simple scheme to use information on compliance and prerandomization covariates to improve analysis of randomized trials with non-compliance. We use the data to determine the effect of randomization on treatment received among various strata defined by pretreatment covariates. When the effect of treatment received on the outcome of interest is the same across strata and pretreatment covariates predict non-compliance, weighting the estimating functions by the effect of randomization on treatment received can improve the precision of explanatory estimates of treatment effect and can increase the power of intent-to-treat tests of the null hypothesis. Efficiency gains under the weighting scheme are a simple increasing function of the variability of these weights. Such weighting schemes will often lead to improvements even when these conditions are not met. We use a randomized trial of cholestyramine to illustrate these points. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: causal inference; noncompliance; instrumental variables; randomized trials; weighting

1. INTRODUCTION

Non-compliance is a pervasive problem in randomized trials. Until recently, analysts have chosen between two methods for analysing such data: the widely accepted and seemingly conservative intent-to-treat (ITT) methods, and the widely and justly maligned ‘as treated’ or ‘per protocol’ analyses and their variants [1].

Recent years have seen the development of alternatives which seek to improve on these familiar approaches; instrumental variables methods, long popular in econometrics, provide the basis for several recent proposals [2–7]. These methods use information on treatment received, but in a seemingly inefficient way, by seeing whether a postulated effect of treatment among the subjects receiving treatment can explain differences between the randomized groups [8]. As

* Correspondence to: Marshall M. Joffe, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Room 602 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, U.S.A.
† E-mail: mjoffe@cceb.upenn.edu

Contract/grant sponsor: United States National Heart, Lung and Blood Institute; contract/grant number: R29 HL59184.

Received February 2001
Accepted July 2002

typically discussed and applied, compliant and non-compliant subjects are weighted equally in the 'adjusted' comparisons of the randomized groups. It seems intuitively obvious that compliant subjects provide more information about the effect of treatment than non-compliant ones and so should receive more weight in the analysis.

This paper presents a simple weighting scheme for instrumental-variables type estimators based on this notion and shows how it can be used to improve efficiency. We extend the scheme to estimators not considered previously [9–12] and consider when such schemes can confer substantial benefits. We illustrate some of these points with data from a randomized trial of cholestyramine [13, 14].

2. CHARACTERIZING NON-COMPLIANCE AND ITS EFFECTS

To motivate our discussion, consider a study of the effect of cholestyramine (CLT) on blood cholesterol levels measured at the end of a fixed follow-up period. Choosing a continuous outcome measured at the end of a fixed follow-up period makes the ideas and calculations most transparent; similar approaches can be taken with failure time and repeated measures outcomes. Let R denote the intervention, assigned by randomization: $R=1$ for subjects randomized to CLT, 0 for others. Let Y denote the outcome of interest (log serum cholesterol at the end of follow-up), and let X represent covariates or attributes measured prior to randomization.

Sometimes compliance refers to a single act or event; other times, as in trials of ongoing drug therapies, it refers to a series of events (for example, whether a subject takes a particular drug on any given day). In the former case, compliance is naturally a scalar quantity. In the latter case, it has often been treated as a scalar quantity not varying over time. For most of our exposition, we will adopt this approach, as it simplifies mathematics and understanding, and yields a tractable analysis that can be performed with standard software. Let A denote the level of consistency of one's observed treatment/exposure pattern with the course of action prescribed by the intervention; here, $A=1$ if a subject takes the level of CLT prescribed to subjects randomized to receive it, and $A=0$ if the subject takes no CLT. More generally, A can take any value between 0 and 1, expressing the overall level of compliance with the regime (for example, proportion of prescribed doses taken for a medication). In order for this representation to be adequate, we must either assume that compliance (for example, CLT use) does not vary over time, or that treatment effects depend only on the average level it takes over the interval.

Consider next defining and describing the effects of CLT use (or, loosely speaking, of compliance). To fix ideas, consider a strong-minded subject who is motivated to take all the prescribed doses and wants to know what effect CLT use will have on his cholesterol level. In other words, this individual wants to compare what would happen to his cholesterol level if he used CLT with what would happen if he did not. This is a comparison of potential outcomes: Y_1, the log serum cholesterol that would be observed at the end of one year were he to take CLT as prescribed, and Y_0, the log cholesterol that would be observed if he does not take CLT; we sometimes call Y_0 the baseline potential outcome. More generally, Y_a is the outcome that would be seen if an individual received level a of CLT; denote the vector of potential outcomes Y_a by $\mathbf{Y}$. At most one of these potential outcomes can be observed. Implicit in our notation is the assumption that randomization has no effect on outcome except that mediated by the dose of CLT received; that is, randomization has no direct effect. An investigator should assess the validity or plausibility of this 'exclusion restriction' [2] in applications [15, 16].

If all subjects complied with their assigned treatment (that is, all subjects assigned to take CLT did so and no others did), a standard ITT analysis could help answer this person's question. A comparison of the mean responses in the randomized groups (that is, $E(Y \mid R=1)-E(Y \mid R=0)$) equals in expectation a comparison of the potential outcomes $E(Y_1) - E(Y_0)$. Under the (non-identifiable) assumption that the effect $Y_1 - Y_0$ is the same in each individual, the ITT estimates from the randomized trial estimate what this motivated subject wants to know.

ITT comparisons are standard even in the presence of non-compliance [1, 17], where they are sometimes considered a *pragmatic* analysis or an analysis of the *effectiveness* of the intervention [18]. The approach compares outcomes under two different plans: one which encourages and provides CLT, and one which does not. This analysis only partially answers the question of interest to this motivated subject; the study shows that CLT is likely to lower serum cholesterol (further assumptions are required even to reach this conclusion [2, 19]), but does not estimate how much.

Analyses which try to answer the question of interest to this subject and so estimate the efficacy of CLT typically compare people based on the levels of treatment they receive, possibly adjusting for pretreatment covariates X. Two variants of this approach include: ignoring randomization status and comparing subjects who received treatment with those who did not ('as treated'); comparing subjects who complied with their assigned regime (that is, subjects assigned to CLT who used it with subjects assigned to placebo, 'per protocol') [1]. An assumption that would justify such analysis is ignorable non-compliance (see Rosenbaum and Rubin [20]), expressed mathematically as $f(A \mid R, X, \mathbf{Y}) = f(A \mid R, X)$; that is, any treatment-arm specific association between the level of treatment received and the potential outcomes $\mathbf{Y}$ may be explained in terms of measured covariates X. Such assumptions make no use of the fact of randomization, and analyses based on them justly provoke scepticism [1].

3. INSTRUMENTAL VARIABLES AS EXPLANATORY ANALYSIS

Instrumental variables analyses provide an alternative for answering the motivated subject's question but do not ignore the fact of randomization. This section outlines one such approach, using G-estimation (an instrumental variables method) for estimating parameters in structural nested distribution and structural nested mean models [6, 11, 19, 21], and shows how it often can provide better-justified explanatory analyses than those sketched above.

Consider simple models for the effect of treatment received, that is, we are interested in comparing observed outcomes Y to the outcome Y_0 that would have been seen had treatment not been provided. Let $\mu_a(W) \equiv E(Y_a \mid W)$ denote the mean of the potential outcome Y_a in the subset of the data defined by observable variables W. A structural mean model (SMM) relates the observable mean outcome $\mu_A(X, R, A)$ in the subset of the population defined by treatment received A, covariates X, and randomization group R to the mean $\mu_0(X, R, A)$ of the baseline potential outcomes Y_0 in the same subset. A simple, one-parameter SMM is

$$\mu_0(X, R, A) = \mu_A(X, R, A) - \Psi_0 A \tag{1}$$

where Ψ_0 parameterizes the effect of treatment received. Under model (1), the realized effect of treatment received is proportional to the amount of treatment received, that is, subjects

who received half of their prescribed dose of CLT received half the cholesterol-lowering effect (on the log scale) of subjects who received the full dose. We consider more flexible, multi-parameter SMMs in Section 5. Let $\varepsilon_{X,R,A;a} \equiv Y_a - \mu_a(X,R,A)$ denote the residual from the regression of the potential outcome Y_a on its mean, and let $\boldsymbol{\varepsilon}_{X,R,A} \equiv \{\varepsilon_{X,R,A;a}\}$ denote the set of potential residuals. Structural distribution models (SDM) place restrictions on the distribution of the potential residuals $\boldsymbol{\varepsilon}$. One possible restriction is that the residuals of the baseline and observable outcomes $\varepsilon_{X,R,A;0}$ and $\varepsilon_{X,R,A;A}$ are identically distributed; that is, that

$$F_{\varepsilon_{X,R,A;0}|X,R,A}(e) = F_{\varepsilon_{X,R,A;A}|X,R,A}(e) \tag{2}$$

If the residuals $\varepsilon_{X,R,A;0}$ and $\varepsilon_{X,R,A;A}$ are identical, models (1)–(2) are deterministic or rank-preserving (RP). A more flexible SDM may allow the distribution of the residuals to vary with the treatment received A, as well as randomization group R and pretreatment covariates X.

Randomization allows consistent estimation of Ψ_0 even when compliance is not ignorable. Let $p \equiv \text{pr}(R=1)$ be the randomization probability, and let $g(\cdot)$ be a known function of its arguments. Let Ψ denote a guess or putative value for the causal parameter Ψ_0, and let $U(\Psi) \equiv Y - A\Psi$. Informally, we call $U(\Psi)$ the putative potential outcome; under a RP SDM, when $\Psi = \Psi_0, U(\Psi) = Y_0$. For SDMs, randomization implies that the $U(\Psi_0)$ and pretreatment covariates X are jointly independent of the initial assignment R. Let $g(u,x)$ be a known, arbitrary function of its arguments. The joint independence implies that $g\{U(\Psi_0), X\}$ is independent of R, and so that

$$E\left[\sum_i (R - p) g\{U(\Psi_0), X\}\right] = 0 \tag{3}$$

For SMMs, $U(\Psi_0)$ is uncorrelated but not necessarily independent of R, thus (3) will hold if $g(u,x) = g^0(x) + g^1(x)u$ is a linear function of u. Replacing the true but unknown Ψ_0 in (3) by a putative value Ψ leads to estimating equations for Ψ_0.

The structural model and the putative potential outcome $U(\Psi)$ both involve treatment received A, and so our estimation procedure (3) uses information on post-randomization treatment to estimate effects. This is acceptable in principle; heuristically, A is used with the structural model (in SDMs) only to remove the effect of treatment and so recover a pretreatment variable $U(\Psi_0)$ uncorrelated with the randomization indicator.

These estimating equations generalize an instrumental variables estimator common in the econometrics literature [2]. To see this, consider using the function $g(u,x) = u$ in (3), which leads to $\sum (R - p)Y = \sum (R - p)A\Psi$; solving for yields Ψ yields $\hat{\Psi} \equiv \{\sum (R - p)Y\}/\{\sum (R - p)A\}$, which is essentially the usual instrumental variables estimator.

4. GIVING MORE WEIGHT TO COMPLIERS

It seems intuitively obvious that compliant subjects should provide more information about the effect of treatment than non-compliant ones. Although, even for binary compliance, we often cannot definitively identify compliant individuals [2, 22], we can identify groups defined by prerandomization covariates X that are less compliant or more so; this identification can improve efficiency. Define $\delta(X) \equiv E(A \mid R=1, X) - E(A \mid R=0, X)$; $\delta(X)$ is a measure of

the effect of randomization on treatment received and has been called the compliance score [23, 24]. If treatment received is all or none (that is $A=0$ or 1) and there are no 'defiers' (people who would take the experimental treatment if and only if assigned not to receive it [2]), $\delta(X)$ is the proportion of compliers among subjects with pretreatment covariates X. One might be tempted to give more weight to strata X in which $\delta(X)$ is large.

Formalizing the weighting scheme and its benefits requires additional structure for the estimating functions. Consider restrictions on the form of the function $g(u,x)$ in (3). We can decompose any arbitrary function $g(u,x)$ into a product of a weight function $w(x)$, dependent only on pretreatment covariates x, and another function $h(u,x)$, that is, $g(u,x)\equiv w(x)h(u,x)$. Define $\varepsilon_{X;0}(\Psi)\equiv Y-\mu_A(X)-A\Psi=U(\Psi)-\mu_0(X)$; for RP SDMs, $\varepsilon_{X;0}(\Psi_0)$ is the baseline residual $\varepsilon_{X;0}$. One set of choices for $h(u,x)$ is the set of functions $q(\cdot)$ of the residual $u-\mu_0(x)$, that is, $h(u,x)=q\{\varepsilon_{X;0}(\Psi)\}$.

The following theorems clarify the benefits of a scheme that weights a function of the residuals $q\{\varepsilon_{X;0}(\Psi)\}$ by the effect of randomization on treatment received $\delta(x)$:

> *Theorem* 1. Suppose that the semi-parametric SDM (1)–(2) holds, and that compliance is ignorable. Consider estimating equations of form (3). Then
> 1. The optimal function $g(u,x)$ is $\delta(x)q^*(\varepsilon_{X;0})$, where $q^*(\varepsilon)\equiv\partial\ln\{f(\varepsilon)\}/\partial\varepsilon$, and $f(\varepsilon)$ is the density of the error function $\varepsilon_{X;0}$.
> 2. For any given function $m(\varepsilon)$ of the residual $\varepsilon_{X;0}$, the most efficient weighting scheme is to choose weights $w(x)=\delta(x)$.

Part 1 of the theorem follows directly from Robins [21]. A sketch of the proof of part 2 appears in the Appendix. The theorem states that the most efficient estimating function $w(x)$ is obtained by combining an optimal function $q^*(\varepsilon)$ of the residuals with compliance-based weights $\delta(x)$. In addition, if one chooses a different function $q(\varepsilon)$ of the residuals, perhaps to improve robustness and decrease sensitivity to large values of ε, compliance-based weights will maximize efficiency for that function $q(\varepsilon)$.

> *Theorem* 2. Suppose that the semi-parametric SMM (1) holds, that compliance is ignorable, that subjects in the placebo group have no access to treatment, and that the residual $\varepsilon_{X;0}(\Psi_0)$ is not associated with covariates X within a given treatment arm R (that is, $f\{\varepsilon_{X,0}(\Psi_0)|X,R\}=f\{\varepsilon_{X,0}(\Psi_0)|R\}$). Consider estimating equations of form (3). The optimal function $g(u,x)$ for estimation in SMMs is $\delta(x)\varepsilon_{X;0}(\Psi_0)$.
>
> Theorem 2 follows directly from Robins [6]. Fischer-Lapp and Goetghebeur [11] also consider the efficiency benefits of using residuals $\varepsilon_{X;0}(\Psi_0)$ in SMMs.

We consider next the efficiency benefits of using compliance-based weights. The simple unweighted estimator proposed elsewhere uses equal weights $w(x)=1$, so we compare compliance-based weights to these. Let η_δ and σ_δ denote the mean and standard deviation of δ, the effect of randomization on treatment received, and let $\Phi_\delta=\sigma_\delta/\eta_\delta$ be the coefficient of variation. Denote by $\hat{\Psi}_{w,q}$ an estimator of Ψ_0 using weights $w(x)$ and transformation $q(\varepsilon)$, and by $\sigma^2_{w,q}$ its asymptotic variance. Theorem 3, proven in the Appendix, quantifies the benefits of compliance-based weights over unweighted estimation in SDMs.

> *Theorem* 3. Under the conditions of theorem 1, the relative efficiency of the weighted and unweighted estimators $\hat{\Psi}_{\delta,q}$ and $\hat{\Psi}_{1,q}$ is $\sigma^2_{1,q}/\sigma^2_{\delta,q}=1+\Phi^2_\delta$.

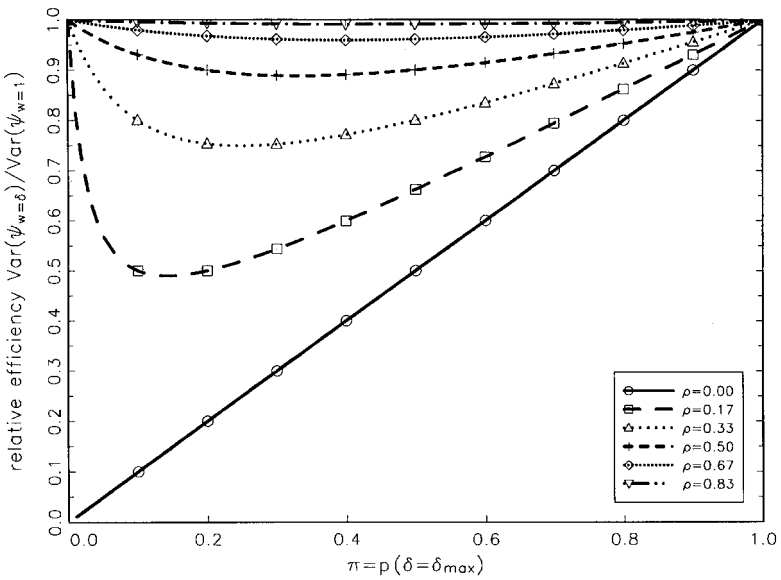

Figure 1. The relative efficiency of the unweighted G-estimator compared to the weighted G-estimator as a function of $\rho \equiv \delta(X=1)/\delta(X=0)$, the ratio of the effects of randomization on treatment received in the better and worse complying strata, for different values of $\pi \equiv \mathrm{pr}(X=1)$, the proportion of the population in the better-complying stratum.

Fischer-Lapp and Goetghebeur [11] have the same findings for the more restricted estimators of SMMs. When the effect of randomization on treatment received $\delta(X)$ is constrained to be non-negative, as one would usually expect in randomized trials, the greatest values of Φ_δ are obtained when a large part of the population has small values of $\delta(X)$ and a small part of the population has relatively large values of $\delta(X)$. Although most trials seek to maximize compliance and so $\delta(X)$, small values of $\delta(X)$ may be common in trials which seek to modify ingrained behaviours (for example, diet or smoking); the MRFIT study is one example. Consider for illustration a binary covariate X; let $\pi \equiv \mathrm{pr}(X=1)$ denote the proportion of the population in the better-complying stratum, and let $\rho \equiv \delta(X=1)/\delta(X=0)$ be the ratio of the effects of randomization on treatment received in the better and worse complying strata. The coefficient of variation Φ_δ is $(\rho-1)\{\pi(1-\pi)\}^{1/2}/\{(\rho-1)\pi+1\}$. Figure 1 plots the relative efficiency against the proportion of subjects π in the better complying stratum for different values of the compliance ratio ρ.

In deriving suggestions for functions $g(\cdot)$ to use in G-estimation, we have assumed ignorable non-compliance, a common distribution of the errors $\varepsilon_{X;A}$ at different values of X, and a correctly specified model for a measure of central tendency $\mu_0(X)$. The efficiency, but not the consistency, of the suggested estimators depends on these assumptions. Under these assumptions (ignorable non-compliance in particular), standard regression estimators which compare groups based on treatment received may be consistent and more efficient than the G-estimators. The primary advantage of G-estimation and other instrumental variables estimators over standard regression methods is that they can provide valid estimates and tests of

treatment effects even when non-compliance is not ignorable. Thus, it is useful to examine the efficiency of the estimating functions discussed above in settings where the derivation does not justify the efficiency of these estimators in general, and under non-ignorable non-compliance in particular.

We performed a simulation study to examine these issues, considering settings where individual compliance is all-or-none. We varied many components of the model used to generate the data: for a scalar binary covariate X, the probability π that $X=1$; for a potential outcome $Y(0)$, we let $\mu_0(X)=X\beta$ and let $\varepsilon_{X;A}=\varepsilon_{X;0}$ be either a standard normal or a standard t (with 5 degrees of freedom) random variable; we assumed that treatment received followed the logistic model $\text{logit}\{pA=1\,|\,X,R=1,Y(0)\}=\alpha_A+X\beta_A+\varepsilon_{X;0}\gamma_A$, and varied all three parameters in the regression; $\gamma_A=0$ implies ignorable non-compliance. For these simulations, we assumed that subjects not assigned to treatment could not obtain treatment outside the study, so $\text{pr}(A=1\,|\,R=0)=0$.

Table I provides some results of the simulations. Notably, the influence of different weighting schemes depended on what function $h(u,x)$ was used in the estimation. When $h(u,x)=w(x)m(\varepsilon_{X;A})$, the weighting by (estimates of) $\delta(X)$ did not decrease and often improved the efficiency of estimators compared with equal weights ($w(X)=1$); this was true even for non-optimal functions $m(\varepsilon)\neq q(\varepsilon)$, as, for example, when $m(\varepsilon)=\varepsilon$ but the error function ε followed a 5 degree-of-freedom t-distribution. However, for other choices of $h(u,x)$ (here, $h(y_0,x)=q(y_0)$), weighting by $\delta(X)$ sometimes reduced efficiency. The efficiency advantages of weighting by $\delta(X)$ appear to extend to settings where non-compliance is not random.

In practical data analysis, neither the appropriate weights $\delta(X)$ nor residuals $\varepsilon_{X;0}(\Psi)$ will be known; none the less, one can estimate them from the data. One can use estimates of expected compliance $\hat{E}(A\,|\,R,X)$ to estimate $\delta(X)$ as $\hat{E}(A\,|\,R=1,X)-\hat{E}(A\,|\,R=0,X)$. Similarly, one can regress $U(\Psi)$ on X to obtain putative means $\mu_0(X)$ and residuals $\varepsilon_{X;0}(\Psi)$. Results from Robins [6] imply that estimation of these regressions from the data does not affect asymptotic efficiency.

To see the potential pitfalls with this in small samples, consider the limit where no two subjects share the same values of covariates X. One might be tempted to replace $\delta(X)$ with the non-parametric estimate $\hat{\delta}(X)=\hat{p}(A=1\,|\,X,R=1)-\hat{p}(A=1\,|\,X,R=0)=A$; in this limit, the approach becomes a 'per protocol' analysis, which can be biased if compliance is not ignorable. Thus, it is important to examine the validity (that is, bias and coverage properties) of the method (using more reasonable estimators of $\delta(X)$) in studies with moderate sample sizes. Similar issues arise in the estimation of $\mu_0(X)$; we estimate $\mu_0(X)$ from the data $\{X,Y(0;\Psi)\}$ for each hypothesized value Ψ for Ψ_0. Figure 2 presents simulation results. For the situations considered, there was little indication of substantial bias or poor coverage. For the settings considered, the mean squared error and the estimated variance increased faster with decreasing sample size when the true $\delta(X)$ was used instead of its estimate, and 95 per cent confidence intervals covered more than at their nominal rate in small samples, especially when the true $\mu_0(X)$ was employed instead of its estimate.

5. THE LRC-CPPT TRIAL

We applied the methods discussed above to the Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT), a randomized placebo-controlled double-blind trial of CLT, a drug

Table I. Root mean squared error of various G-estimators and least-squares estimators.

Data generation process characteristics				Estimators							
Logistic regression parameters for compliance in experimental group			Error function ε; d.f. of t distribution	G-estimators; functions $g\{Y(0;\Psi),X\}$						Least squares (with covariate control)	
Intercept	Baseline covariates	Potential outcome		$Y(0;\Psi)$	$Y(0;\Psi)\hat{\delta}(X)$	$Y(0;\Psi)-\mu(X)$	$\{Y(0;\Psi)-\mu(X)\}\hat{\delta}(X)$	tanh $\{Y(0;\Psi)\}$	$\tanh\{Y(0;\Psi)-\mu(X)\}\hat{\delta}(X)$	Intent-to-treat	As treated
−1.5	0.0	0	∞	0.270	0.269	0.241	0.239	0.290	0.249	0.817	0.075
			5	0.338	0.337	0.315	0.314	0.321	0.282	0.819	0.103
		1	∞	0.179	0.185	0.165	0.158	0.248	0.177	0.727	0.701
			5	0.220	0.221	0.205	0.196	0.287	0.196	0.719	0.987
	1.5	0	∞	0.183	0.181	0.169	0.151	0.203	0.158	0.723	0.068
			5	0.223	0.214	0.211	0.186	0.224	0.168	0.720	0.090
		1	∞	0.133	0.133	0.120	0.104	0.172	0.107	0.635	0.579
			5	0.164	0.158	0.155	0.137	0.208	0.126	0.634	0.824
	3.0	0	∞	0.134	0.119	0.123	0.096	0.160	0.100	0.630	0.059
			5	0.169	0.143	0.161	0.121	0.174	0.108	0.627	0.080
		1	∞	0.116	0.108	0.105	0.084	0.146	0.087	0.577	0.456
			5	0.144	0.130	0.135	0.111	0.158	0.099	0.573	0.685
0.0	0.0	0	∞	0.098	0.098	0.090	0.090	0.104	0.095	0.504	0.054
			5	0.121	0.121	0.113	0.113	0.113	0.101	0.499	0.067
		1	∞	0.090	0.094	0.081	0.080	0.102	0.082	0.444	0.499
			5	0.113	0.115	0.106	0.104	0.115	0.093	0.449	0.692
	1.5	0	∞	0.081	0.086	0.074	0.073	0.086	0.076	0.408	0.049
			5	0.104	0.105	0.098	0.095	0.100	0.085	0.407	0.064
		1	∞	0.082	0.085	0.074	0.071	0.092	0.073	0.388	0.414
			5	0.098	0.100	0.092	0.090	0.097	0.078	0.389	0.592
	3.0	0	∞	0.077	0.081	0.070	0.068	0.084	0.069	0.365	0.051
			5	0.100	0.101	0.093	0.088	0.095	0.076	0.368	0.063
		1	∞	0.074	0.075	0.069	0.065	0.082	0.067	0.361	0.373
			5	0.098	0.099	0.092	0.088	0.095	0.077	0.366	0.532

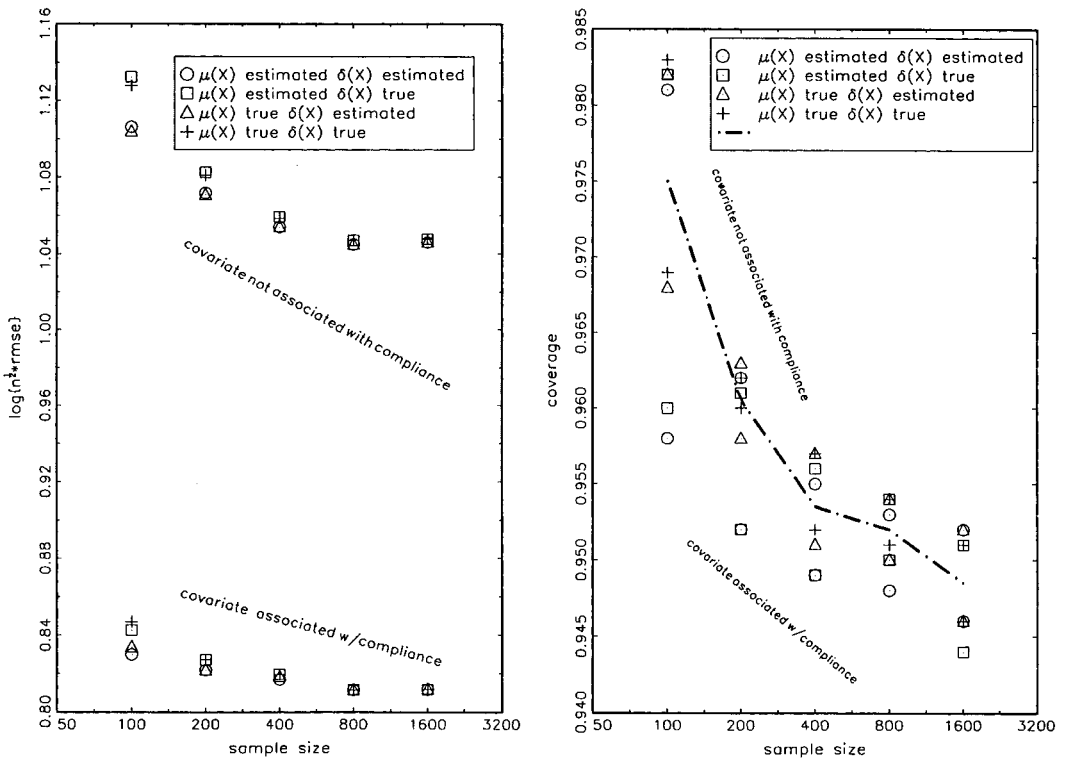

Figure 2. Empirical root-mean squared error and confidence interval coverage from simulations. The left-hand panel plots the empirical root mean squared error (scaled by the square root of the sample size) of different G- estimators of Ψ, and the right-hand panel plots the coverage proportion of confidence intervals from the same estimators, under two sets of circumstances. In each case the estimators differ only in whether they use the true or estimated values of the covariate X-specific effect $\delta(X)$ of randomization on treatment received and the conditional mean of $U(\Psi)$ (or $Y(0)$) given covariates X.

used to lower cholesterol levels and, thereby, the risk of coronary events and death [13, 14]. Efron and Feldman [25] used a subset of this study to examine the effect of compliance; we apply different methods towards the same end.

In the LRC-CPPT, serum cholesterol was measured at several prerandomization and at bimonthly postrandomization clinic visits for the duration of follow-up, which averaged 7.4 years. These clinic visits were also used to dispense packets of their assigned medication; patients returned unused packets at each clinic visit. The proportion of packets used is a measure of compliance.

As above, our endpoint is the serum cholesterol at the end of a fixed follow-up period, here one year (400 days). We use as a scalar measure of the treatment received over the year (A) the proportion of the total potential number of packets (number of days $\times$ 6 packets/day) used; for subjects who were prescribed reduced doses because they could not tolerate the drug, this variable is not identical to per cent compliance, such as used by Efron and Feldman [25].

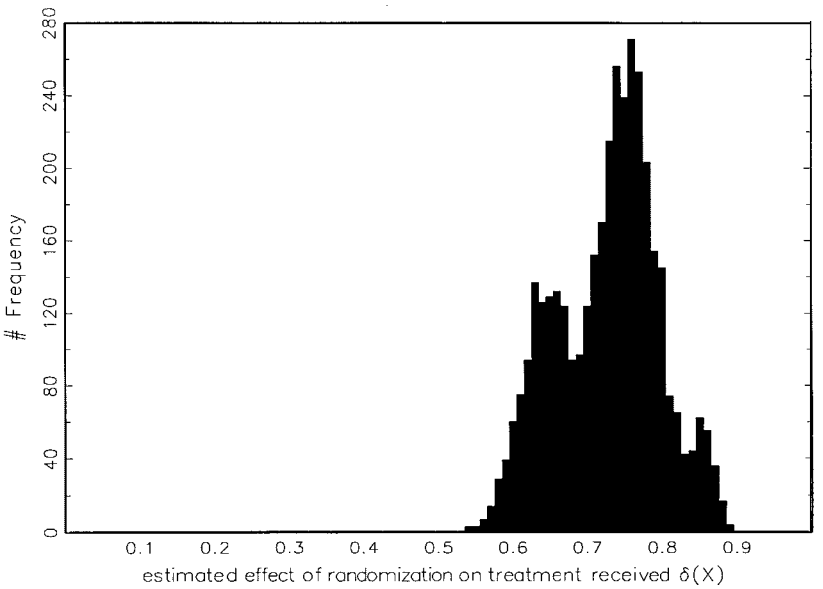

Figure 3. A histogram of the distribution of the model-based values of the covariate-X specific effect $\delta(X)$ of randomization on treatment received from the LRC-CPPT.

We used the continuation ratio model [26] to model compliance with CLT (divided into several ordered categories) in the treatment arm as a function of pretreatment covariates X, which included age, race, sex and centre. We used the model coefficients to compute $\hat{\delta}(X)$, the stratum-specific effect of randomization on CLT received. Figure 3 illustrates the distribution of $\hat{\delta}(X)$ for all subjects in the study; although there is substantial variation, the coefficient of variation $\Phi_{\hat{\delta}}$ is only 9.5 per cent. We used standard regression methods to model $U(\Psi)$ as a function of covariates; the strongest predictor of post-treatment cholesterol is pre-treatment cholesterol, which is included in X.

Table II presents different G-estimates of Ψ_0. Because of the large number of subjects in the study, efficiency is not an important concern here; none the less, the estimator using the natural transformation $m(\varepsilon)=\varepsilon$, compliance-based weighting, and, most importantly, residuals from the regression of $Y(\Psi)$ on X has the smallest estimated variance. As would be expected when Φ_{δ} is small, weighting by $\hat{\delta}(X)$ improves efficiency only slightly. Using the residuals from a regression of $U(\hat{\Psi})$ on X (that is, using $h(u,x)=m(\varepsilon)w(x)$ instead of $h(u,x)=m(u)w(x)$) did improve efficiency. Use of 'robust' estimators $m(\varepsilon)=\tanh^{-1}(\varepsilon/\kappa)$ did not change the point estimates substantially but did increase the estimated variance.

Figure 4 plots the residuals $\varepsilon_{X;0}(\hat{\Psi})$ against the proportion of the nominal dose of the assigned treatment taken (in the treatment arm, this is A) in both arms and restricted quadratic regression spline estimates of the mean residuals at different compliance levels. These plots have implications both for G-estimation analysis using structural nested models and for standard regression approaches. The variability of the residuals is greater in the treatment than the placebo arm, especially for subjects with more than the lowest level of compliance. Efron

Table II. Various estimates of the effect of cholestyramine on log cholesterol, using LRC data.

Characteristics of estimator				Estimates	
Use residual from regression	Robust transformation $m(\varepsilon)=\tanh^{-1}(\varepsilon)$ used?	If so, divisor coefficient	Compliance used to weight	Point estimate	Standard error
		G-estimators			
No	No		No	−0.07866	0.00290
			Yes	−0.08019	0.01068
Yes	No		No	−0.07994	0.00233
			Yes	−0.08010	0.00232
	Yes	0.1	No	−0.07994	0.00248
		1	Yes	−0.08010	0.00232
		0.1	Yes	−0.08016	0.00247
		0.02	Yes	−0.07778	0.00314
Least-squares estimators					
	Control for baseline cholesterol?				
Intent to treat		No		−0.05763	0.00224
As treated		No		−0.08763	0.00256
		Yes		−0.08889	0.00205

and Feldman [25] noted the same phenomenon. These observations are not consistent with (even a non-deterministic version of) SDM, (1)–(2) under which the variability in both arms should be the same. This provides reason to prefer the weaker SMM and so G-estimators that use linear functions $m(\varepsilon)=\varepsilon$.

In both arms, subjects who comply well tend to have lower residuals $\varepsilon_{X;0}(\hat{\Psi})$ than poor compliers. If the model for treatment effect is true (that is, in the placebo arm, compliance has no effect), both arms provide some evidence against ignorable compliance. Thus, standard regression analyses using treatment received as a regressor overestimate the benefit of CLT for lowering cholesterol (Table II).

There are also graphical checks of model assumptions. For example, Joffe *et al.* show elsewhere [24] that, under our simple SDM or SMM (1) $E\{Y \mid \delta(X), R=1\}-E\{Y \mid \delta(X), R=0\} = \delta(X)\Psi$. Thus, one could plot the difference between smoothed estimates, $E\{Y \mid \delta(X), R=1\}-E\{Y \mid \delta(X), R=0\}$, against the compliance score $\delta(X)$. If the model is correct, the plot should be a straight line through the origin with slope Ψ. Figure 5 provides such a plot; there is slight evidence of deviation from linearity and of a non-zero origin. Alternatively, one could plot the smoothed difference $E\{U(\hat{\Psi})|X, R=1\}-E\{U(\hat{\Psi})|X, R=0\}$ against a scalar component of X or some function of X; for example, $E\{U(\hat{\Psi}) \mid X\}$, or the compliance score. Under the model (or any correctly specified SDM), the plot should be a horizontal line.

6. MODEL ELABORATION

This section considers various elaborations on our simple structural model (1). These more elaborate models may be used to check the assumptions of the simpler model.

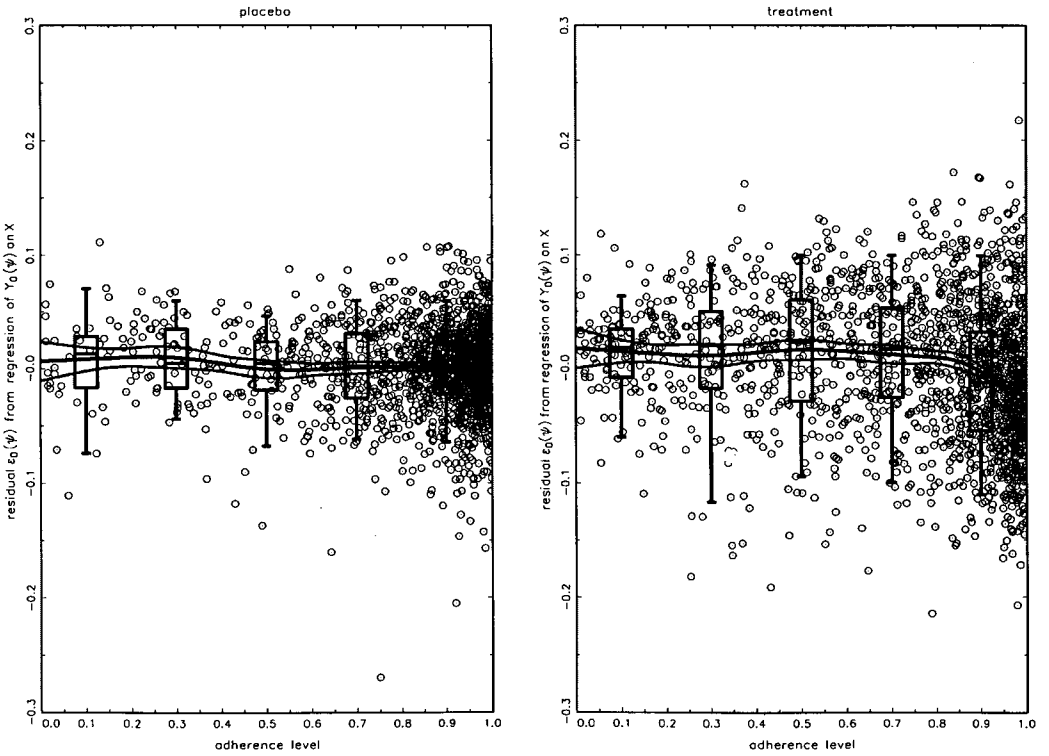

Figure 4. Scatter plots of compliance with the prescribed regimen (per cent of prescribed doses taken; this is A for subjects in the treatment arm) versus the residual from the regression of $U(\Psi)$ on covariates X. Point estimates and pointwise 95 per cent confidence intervals from a regression spline of the residual on compliance are overlaid. The end of the boxplot whiskers show the 5th and 95th percentiles of the residual for subjects in each category of compliance (0–20 per cent, 20–40 per cent, 40–60 per cent, 60–80 per cent, 80–100 per cent).

In many settings, analysts may wish to allow treatment effects to vary with levels of other baseline covariates. A generalization of the one parameter model (1) is

$$\mu_0(X,R,A)=\mu_A(X,R,A) - \Psi_0 X^* A \tag{4}$$

where X^* is a vector function of pre-treatment covariates X (which can include a constant 1) and Ψ_0 is a vector parameter. In this case, replace the scalar $g(u,x)$ by a vector function $g(u,x)$ and sum in (3). One choice of functions is $g(u,x)=w(x)q(\varepsilon)x^*$; again a product of a scalar weight, a function of the residual, and the vector function x^* of x. As previously, under the same additional assumptions about the residual variance, the most efficient estimating function is obtained by choosing compliance-based weights $\delta(X)$ and optimal function $q^*(\varepsilon)$.

We chose to fit model (4) to the LRC-CPPT data. In particular, we tested whether (centred) baseline log cholesterol, the most potent baseline predictor of outcome Y, modified the effect of treatment received and so included it in model (4); that is, X^* here is the two-vector $\{1, \text{cblogchl}\}$. The point estimate (standard error) for the interaction term is 0.060 (0.056).

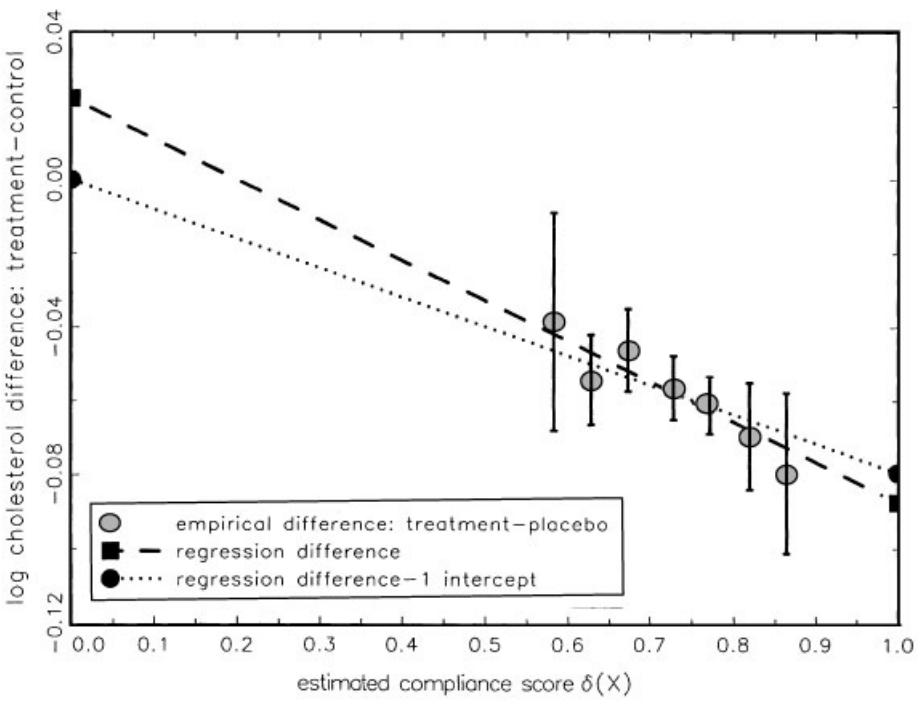

Figure 5. Plots of the empirical ITT difference between treatment and control groups within strata of the estimated compliance score $\hat{\delta}(X)$. There are also two separate linear regression-based estimates; one constraining the line to go through the origin, one not.

In models (1) and (4), the effect of treatment received is proportional to the dose received. To test this assumption, we can consider non-linear 'dose–response' relationships. Such a two-parameter generalization of (1) is

$$\mu_0(X,R,A) = \mu_A(X,R,A) - \Psi_0 A - \Psi_1 A^2 \tag{5}$$

The model can again be fit by G-estimation. Consider again functions of the form $g(u,x) = w(x)q(\varepsilon)$; here, $w(x)$ is now a vector. The optimal function $w(x)$ is $\{\delta(X), \tau(X)\}$, where $\tau(X) \equiv E(A^2|X,R=1) - E(A^2|X,R=0)$.

In the LRC-CPPT, we tested whether the effect of treatment received is proportional to the dose received; that is, if $\Psi_1 = 0$. Using the optimal weights, we obtain point estimates (standard errors) for Ψ_0 and Ψ_1 of -0.006 (0.085) and -0.100 (0.113); there is little evidence ($p = 0.37$) against the simpler model.

7. EXPLANATORY AND PRAGMATIC ESTIMATES

We have so far adopted an explanatory attitude towards G-estimation and structural nested models, in which the analysis attempts to explain differences between randomized groups in

terms of the effects of treatment received [8]. When the goal is merely explanatory, it makes little difference whether changes from the nominally assigned dose should be classified as non-compliance, or whether the effects in compliant and non-compliant individuals are the same.

Randomized trials are usually run to provide guidance in how to treat patients. These explanatory analyses do not directly answer the question of our motivated subject, or the questions of the planner who would like to anticipate what would happen were the treatment to be introduced into general practice. After a trial is completed the patterns of treatment received among people assigned to it may differ from the patterns observed in the treatment arm in the trial. For example, treatments to palliate unpleasant side-effects may become available, making it feasible and possibly desirable for subjects to continue with the assigned treatment [27]; further, compliance even in a similar population could be higher after a treatment is accepted as efficacious or lower without the encouragement given in a randomized trial [28].

To anticipate what might happen in practice, one must compare a wider variety of potential outcomes and consider more than the observed outcome $Y(A)$ and baseline outcome $Y(0)$. It is thus tempting to extend the models for treatment effects to all potential outcomes Y_a; for, example, (1) generalizes naturally to

$$\mu_a(X,R,A)=\mu_0(X,R,A)+\Psi_0 a \tag{6}$$

Such models attempt to determine the potential effects of any arbitrary treatment regime on outcome. Note that formulation (6), which is more ambitious, implies (1), but not vice versa. Comparisons of $E(Y_1)$ and $E(Y_0)$, including average treatment effects $E(Y_1-Y_0)=E(Y_1)-E(Y_0)$, are implied by (6) but not by (1). (6) makes the untestable (given the data as described) assumption that the effects of additional increments of treatment are the same among people whose observed treatment $A=a$ and those whose observed treatment $A<a$; that is, among subjects with better and worse compliance.

Suppose that the reasons for non-compliance in the treatment group are associated with side-effects, and that the level of side-effects is related to the anti-cholesterol effects at a given dose (both might be influenced by variability in metabolism of the drug). Then, one might expect the magnitude of expected individual treatment effects $E(Y_a - Y_0 \mid X,R,A)$ to be decreasing (or at least non-increasing) in A, and ignorable compliance would fail to hold. This does not necessarily invalidate structural model (1) (or more complicated functions like (5)), whose applicability is subject to empirical investigation, as above. However, under this scenario of non-ignorable non-compliance, (6) will not generally hold; Robins [19] terms this phenomenon current treatment interaction.

Consider examining the sensitivity of pragmatic conclusions to the assumption. We generalize (6) to allow the effect that additional treatment (beyond A, what the subject has in fact taken) to be a proportionally smaller than the realized effect of the treatment already taken. One such model is

$$\begin{aligned} \mu_a(X,R,A) &= \mu_0(X,R,A)+\Psi_0 a, && a \leqslant A \\ \mu_a(X,R,A) &= \mu_0(X,R,A)+\Psi_0 \theta(a-A), && a > A \end{aligned} \tag{7}$$

where θ is the relative effect of additional treatment, compared to the realized effect of treatment received. The data provide no information about θ.

Table III. Estimates of effect of encouragement to use CLT on log cholesterol, under different assumptions about effects of additional CLT, on how much compliance can be improved.

How much can encouragement affect compliance?	Extra effect (θ) of additional compliance for subjects with improvable compliance		
	0	36892	1
None ($a=A$)	−0.058	−0.058	−0.058
30 per cent improvement ($a=(A+1)/2$)	−0.058	−0.061	−0.064
100 per cent improvement ($a=1$)	−0.058	−0.068	−0.079

In a pragmatic analysis, one might want to predict the mean outcome under different but attainable compliance conditions than those observed in the study. This involves speculation not only about the relative extra effect θ but also about what compliance levels are attainable.

We used model (7), together with different assumptions about the attainable improvement in compliance, to predict the mean (log) cholesterol that would have been seen had subjects complied differently with their assigned treatments and so speculate what gains might be attainable in practice. Table III shows the expected means under different assumptions about these unknown quantities. If $\theta=0$ or no improvement in compliance is possible, the intent-to-treat effect is also the pragmatic effect. If $\theta=1$ and full compliance is possible, the G-estimate of Ψ_0 is also the pragmatic estimate. Intermediate values of the relative effect θ and of how much compliance can be improved produce intermediate attainable gains due to CLT.

8. DISCUSSION

The increased efficiency of weighted estimators is potentially useful for both explanatory and pragmatic analysis. Better weights can result in more power against the null. It is often claimed in trials with negative results that, had compliance been better, the experimental treatment might have appeared more beneficial than the analysis performed suggests. The approach suggested here focuses attention on strata with better compliance, in which poor compliance cannot easily explain the absence of differences between randomized groups; thus, the weighting approach suggested here can permit better assessment of the reasonableness of such explanations of negative findings.

Our approach is complementary to the more standard ITT approach. In fact, tests of the null using both compliance-based weighting and the adjustments we have indicated for pre-treatment covariates X as predictors of the potential outcome Y_0 are still ITT tests of the null hypothesis. The tests have the correct size under the null, since only their efficiency, but not their validity, depends on the distributions of compliance $f(A|X,R)$ and the potential outcome $f(Y_0|X)$. Thus, our approach will not be inconsistent with an appropriately powerful ITT test of the null hypothesis, and will not find a positive effect if an appropriate ITT test does not.

Our approach may sometimes be inconsistent with non-parametric ITT estimates of effect (that is, simple comparisons of the randomized groups). Our measures of treatment efficacy will generally be farther from the null than simple ITT estimates of effectiveness. Because

our estimates depend on modelling assumptions while the simpler ITT estimates do not, we suggest that both be reported in the presence of non-compliance.

Other investigators have also proposed compliance-based weighting schemes [9–12]. Our approach generalizes the work of Goetghebeur and Lapp [10] and Fischer-Lapp and Goetghebeuer [11], who consider only the identity transformation of the error function, $m(\varepsilon)=\varepsilon$. Their restriction of the functions $m(\varepsilon)$ is dictated by the use of a structural nested mean model [6] rather than the stronger structural nested distribution model [19, 29]. Using non-identity functions $m(\varepsilon)$ can increase efficiency and reduce sensitivity to outliers in Y or ε when the stronger SDM holds.

Most of the methodologic findings are consequences of the very general theory developed by Robins. None the less, the application of this theory to specific situations has not been worked out explicitly in many cases. In particular, formulae developed previously [11] do not explicitly identify the compliance score and its role or importance in the broader array of settings considered here (that is, in settings where robust estimators are used or placebo-group subjects have access to the active treatment).

There are competing considerations in the choice of the function $g(u,x)$ to include in the estimating equations (3) and whether to use compliance-based weights. Analysts using G-estimation for structural nested models may need to formulate and fit several models: the structural models themselves; the model for treatment assignment/randomization, the model for compliance, and the model for the baseline outcome Y_0 given baseline covariates. Each model to be fit takes the some of the investigator's limited resources and complicates explanation of the procedures used. Thus, it is useful to know when the effort invested is likely to yield substantial efficiency benefits. Our results help determine when the extra effort will yield those substantial benefits. However, even when compliance-based weights will not substantially improve efficiency, understanding why subjects fail to comply may be of scientific and practical interest.

It is worth comparing our assumptions to those of other methods proposed for dealing with non-compliance. Efron and Feldman [25] proposed an approach not explicitly using the potential outcomes approach for derivation. A primary assumption in their approach is that someone in the jth percentile of the compliance distribution in the treatment arm would have been in the jth percentile of compliance had he been in the placebo arm, and vice versa. Our approach makes no corresponding assumption, which is hard to justify if side-effects of the treatment are important determinants of non-compliance.

The approach of Angrist *et al.* [2] and Imbens and Rubin [22] is more similar. Their approach assumes that compliance is all-or-none; this is a special case in our approach. They sometimes further assume that there are no defiers (that is, subjects who would receive treatment if and only if randomized to placebo); our corresponding assumption is that the effect in defiers and compliers (subjects who would receive treatment if and only if randomized to treatment) is the same. If, as in the LRC-CPPT, subjects randomized to placebo cannot receive treatment, the assumption is not necessary. Under the assumption that there are no defiers, their estimator and our simple estimator (no weights or residuals) are asymptotically equivalent.

In most studies of pharmaceutical agents, compliance is not a one-time event but rather may vary over time, even from day to day. The discussion above postulates that the average level of compliance with the prescribed regimen, but no other aspect of treatment received, determines response to treatment. This will often not be the case; in particular, when post-randomization clinical course influences treatment received, there will likely be non-random

fluctuations in compliance level over time. In such settings, one might want to consider explanatory alternatives to the simple model that allow the effects of treatment received at different times to depend on the time of treatment (in relation to the end of the study); Robins provides a class of such models, the structural nested models, and extends the estimation procedures described here to those settings [30]. Again, it is possible to develop G-estimators which provide more weight to compliant strata than non-compliant ones. Here, the form of the optimal estimators will often be intractable; we would expect compliance-based estimators to be both simple and often reasonably efficient.

Our use of the terms *explanatory* and *pragmatic* differs somewhat from previous uses in the literature [18]. We do not identify 'as treated' as explanatory, because invalid estimates of treatment effect explain nothing. Similarly, for the reasons noted above, intent-to-treat are not necessarily pragmatic. The G-estimation/instrumental variables approach presented here provides a unified approach to explaining associations observed in randomized trials and to predicting the effects of a future intervention or treatment; the latter task is more ambitious and requires additional assumptions or a sensitivity analysis.

For both goals, information on treatment received and on pre-treatment covariates is potentially useful. Information on treatment received allows for better-justified explanatory analysis than 'as treated' analyses and its variants, permits more powerful ITT analyses, and provides a basis for better justified and understood pragmatic analyses. The gain in power and precision for the approach sketched here requires measurement of pre-treatment covariates X that predict compliance; the gain will likely be most dramatic in studies with substantial non-compliance and identifiable strata with substantially better compliance.

APPENDIX: OUTLINE OF PROOFS OF THEOREMS

Theorem 1

Using a Taylor series expansion, rewrite the estimating equations (3) as $0=\sum\{R-p\}q\{\varepsilon(\Psi)\}w(X)=\sum\{R-p\}q\{\varepsilon(\Psi_0)\}w(X)+(\Psi-\Psi_0)\sum\{R-p\}[\partial q\{\varepsilon(\Psi_0)\}/\partial\Psi_0]w(X)$. Let $B(w)\equiv\sum\{R-p\}q\{\varepsilon(\Psi_0)\}w(X)$ and let $C(w)\equiv\sum\{R-p\}[\partial q\{\varepsilon(\Psi_0)\}/\partial\Psi_0]w(X)$. For a given choice of weights, the variance $V(w)$ of the resulting estimator is $\mathrm{var}\{B(w)\}/[E\{(C(w))\}]^2$. Now $\mathrm{var}\{B(w)\}=E[\mathrm{var}\{B(w)\,|\,X,\varepsilon\}]+\mathrm{var}[E\{B(w)\,|\,X,\varepsilon\}]=E[\mathrm{var}\{B(w)\,|\,X,\varepsilon\}]$ (because randomization guarantees that $E(B\,|\,X,\varepsilon)=0$). Let $B_i\equiv\{R-p\}q\{\varepsilon(\Psi_0)\}w(X)$; because the subjects are i.i.d., write $\mathrm{var}\{B(w)\,|\,X,\varepsilon\}=\sum\mathrm{var}(B_i\,|\,X,\varepsilon)$. Now $\mathrm{var}(B_i\,|\,X,\varepsilon)=w(X)^2q(\varepsilon)^2\mathrm{var}(R-p\,|\,X)=w(X)^2q(\varepsilon)^2p(1-p)$; then $\mathrm{var}(B_i\,|\,X)=E\{\mathrm{var}(B_i\,|\,X,\varepsilon)\}=K_1w(X)^2p(1-p)$, where $K_1\equiv E\{q(\varepsilon)^2\}$, a constant not dependent on X. Treating the X's as fixed, we obtain $\mathrm{var}(B)=K_1\sum w(X)^2p(1-p)$.

Let $C_i\equiv\{R-p\}[\partial q\{\varepsilon(\Psi_0)\}/\partial\Psi_0]w(X)$; then $E(C)=\sum E(C_i)$. Because $\partial\varepsilon(\Psi_0)/\partial\Psi_0=-A$, $E(C_i\,|\,R,X,\varepsilon)=\{R-p\}w(X)E([\partial q\{\varepsilon(\Psi_0)\}/\partial\Psi_0]\,|\,R,X,\varepsilon)=-\{R-p\}w(X)q'(\varepsilon)E[A\,|R,X,\varepsilon]$. Under random non-compliance, $E(A\,|\,R,X,\varepsilon)=E(A\,|\,R,X)$, so $E(\text{Confidence interval}\,|\,R,X,\varepsilon)$ simplifies to $-\{R-p\}w(X)q'(\varepsilon)E(A\,|\,R,X)$. Let $K_2\equiv-E\{Q'(\varepsilon)\}$, which is independent of R and X under random non-compliance. Summing over the distributions of ε and R successively, we obtain $E(C_i\,|\,R,X)=K_2(R-p)w(X)E(A\,|\,R,X)$ and $E(C_i\,|\,X)=K_2w(X)E\{(R-p)E(A\,|\,R,X)|X\}=K_2w(X)p(1-p)\delta(X)$. Treating the X's as fixed leads to $E(C)=K_2\sum p(1-p)w(X)\delta(X)$.

We now seek to find the weights **w** which minimize the variance of $\mathrm{var}(B)/\{E(C)\}^2$. Let X_0 denote an arbitrary reference stratum defined by covariates X. Without loss of generality, let $w(X_0)=1$. The variance is maximized when the derivatives $\partial V(w)/\partial w(X)=0$ for all $X \neq X_0$, which in turn happens when $[\partial \mathrm{var}\{B(w)\}/\partial w(X)]\{E(C(w))\}^2 = 2\mathrm{var}\{B(w)\}E(C(w))[\partial E\{C(w)\}/\partial w(X)]$. which in turn implies $[\partial \mathrm{var}\{B(w)\}/\partial w(X)]E\{C(w)\} = 2\mathrm{var}\{B(w)\}[\partial E\{C(w)\}/\partial w(X)]$. Substituting in the expressions for $B(w)$ and $C(w)$ leads to $2K_1 w(X)p(1-p)K_2 \sum p(1-p)w(X)\delta(X) = 2K_2 p(1-p)\delta(X)K_1 \sum w(X)^2 p(1-p)$. Let $K_3 \equiv \sum p(1-p)w(X)\delta(X)$ and $K_4 \equiv \sum w(X)^2 p(1-p)$. Further reorganization leads to $w(X)=\delta(X)(K_4/K_3) \propto \delta(X)$, yielding a solution $w(X)=\delta(X)/\delta(X_0)$.

Although in practice neither the effect of randomization on treatment received $\delta(X)$ nor the residual $\varepsilon_0(\Psi)$ will be known precisely, it follows from Robins *et al.* [31] that replacing these quantities by consistent estimates thereof does not affect the asymptotic variance.

Theorem 3

From above, we have, for an estimator $\hat{\Psi}_{w,q}$ with weights $w(X)$ and a common distribution of errors that $\mathrm{var}(\hat{\Psi}_{w,q}) = K_1 \sum w(X)^2 p(1-p)/\{K_2 \sum p(1-p)w(X)\delta(X)\}^2$. Because K_1 and K_2 do not depend on the weighting scheme, and, in a randomized trial, there is a common value for p, we can write, taking expectations $\mathrm{var}(\hat{\Psi}_{w,q}) \propto E\{w(X)^2\}/[E\{w(X)\delta(X)\}]^2$; the relative efficiencies are then $\mathrm{var}(\hat{\Psi}_{1,q})/\mathrm{var}(\hat{\Psi}_{\delta,q}) = [E\{\delta(x)\}]^{-2}/[E\{\delta(x)^2\}]^{-1} = 1 + \Phi^2_{\delta(X)}$.

ACKNOWLEDGEMENTS

The author thanks Drs Basil Rifkind, David Gordon and Michael Hartman for assistance in obtaining data from the LRC-CPPT study, and Dr James Robins and a referee for useful comments. This work was funded by a grant from the National Heart, Lung and Blood Institute (R29 HL59184).

REFERENCES

1. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Statistics in Medicine* 1991; **10**:1595–1605.
2. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association* 1996; **91**:444–472.
3. Baker SG, Lindeman KS. The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statistics in Medicine* 1994; **13**:2269–2278.
4. Mark SD, Robins JM. A method for the analysis of randomized trials with compliance information: an application to the multiple risk factor intervention trial. *Controlled Clinical Trials* 1993; **14**:79–97.
5. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics–Theory and Methods* 1991; **20**:2609–2631.
6. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics – Theory and Methods* 1994; **23**:2379–2412.
7. Sommer A, Zeger SL. On estimating efficacy from clinical trials. *Statistics in Medicine* 1991; **10**:45–52.
8. White IR, Goetghebeur EJ. Clinical trials comparing two treatment policies: which aspects of the treatment policies make a difference. *Statistics in Medicine* 1998; **17**:319–339.
9. Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. *Statistics in Medicine* 1997; **16**:1017–1029.
10. Goetghebeur E, Lapp K. The effect of treatment compliance in a placebo-controlled trial: regression with unpaired data. *Applied Statistics* 1997; **46**:351–364.
11. Fischer-Lapp K, Goetghebeur E. Practical properties of some structural mean analyses of the effect of compliance in randomized trials. *Controlled Clinical Trials* 1999; **20**:531–546.
12. Zelen M. Randomized consent designs for clinical trials: an update. *Statistics in Medicine* 1990; **9**:645–656.

13. Lipid Research Clinic Program. The Lipid Research Clinics Coronary Primary Prevention Trial results. I. Reduction in incidence of coronary heart disease. *Journal of the American Medical Association* 1984; **251**: 351–364.
14. Lipid Research Clinic Program. The Lipid Research Clinics Coronary Primary Prevention Trial results. II. The relationship of reduction in incidence of coronary heart disease to cholesterol lowering. *Journal of the American Medical Association* 1984; **251**:365–374.
15. Heckman J. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources* 1997; **32**:441–462.
16. Philipson T, Desimone J. Experiments and subject sampling. *Biometrika* 1997; **84**:619–630.
17. Friedman L, Furberg C, DeMets D. *Fundamentals of Clinical Trials*. 3rd edn. Springer-Verlag: New York 1998.
18. Newcombe RG. Explanatory and pragmatic estimates of the treatment effect when deviations from alloated treatment occur. *Statistics in Medicine* 1988; **7**:1179–1186.
19. Robins JM, Rotnitzky A, Scharfstein DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology*, Halloran E, Berry D (eds). Springer-Verlag, New York 2000:1–99.
20. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**:41–55.
21. Robins JM. Causal inference from complex longitudinal data. In *Lecture Notes in Statistics-Latent Variable Modeling with Applications to Causality*, Berkane M (ed.). Springer-Verlag: New York, 1997;69–117.
22. Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* 1997; **25**:305–327.
23. Follmann DA. On the effect of treatment among would-be treatment compliers: an analysis of the multiple risk factor intervention trial. *Journal of the American Statistical Association* 2000; **95**:1101–1109.
24. Joffe MM, Ten Have TR, Brensinger C. The compliance score as a regressor in randomized trials. *Biostatistics* 2003;(in press).
25. Efron B, Feldman D. Compliance as an explanatory variable in clinical trials (with discussion). *Journal of the American Statistical Association* 1991; **86**:9–26.
26. Greenland S. Alternative models for ordinal logistic regression. *Statistics in Medicine* 1994; **13**:1665–1677.
27. Robins JM, Greenland S. Adjusting for differential rates of prophylaxis therapy for PCP in high- versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association* 1994; **89**:737–749.
28. Robins JM. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology: A Focus on AIDS*, Sechrest LA (ed.). NCHSR, U.S. Public Health Service: 1989; 113–159.
29. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: *Statistical Models in Epidemiology: The Environment and Clinical Trials*. Halloran ME, Berry D (eds.). Springer-Verlag: New York, 1999; 95–134.
30. Robins JM. Testing and estimation of direct effects by reparameterizing directed acyclic graphs with structural nested models. In *Computation*, *Causation*, *and Discovery*, Glymour C, Cooper G (eds.). AAAI Press/The MIT Press: Menlo Park, CA, 1999; 349–405.
31. Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* 1992; **48**:479–495.

