STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:2639–2652
Published online 3 July 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3636

Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity

Roseanne McNamee*, †

Biostatistics Group, School of Community-based Medicine, University of Manchester, Oxford Road, Manchester M13 9PL, England

SUMMARY

We consider the behaviour of three approaches to efficacy estimation—using so-called 'as treated' (AT), 'per protocol' (PP) and 'instrumental variable' (IV) analyses—and of the Intention to Treat estimator, in a two-arm randomized treatment trial with a Normally distributed outcome when there is treatment effect heterogeneity and non-random compliance with assigned treatment. Formulae are derived for the bias of estimators when used either to estimate average treatment effect (ACE) or to estimate complier average treatment effect (CACE) under several models for the relationship between compliance and potential outcomes. These enable the expected values of AT, PP and IV estimators to be ranked in relation to ACE, and show that AT and PP estimators are generally biased for both ACE and CACE even under homogeneity. However, we show that the difference between any pair of (AT, PP, IV) estimates can be used to estimate the correlation between the latent variable determining compliance behaviour and one potential outcome. In the absence of measures that predict compliance, bounds for ACE can only be set given strong assumptions. Regarding the Intention to Treat estimator, while this is 'biased towards the null' if viewed as a measure of CACE, we show that it is not always so in relation to ACE. Finally we discuss the behaviour of the estimators under weak and strong null hypotheses. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: RCT; efficacy; bias; per protocol analysis; instrumental variable

1. INTRODUCTION

Treatment received in a randomized treatment trial may differ from that allocated by randomization, but investigators may be interested in the effect if treatment had been taken as intended, i.e. in treatment *efficacy*. If deviations from intended treatment are non-random with respect to prognostic factors, then an 'as treated' (AT) analysis, which groups patients according to treatment received rather than intended, may be confounded by these factors. For similar reasons, a 'per protocol'

*Correspondence to: Roseanne McNamee, Biostatistics Group, School of Community-based Medicine, University of Manchester, Oxford Road, Manchester M13 9PL, England.
†E-mail: rmcnamee@manchester.ac.uk

Received 20 December 2008
Accepted 14 May 2009

(PP) analysis, which groups by randomized arm but drops patients who do not follow the protocol, may be biased. An 'intention to treat' estimator (ITTe), which compares average outcomes across all patients between arms, does not suffer in this way if there is no loss of data, but estimates treatment *effectiveness* [1] not efficacy. In recent years, new methods for efficacy estimation under non-compliance, which preserve the advantages of randomization, have been proposed: see for example, a *Statistics in Medicine* issue [2] and, more recently, White [3]. One class of methods consists of 'instrumental variable' estimators.

In a randomized trial comparing an active and a control treatment on a continuous variable, Y, the IVe estimator (IVe) is defined as $(\bar{y}_1-\bar{y}_0)/(p_1-p_0)$, where $\bar{y}_1$ and $\bar{y}_0$ are the observed means in the active and control arms, and p_1, p_0 the proportions of subjects in each arm who take the active treatment. It is important to realize that, subject to certain assumptions [4], this IVe is strictly an estimate of average efficacy in the 'complier' subpopulation of patients, known as the complier average causal effect (CACE). (Compliers are subjects which would comply with either treatment programme.) The average potential efficacy across all patients, i.e. the average causal effect (ACE), might also be of interest; for example, if it were thought that the subpopulation represented by non-compliers in a trial could in future be persuaded to become compliers [5]. ACE and CACE may be different when treatment effects vary. In their landmark paper on IVe and ensuing discussion, Angrist *et al.* [4, 6] acknowledged that the IVe might be biased for ACE; yet the IVe has recently [7] been criticized again on these grounds, drawing further comment on its favourable properties relative to CACE [8]. Indeed, without further strong assumptions or weaker assumptions in combination with predictors of compliance, point estimation of ACE is not possible although bounds have been derived [9–11] for the binary outcome case. Nevertheless, there is evidence [12] that readers may generalize from CACE to ACE regardless; hence, the relationship between these two parameters under various models for non-compliance is of interest.

Although authors [13, 14] have drawn attention to the limitations of PP and AT analyses, the PP approach is sometimes cautiously recommended along with ITTe, especially for equivalence trials [15, 16]. The premise is that, as a measure of efficacy, the ITTe is 'biased towards the null'—which is undesirable in equivalence trials—and that a PP estimator (PPe) is less so. A systematic review of publications found that ITT and PP results were frequently very different, with PP tending to be further from the null [17]. Whether PPe is to be seen as an estimator of CACE or ACE is usually not stated but, by analogy with the IVe, one might argue that both PPe and ATe should at best be considered estimators of CACE. However, a recent paper by Bang and Davis (B&D) that considered PP and AT estimators in relation to ACE concluded that, compared to IVe, the PP and AT estimators 'may provide the smallest bias in some situations' [7]. If true, this would be an important advantage when ACE is of interest.

This paper aims to clarify the relationships between the two potential targets, CACE and ACE, and the expected values, E[PPe], E[ATe] and E[ITTe]. It is assumed that the conditions hold (e.g. exclusion restriction [4]) under which E[IVe]=CACE. It will emerge that, in terms of bias, there is no reason to prefer ATe or PPe over IVe as estimators of CACE or ACE: the good performance reported [7] for PPe is shown to be specific to a very restricted set of assumptions. It is also shown that E[PPe] may sometimes be closer to the null than E[ITTe]. However we show that differences between ATe, PPe and IVe may be informative. The focus then shifts to the ITTe: it is shown that, as an estimator of ACE, ITTe can be biased away from zero or even have a different sign, but has the reassuring property of 'bias towards the null' [18] in relation to CACE. Finally, we address the behaviour of the estimators under three previously defined 'null' hypotheses: two of which we term as 'weak' and one 'strong'. A previous suggestion that ITT can be biased 'when

DOI: 10.1002/sim

the hypothesis of no treatment effect is true' [7] is shown to apply to a weak null hypothesis, which, we argue, has little practical relevance.

Results are derived under several very general, although parametric, models for compliance behaviour. Our work is complimentary to a recent, 'distribution-free' approach [19], where the authors considered what assumptions would be necessary so that IVe, PPe and ATe are valid as estimators of CACE (but did not address ACE). The results are discussed in the context of a depression trial in Section 7. Attention is restricted to trials where patients may cross over to the non-allocated treatment, but not opt out of both. Subjects are assumed to either comply totally with their allocated regime or not at all. We focus on the expected values of estimators and ignore sampling error.

2. ACE AND CACE

Following Rubin [20], we define the *causal effect* of active treatment on an outcome Y in an individual i as $Y_i(1)-Y_i(0)$ where $Y_i(1)$ and $Y_i(0)$ are the *potential* values of Y for individual i under active and control treatments, respectively. We assume that causal effects vary across individuals, i.e. *treatment effect heterogeneity.* It may be helpful to think of Y as a measure of severity of illness, so that low values are desirable. The ACE is $E[Y(1)-Y(0)]$ where, here and throughout, the notation $E[\,]$ is used to refer to the average over individuals in the trial. The CACE is $E[Y(1)-Y(0)|C=1]$ where C is a 1/0 variable: $C=1$ for the subpopulation of individuals (compliers) who would always comply with their allocated treatment regime, regardless of whether it is active or control. Since C is a measure of *two* potential behaviours, it is not generally observed in all patients; it should be distinguished from observed compliance. It is easy to show that: $\text{CACE}\neq\text{ACE}\Leftrightarrow E[Y(1)-Y(0)|C=1]\neq E[Y(1)-Y(0)|C=0]$. A related condition, stating that treatment effects vary between those receiving and not receiving active treatment, has been described as 'essential heterogeneity' [21].

3. MODELS FOR POTENTIAL OUTCOMES AND NON-COMPLIANCE

We assume that subjects are randomly allocated to treatment arms with equal probability and that $Y(1)$ and $Y(0)$ have a bivariate Normal distribution with means μ_1 and μ_0, respectively, variances σ_1^2, σ_0^2 and correlation ρ_{01}; $\text{ACE}=\mu_1-\mu_0$. The correlation ρ_{01} is also an inverse measure of treatment effect heterogeneity. To see this, let $\lambda=\sigma_1/\sigma_0$; then a standardized measure of heterogeneity relative to variation in Y is $V[Y(1)-Y(0)]/\sigma_0\sigma_1=(1+\lambda^2-2\lambda\rho_{01})/\lambda=k$, say. If $\lambda=1$, then if $\rho_{01}=1$, we have $k=0$, i.e. a homogeneous treatment effect.

Non-compliance behaviour, C, is assumed to be completely determined by an unobserved latent variable, B, which has a N(0,1) distribution. Six possible models (MBa–MBf) are considered:

- MBa: non-compliers ($C=0$) are subjects with $B<-\delta$ and who would never take active treatment regardless of allocation ('never takers'). All others are compliers ($C=1$).
- MBb: non-compliers ($C=0$) are subjects with $B\geqslant\delta$ who would always get and take the active treatment: they are 'always takers'. All others are compliers with $C=1$.
- MBc: $C=1$ if $-\delta_1\leqslant B<\delta_2$; there are two different non-complier groups defined as in MBa and MBb, respectively.

DOI: 10.1002/sim

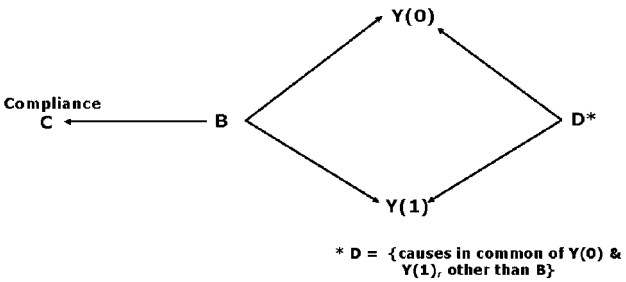

Figure 1. Possible causal relationships underlying non-random compliance.

MBd–MBf are similar to MBa–MBc except that never takers/always takers are taken from the opposite tail:

MBd: $C=0$ (never takers) if $B>\delta$;
MBe: $C=0$ (always takers) if $B<-\delta$;
MBf: $C=1$ if $-\delta_2 \leqslant B<\delta_1$ with non-compliers as in MBd and MBe.

In MBa–MBc, there is a positive correlation between B and taking active treatment while in MBd–MBf, it is negative.

Next, we suppose B to be correlated with $Y(0)$ and $Y(1)$ with coefficients $\rho_{B0} \geqslant 0$ and $\rho_{B1} \geqslant 0$, respectively. These non-negative values are not restrictive since we could replace B by $-B$ to achieve them. Figure 1 illustrates possible causal relations between $C, B, Y(0)$ and $Y(1)$: the correlation of the latter pair can arise not only through B but also possibly through a set of other variables, D, not correlated with compliance. We also assume that the partial correlation between $Y(0)$ and $Y(1)$, given B, is non-negative; this means that $\rho_{B0}\rho_{B1} \leqslant \rho_{01}$, with equality when there is no partial correlation (i.e. no D). The models simulated by B&D [7] are a special case of MBa–MBc in which $\rho_{B0}=1, \rho_{B1}=0=\rho_{01}$: B is redundant here as now C is completely determined by the tails of $Y(0)$, and $Y(1)$ and $Y(0)$ are completely uncorrelated. They also assumed $\delta_1=\delta_2$ in MBc.

Under (a),(b),(d) and (e), the proportion of non-compliers is $1-\Phi(\delta)$ where $\Phi(x)$ is the cumulative Normal distribution. In (c) and (f), the proportions of never takers and always takers are $1-\Phi(\delta_1)$ and $1-\Phi(\delta_2)$, respectively. In the Appendix, formulae for CACE, E[PPe], E[ATe] and E[ITTe] are derived for all models. For the presentation below, we simplify by assuming $\sigma_0^2=\sigma_1^2=\sigma^2$, whence $k=V[Y(1)-Y(0)]/\sigma^2=2(1-\rho_{01})$.

4. RESULTS

4.1. Difference between CACE and ACE

It can be shown (see Appendix) that, under MBa–MBf, CACE is equal to

$$\text{MBa, MBe: } E[Y(1)-Y(0)|C=1]=\text{ACE}-\sigma(\rho_{B0}-\rho_{B1})\frac{f(\delta)}{\Phi(\delta)} \tag{1}$$

$$\text{MBb, MBd: } E[Y(1)-Y(0)|C=1]=\text{ACE}+\sigma(\rho_{B0}-\rho_{B1})\frac{f(\delta)}{\Phi(\delta)} \tag{2}$$

DOI: 10.1002/sim

$$\text{MBc: } E[Y(1)-Y(0)|C=1] = \text{ACE} - \sigma(\rho_{B0}-\rho_{B1})\frac{f(\delta_1)-f(\delta_2)}{\Phi(\delta_2)+\Phi(\delta_1)-1} \tag{3}$$

$$\text{MBf: } E[Y(1)-Y(0)|C=1] = \text{ACE} + \sigma(\rho_{B0}-\rho_{B1})\frac{f(\delta_1)-f(\delta_2)}{\Phi(\delta_2)+\Phi(\delta_1)-1} \tag{4}$$

where $f(x)$ is the Normal density function. CACE=ACE when $\rho_{B0}=\rho_{B1}$; also, in (c) and (f), CACE=ACE when $\delta_1=\delta_2$ even if $\rho_{B0}\neq\rho_{B1}$. More generally, the sign of CACE–ACE depends on which tail of B the non-compliers are drawn from, and whether compliance is more strongly correlated with $Y(0)$ or with $Y(1)$. To use these formulae to infer ACE from an estimate of CACE, we would need to know $\rho_{B0}-\rho_{B1}$. Without such information, we can only place an upper bound on $|\text{CACE}-\text{ACE}|/\sigma$ if we are prepared to make assumptions: for example if we specify k, and assume $Y(0)$ and $Y(1)$ have no causes in common except B (i.e. no D in Figure 1), then $\rho_{B0}\rho_{B1}=\rho_{01}\Rightarrow|\rho_{B0}-\rho_{B1}|\leqslant 1-\rho_{01}=k/2$. This situation contrasts with that for binary outcome data where bounds can be set without any such specification [9–11].

4.2. Expected value of PPe

For the PP estimators, those *observed* not to comply are dropped from the analysis. Under MBa–MBf, the expected values are (see Appendix):

$$\text{MBa: } E[\text{PPe}] = \text{ACE} + \rho_{B1}\frac{\sigma f(\delta)}{\Phi(\delta)} = \text{CACE} + \rho_{B0}\frac{\sigma f(\delta)}{\Phi(\delta)} \tag{5}$$

$$\text{MBb: } E[\text{PPe}] = \text{ACE} + \rho_{B0}\frac{\sigma f(\delta)}{\Phi(\delta)} = \text{CACE} + \rho_{B1}\frac{\sigma f(\delta)}{\Phi(\delta)} \tag{6}$$

$$\text{MBc: } E[\text{PPe}] = \text{ACE} + \rho_{B1}\frac{\sigma f(\delta_1)}{\Phi(\delta_1)} + \rho_{B0}\frac{\sigma f(\delta_2)}{\Phi(\delta_2)} \tag{7}$$

The results for MBd–MBf correspond to those for MBa, MBb and MBc, respectively, except that all positive signs in (5–7) are changed to negative.

On average, when $\rho_{B0}>0$ and $\rho_{B1}>0$, PPe exceeds ACE under MBa–MBc, but underestimates it under MBd–MBf. For the 2-sided non-compliance models (c) and (f), the bias relative to ACE is equal to the sum of the corresponding one-sided biases. On average, PPe also exceeds CACE under (a) and (b), but the sign of the difference $E[\text{PPe}]-\text{CACE}$ is not fixed for (c) and (f).

If we set $\rho_{B0}=1, \rho_{B1}=0$, as in the models simulated by B&D then, under (a), $E[\text{PPe}]=$ ACE and under (b), $E[\text{PPe}]=\text{CACE}$: their conclusion [7] that PPe could be superior to IV for ACE estimation was based on this special case of MBa. However, our more general results suggest that, at least in expectation, IVe is closer to ACE: for one-sided non-compliance models, we have $|E[\text{IVe}]-\text{ACE}|*\Phi(\delta)/\rho f(\delta)=|\rho_{B0}-\rho_{B1}|$ while $|\text{E[PPe]}-\text{ACE}|*\Phi(\delta)/\rho f(\delta)=|\rho_{B0}|$ or $|\rho_{B1}|$.

4.3. Expected value of ATe

The results for ATe depend on the number of subjects in each arm, except when they are equal. We assume equal numbers here with the general case given in the Appendix. We have

$$\text{MBa: } E[\text{ATe}] = \text{ACE} + \sigma f(\delta)\left[\frac{\rho_{B1}}{\Phi(\delta)} + \frac{\rho_{B0}}{2-\Phi(\delta)}\right] = E[\text{PPe}] + \rho_{B0}\frac{\sigma f(\delta)}{2-\Phi(\delta)} \tag{8}$$

DOI: 10.1002/sim

$$\text{MBb: } E[\text{ATe}] = \text{ACE} + \sigma f(\delta)\left[\frac{\rho_{B1}}{2-\Phi(\delta)} + \frac{\rho_{B0}}{\Phi(\delta)}\right] = E[\text{PPe}] + \rho_{B1}\frac{\sigma f(\delta)}{2-\Phi(\delta)} \tag{9}$$

$$\text{MBc: } E[\text{ATe}] = \text{ACE} + \sigma[f(\delta_1)+f(\delta_2)]\left[\frac{\rho_{B1}}{1-\Phi(\delta_2)+\Phi(\delta_1)} + \frac{\rho_{B0}}{1-\Phi(\delta_1)+\Phi(\delta_2)}\right] \geqslant \text{PPe} \tag{10}$$

Results for MBd–MBf are (8)–(10), respectively, but with the plus signs directly after ACE and after E(PPe] replaced by negative signs. Overall, the direction of difference between E[ATe] and ACE shows a similar dependence on model assumptions to that for E[PPe]−ACE, but the absolute differences with ACE are greater for ATe.

Equations (1–10) together enable a ranking of E[ATe], E[PPe], CACE and ACE for the four 'one-sided' non-compliance models, MBa–MBb and MBd–MBe; see Table I. For the two-sided models, we can rank all but CACE: for example under MBc, E[ATe]>E[PPe]>ACE except if $\delta_1 = \delta_2 = 0$ —which corresponds to half of each arm being non-compliant—hence E[ATe]=E[PPe].

Equations (5–9) suggest that we could estimate either ρ_{B0} under MBa/MBd or ρ_{B0} under MBb/MBe from the observed difference between any pair of (ATe, PPe and IVe), and estimated non-compliance rates. For example, (5) suggests the estimator $\hat{\rho}_{B0} = (\hat{\text{PPe}} - \hat{\text{IVe}})\hat{\Phi}(\delta)/\hat{\sigma} f(\delta)$ under MBa. But we cannot estimate both correlations from these statistics alone.

4.4. Expected value of ITTe

$$\text{MBa, MBe: } E[\text{ITTe}] = \Phi(\delta)\text{ACE} - \sigma(\rho_{B0}-\rho_{B1}) f(\delta) = \Phi(\delta)\text{CACE} \tag{11}$$

$$\text{MBb, MBd: } E[\text{ITTe}] = \Phi(\delta)\text{ACE} + \sigma(\rho_{B0}-\rho_{B1}) f(\delta) = \Phi(\delta)\text{CACE} \tag{12}$$

$$\begin{aligned}\text{MBc: } E[\text{ITTe}] &= [\Phi(\delta_2)+\Phi(\delta_1)-1]\text{ACE} - \sigma(\rho_{B0}-\rho_{B1})[f(\delta_1)-f(\delta_2)] \\ &= [\Phi(\delta_2)+\Phi(\delta_1)-1]\text{CACE}\end{aligned} \tag{13}$$

$$\begin{aligned}\text{MBf: } E[\text{ITTe}] &= [\Phi(\delta_2)+\Phi(\delta_1)-1]\text{ACE} + \sigma(\rho_{B0}-\rho_{B1})[f(\delta_1)-f(\delta_2)] \\ &= [\Phi(\delta_2)+\Phi(\delta_1)-1]\text{CACE}\end{aligned} \tag{14}$$

Clearly, as an estimator of CACE, ITTe is always 'biased towards the null' provided that $p_1 - p_0 = [\Phi(\delta_2)+\Phi(\delta_1)-1] \geqslant 0$. The *attenuation factor* is either $\Phi(\delta)$ or $\Phi(\delta_2)+\Phi(\delta_1)-1$; attenuation occurs whether non-compliance is random or not. However, when viewed as an estimator for ACE, there is a second aspect of bias specific to non-random situations with $\rho_{B0} \neq \rho_{B1}$, and it is no longer clear if the bias is always towards the null; this is examined further later.

The ranking of E[ITTe] relative to E[PPe] and E[ATe] is also not obvious, but examples where $|E[\text{ITTe}]| > |E[\text{PPe}]|$ are easy to find. In Figure 2 it is assumed that ACE is negative and non-compliance follows model MBa. All the statistics have been standardized by division by σ; $\text{ACE}/\sigma = -0.5$. E[PPe]$/\sigma$ is shown when ρ_{B1} is 0.2 or 0.4. If $\rho_{B0} - \rho_{B1} = 0.2$, $|E[\text{ITTe}]| > |E[\text{PPe}]|$ for both values of ρ_{B1} over most of the non-compliance range shown, but when $\rho_{B0} - \rho_{B1} = -0.2$, $|E[\text{ITTe}]| < |E[\text{PPe}]|$.

DOI: 10.1002/sim

Table I. Ranking of E[ATe], E[PPe], ACE, CACE under models with non-compliance in one arm only.

(i) Observed non-compliance in active arm: models MBa and MBd.

MBa $\rho_{B0}>\rho_{B1}$	MBa $\rho_{B0}<\rho_{B1}$	MBa $\rho_{B0}=\rho_{B1}$	MBd $\rho_{B0}>\rho_{B1}$	MBd $\rho_{B0}<\rho_{B1}$	MBd $\rho_{B0}=\rho_{B1}$
E[ATe] E[PPe]	E[ATe] E[PPe] CACE	E[ATe] E[PPe] CACE			
ACE	ACE	ACE=CACE	ACE	ACE	ACE=CACE
CACE			E[PPe] E[ATe]	CACE E[PPe] E[ATe]	 E[PPe] E[ATe]

(ii) Observed non-compliance in control arm: models MBb and MBe.

MBb $\rho_{B0}>\rho_{B1}$	MBb $\rho_{B0}<\rho_{B1}$	MBb $\rho_{B0}=\rho_{B1}$	MBe $\rho_{B0}>\rho_{B1}$	MBe $\rho_{B0}<\rho_{B1}$	MBe $\rho_{B0}=\rho_{B1}$
E[ATe] E[PPe] CACE	E[ATe] E[PPe]	E[ATe] E[PPe]		CACE	
ACE	ACE	ACE=CACE	ACE	ACE	ACE=CACE
	CACE		CACE E[PPe] E[ATe]	 E[PPe] E[ATe]	 E[PPe] E[ATe]

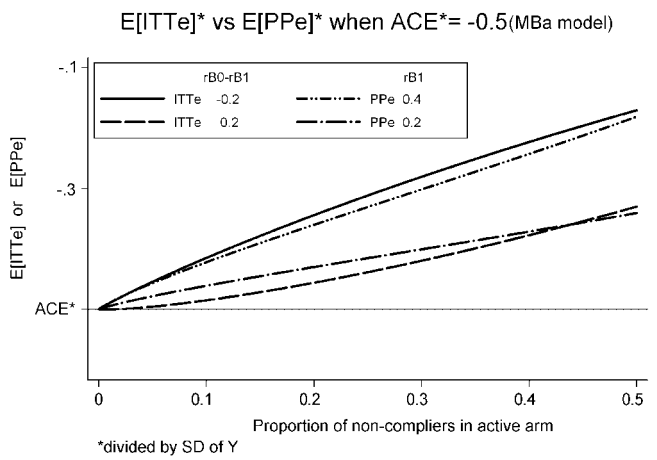

Figure 2. Relationship between ITTe and PPe when non-compliers in active arm are drawn from lower tail of B.

DOI: 10.1002/sim

4.5. Degree of treatment effect heterogeneity

The differences (1–14) are strongly dependent on the degree of heterogeneity, but in opposite ways for CACE and E[ITTe] on the one hand and E[PPe] and E[ATe] on the other. Maximum heterogeneity occurs when $\rho_{01}=0$, which requires either ρ_{B0} or ρ_{B1} to be zero in our models; this maximizes the difference between ACE on the one hand and CACE or E[ITTe] on the other. However, it tends to *minimize* the difference between ACE and both E[PPe] and E[ATe]. The models simulated by B&D assumed $\rho_{01}=0$. If $\rho_{01}=1$, then treatment effects do not vary at all (homogeneous model); this is achieved in our models if either $\rho_{B0}=\rho_{B1}$ or, trivially, both are zero. Under homogeneity, CACE = ACE. However, PPe and ATe are biased for ACE and CACE unless both correlations are zero.

4.6. Which null hypothesis?

It we accept heterogeneity when thinking about non-null treatment effects, how should we frame the null hypothesis? 'Strong' and 'weak' versions are that $Y_i(1)=Y_i(0)\ \forall i$ and $E[Y(1)-Y(0)]=0$, respectively. Our view, elaborated in the Discussion, is that this 'weak' version should not be termed null, but we include it here as it has been considered by others [5, 7].

The strong null is a special case of the homogeneous model; hence, CACE = ACE = 0, but PPe and ATe will be biased away from the null unless $\rho_{B0}=\rho_{B1}=0$. The behaviour under the weak null hypothesis is derived by setting ACE = 0 and $\rho_{01}\neq 1$. CACE can now be positive or negative depending on the subpopulation represented by the compliers and therefore E[IVe] $\neq 0$ and E[ITTe] $\neq 0$. This bias was noted by B&D [7] in simulations of this weak null hypothesis. Liu [8], who sought to defend the IV method against this charge of bias, implied a third null hypothesis, which we also term 'weak': $E[Y(1)-Y(0)|C=1]=0=$CACE. Under this CACE-centred 'null', both IVe and ITTe are no longer biased, but ATe and PPe will be in general.

4.7. Is ITTe always biased towards the null for ACE?

Suppose that there is effect heterogeneity and the target is ACE; also suppose that ACE is negative. For bias only towards the null, we would need:

$$\text{ACE} \leqslant E[\text{ITTe}] \leqslant 0 \tag{15}$$

If $\rho_{B0}>\rho_{B1}$ and non-compliance follows MBa or MBe, then E[ITTe] can be farther from the null than ACE, unless:

$$\frac{|\text{ACE}|}{\sigma} \geqslant (\rho_{B0}-\rho_{B1})\frac{f(\delta)}{1-\Phi(\delta)} \tag{16}$$

If $\rho_{B0}>\rho_{B1}$ and non-compliance follows MBb or MBd then, to ensure E[ITTe]<0, we need:

$$\frac{|\text{ACE}|}{\sigma} \geqslant (\rho_{B0}-\rho_{B1})\frac{f(\delta)}{\Phi(\delta)} \tag{17}$$

Similar constraints can be derived for the two-sided non-compliance models. Figure 3 shows E[ITTe] (standardized by σ) under MBa and MBd models when ACE$/\sigma=-0.33$ and $\rho_{B0}-\rho_{B1}$ is 0.2 or 0.5. In the extreme case $\rho_{B0}-\rho_{B1}=0.5$, $|E[\text{ITTe}]|>|\text{ACE}|$ under MBa throughout the range shown. Also under MBd, if $p_1>0.4$, we have ACE$<0<E$[ITTe]. The same results apply when $\rho_{B0}-\rho_{B1}<0$ except that the implications of MBa and MBd are swopped.

DOI: 10.1002/sim

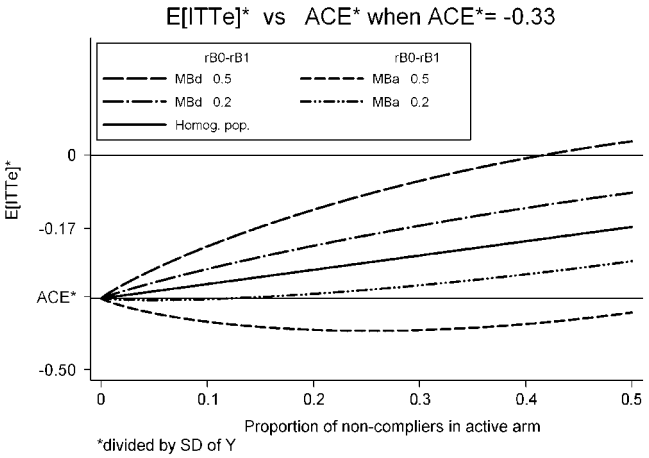

Figure 3. Relationship between ITTe and ACE with non-compliers in active arm: in MBa (MBb) they are drawn from lower (upper) tail of B.

5. EXAMPLE

ODIN was a trial of psychological interventions at the primary care level that aimed to reduce depression symptoms as measured by the Beck Depression Inventory (BDI). There were three treatment arms but, as outcomes were similar for the two types of psychological intervention, they are combined here into a single 'active treatment' arm (arm 1) to be contrasted with 'care as usual' (arm 0). We will ignore data features that might complicate this brief illustration, namely that randomization was stratified by centre and outcome data was missing for around 26 per cent of each arm; only subjects with complete data are included here. A full analysis that addressed problems allied to missing data has been published elsewhere [22]. Patients in arm 0 did not have access to the active treatments. As it was impossible not to comply with 'care as usual', C refers to potential compliance with active treatment only. Because of randomization, the observed compliance proportion in the active arm estimates the population proportion, Φ. Compliance is defined as attendance at all the active treatment sessions; two thirds of arm 1 complied; hence $\delta = 0.431$, $f(\delta)/\Phi(\delta) = 0.545$. The outcome variable was change in BDI at 6 months compared with baseline (BDI6–BDI0). The CIs below were all derived using bootstrap methods.

Table II summarizes the change score and baseline values for three groups: arm 0, arm1 non-compliers and arm 1 compliers. It seems reasonable to assume $\sigma_0 = \sigma_1 = \sigma$; an ANOVA estimate

Table II. ODIN RCT of psychological intervention for depression: summary patient data on Beck Depression Inventory.

		Baseline BDI: BDI0		Change score, $Y = \text{BDI6} - \text{BDI0}$	
	n	Mean	SD	Mean	SD
Arm 0	140	22.0	8.0	−6.8	10.2
Arm 1 non-compliers	59	21.7	8.3	−8.5	8.4
Arm 1 compliers	118	23.2	8.2	−9.8	9.6

DOI: 10.1002/sim

of σ is 9.6. Then ITTe/$\hat{\sigma}=-0.27$ (95 per cent CI: -0.49, -0.05), ATe/$\hat{\sigma}=-0.26$, PPe/$\hat{\sigma}=-0.32$ and IVe/$\hat{\sigma}=-0.40$ (95 per cent CI: -0.73, -0.06). The rank order of these statistics is consistent with MBa where non-compliers are drawn from the left tail of the latent variable B. An estimate of ρ_{B0} based on (5) is $\hat{\rho}_{B0}=[(\text{PPe}-\text{IVe})/\hat{\sigma}]*\{\Phi(\delta)/f(\delta)]=0.16$, (95 per cent CI: -0.10, 0.43). The same estimate can be obtained using either a formula based on the difference ATe$-$PPe or one based on ATe$-$IVe: for example, the counterpart of (8) for unequal n, given in the Appendix, yields $\hat{\rho}_{B0}=[(\text{ATe}-\text{PPe})/\hat{\sigma}]*[n_0+n_1(1-\Phi(\delta))]/[n_1 f(\delta)]=0.16$. We cannot estimate ρ_{B1}.

To make progress in estimating ACE using (1), one might be willing to place a bound on $|\rho_{B0}-\rho_{B1}|$ in the light of $\hat{\rho}_{B0}$ and subject matter knowledge. An alternative route to a bound on $|\rho_{B0}-\rho_{B1}|$, as outlined in Section 4, is to set an upper limit on the maximum degree of treatment heterogeneity, and assume no common causes of $Y(0)$ and $Y(1)$ except B. With $k=1$ for example, implying that the treatment effect distribution has an SD of 9.6, these assumptions lead to $\rho_{01}=0.5$ implying $|\rho_{B0}-\rho_{B1}|\leqslant 0.5$ and hence $|\text{CACE}-\text{ACE}|<0.27\sigma$. However, the assumption of no common causes seems highly questionable in this case. Alternatively, one might try to study correlations with B directly, using baseline predictors of compliance. In principle, B is that function of all such predictors that perfectly discriminates between compliers and non-compliers and is positively associated, if at all, with $Y(0)$ and $Y(1)$. Candidate components of B could be evaluated for their ability to predict C in the active arm, and for their correlations with $Y(0)$ in arm 0 and with $Y(1)$ in compliers. We do not develop this idea further here, but it may be of interest to compare it with other methods utilizing predictors of compliance to improve estimation of efficacy [23–25].

6. DISCUSSION

According to Senn [26], no one really believes that treatment effects are homogeneous (which condition he terms 'additivity') although it is often a necessary assumption in analyses. But acceptance of heterogeneity raises the awkward problem that efficacy estimated for one subgroup—e.g. compliers—may not be the same as in another, even in the same trial. So far as patients entering a trial are 'typical', the overall average measure, ACE, would seem to be of most interest—but perhaps not in situations where non-compliance is due to unacceptable symptoms [19]. Some authors appear to insist on the primacy of ACE [5, 27] and highlight the problem of identifying compliers as they are defined for CACE. Here we assumed that both ACE and CACE are of interest.

In our model, the essential feature of 'non-random' compliance is that $\rho_{B0}-\rho_{B1}\neq 0$ (or more generally $\sigma_0\rho_{B0}-\sigma_1\rho_{B1}\neq 0$). An important insight is that the difference between CACE and ACE is a function of $\rho_{B0}-\rho_{B1}$, suggesting that analyses that try to measure this, or its sign, directly from predictors of compliance—as discussed in Section 5—might be worthwhile. On the other hand, the differences between $E[\text{PPe}]$ (or $E[\text{ATe}]$) and the targets CACE or ACE are *not* a function of the correlation difference and therefore have quite different properties from IVe. We note, however, that the IVe is not without problems: for example, Little *et al.* found that IVe was markedly less efficient, in terms of root mean-squared error, than PPe and ATe when there was a large proportion of non-compliers, although, efficiency was increased in analyses that used predictors of compliance [19]. We also showed that PPe may be more biased towards the null than ITT; this has been noted before [3], although the author did not rule out PP analyses.

DOI: 10.1002/sim

The distinction between CACE and ACE is important even for trialists who are interested in ITT analyses only and consider it as a conservative approach. This is because, when there is non-compliance, the view that ITT is biased towards the null is only supportable if we wish to restrict our interest to the complier sub-population of patients. If the whole population is of interest, this supposition no longer holds in general. We should also note that ITT may not be biased towards the null, even as a measure of CACE, in trials where patients may access non-trial treatments.

Previous observations [7] that E[ITTe] can be biased for ACE under the 'null hypothesis', $E[Y(1)-Y(0)]=0$, were replicated here for our more general models; however in our view, this null hypothesis is not of any great interest nor is $E[Y(1)-Y(0)|C=1]=0$. Such hypotheses suppose that active treatment has a positive effect in some people and negative in others, but the population distribution is such that, on average, they would cancel out. Given that 'cancelling out' depends on a specific population distribution, this hypothesis is very tenuous: even with full compliance we might not expect it to hold in a future population that might have a different mix. There are resonances here with Neyman's hypothesis in an agricultural context that 'on average all treatments are equal', but they could differ in effect on given plots. Senn [26] considered this hypothesis 'not sensible'. We concur in the sense that such hypotheses should not be termed 'null'. Robins and Greenland [5] also noted that the 'ITT null hypothesis' might be rejected even though ACE$=0$, but in the context of an argument favouring ACE estimation rather than any concern about the meaning of such a null hypothesis.

APPENDIX A

All results follow from conditional probability properties of the Normal distribution. For any two bivariate normal variables X and Z with means μ_X, μ_Z, variances σ_X^2, σ_Z^2 and correlation ρ, we can show that

$$E[X|Z>\mu_Z+\delta\sigma_Z]=\mu_X+\rho\sigma_X\frac{f(\delta)}{1-\Phi(\delta)}$$

$$E[X|Z>\mu_Z-\delta\sigma_Z]=\mu_X+\rho\sigma_X\frac{f(\delta)}{\Phi(\delta)}$$

$$E[X|\mu_Z-\delta_1\sigma_Z\leqslant Z\leqslant\mu_Z+\delta_2\sigma_Z]=\mu_X+\rho\sigma_X\frac{f(\delta_1)-f(\delta_2)}{\Phi(\delta_2)+\Phi(\delta_1)-1}$$

where $\Phi(x)$ and $f(x)$ are the cumulative Normal and Normal density functions, respectively. These results are used below always with $Z=B\sim N(0,1)$. In some cases, we set $X=Y(0)$, $\rho=\rho_{B0}$ and in others $X=Y(1)$, $\rho=\rho_{B1}$.

A.1. CACE

Under compliance models MB, we have

$$\text{MBa: } E[Y(1)-Y(0)|C=1]=E[Y(1)-Y(0)|B\geqslant-\delta]=\mu_1-\mu_0-(\rho_{B0}\sigma_0-\rho_{B1}\sigma_1)\frac{f(\delta)}{\Phi(\delta)} \quad \text{(A1)}$$

$$\text{MBb: } E[Y(1)-Y(0)|C=1]=E[Y(1)-Y(0)|B\leqslant+\delta]=\mu_1-\mu_0+(\rho_{B0}\sigma_0-\rho_{B1}\sigma_1)\frac{f(\delta)}{\Phi(\delta)} \quad \text{(A2)}$$

DOI: 10.1002/sim

$$\text{MBc: } E[Y(1)-Y(0)|C=1] = E[Y(1)-Y(0)|-\delta_1 \leqslant B \leqslant +\delta_2]$$
$$= \mu_1-\mu_0-(\rho_{B0}\sigma_0-\rho_{B1}\sigma_1)\frac{f(\delta_1)-f(\delta_2)}{\Phi(\delta_2)+\Phi(\delta_1)-1}$$

The results for MBd and MBe are the same as for MBb and MBa, respectively; for MBf we have

$$\text{MBf: } E[Y(1)-Y(0)|C=1] = E[Y(1)-Y(0)|-\delta_2 \leqslant B \leqslant +\delta_1]$$
$$= \mu_1-\mu_0+(\rho_{B0}\sigma_0-\rho_{B1}\sigma_1)\frac{f(\delta_1)-f(\delta_2)}{\Phi(\delta_2)+\Phi(\delta_1)-1}$$

A.2. PPe

Those who do not comply are dropped from PP analyses: in (MBa) these are subjects with $Y(0) \leqslant \mu_0-\delta\sigma_0$ in the active arm only, in (MBb) subjects with $Y(0) \geqslant \mu_0+\delta\ \delta\sigma_0$ in the control arm only and in (MBc) both of these groups with δ_1 in (MBa) and δ_2 in (MBb). Hence

$$\text{MBa: } E[\text{PPe}] = E[Y(1)|B \geqslant -\delta]-E[Y(0)] = \mu_1-\mu_0+\rho_{B1}\frac{\sigma_1 f(\delta)}{\Phi(\delta)}$$

$$\text{MBb: } E[\text{PPe}] = E[Y(1)]-E[Y(0)|B \leqslant +\delta] = \mu_1-\mu_0+\rho_{B0}\frac{\sigma_0 f(\delta)}{\Phi(\delta)}$$

$$\text{MBc: } E[\text{PPe}] = E[Y(1)|B > -\delta_1]-E[Y(0)|B \leqslant +\delta_2] = \mu_1-\mu_0+\rho_{B1}\frac{\sigma_1 f(\delta_1)}{\Phi(\delta_1)}+\rho_{B0}\frac{\sigma_0 f(\delta_2)}{\Phi(\delta_2)}$$

MBd–MBf results correspond to those of MBa–MBc, respectively, but with a change of sign after $\mu_1-\mu_0$, e.g.:

$$\text{MBd: } E[\text{PPe}] = E[Y(1)|B \leqslant +\delta]-E[Y(0)] = \mu_1-\mu_0-\rho_{B1}\frac{\sigma_1 f(\delta)}{\Phi(\delta)}$$

By substituting for $\mu_1-\mu_0$ from (A1)–(A2), the bias of the PPe estimators for CACE is found, e.g.:

$$\text{under MBa} \quad E[\text{PPe}]-\text{CACE} = \rho_{B0}\frac{\sigma_0 f(\delta)}{\Phi(\delta)}$$

$$\text{under MBb} \quad E[\text{PPe}]-\text{CACE} = \rho_{B1}\frac{\sigma_1 f(\delta)}{\Phi(\delta)}$$

A.3. ATe

For the AT estimators, the observed non-compliers, which are a proportion, $1-\Phi(\delta)$, of their respective arm under (a) and (b), are added to the other arm before analysis. Under MB we have

$$\text{MBa: } E[\text{ATe}] = E[Y(1)|B \geqslant -\delta]-\frac{n_0 E[Y(0)]+n_1[1-\Phi(\delta)]E[Y(0)|B \leqslant -\delta]}{n_0+n_1[1-\Phi(\delta)]}$$
$$= \mu_1-\mu_0+\left[\rho_{B1}\frac{\sigma_1 f(\delta)}{\Phi(\delta)}+\rho_{B0}\frac{n_1\sigma_0 f(\delta)}{n_0+n_1[1-\Phi(\delta)]}\right]$$

DOI: 10.1002/sim

$$\text{MBb: } E[\text{ATe}] = \frac{n_1 E[Y(1)] + n_0[1-\Phi(\delta)]E[Y((1)|B>+\delta]}{n_1 + n_0[1-\Phi(\delta)]} - E[Y(0)|B\leqslant+\delta]$$

$$= \mu_1 - \mu_0 + \left[\rho_{B1}\frac{n_0\sigma_1 f(\delta)}{n_1 + n_0[1-\Phi(\delta)]} + \rho_{B0}\frac{\sigma_0 f(\delta)}{\Phi(\delta)}\right]$$

$$\text{MBc: } E[\text{ATe}] = \frac{n_1\Phi(\delta_1)]E[Y(1)|B>-\delta_1] + n_0[1-\Phi(\delta_2)]E[Y((1)|B>+\delta_2]}{n_1\Phi(\delta_1) + n_0[1-\Phi(\delta_2)]}$$

$$- \frac{n_0\Phi(\delta_2)E[Y(0)|B\leqslant+\delta_2] + n_1[1-\Phi(\delta_1)]E[Y(0)|B<-\delta_1]}{n_0\Phi(\delta_2) + n_1[1-\Phi(\delta_1)]}$$

$$= \mu_1 - \mu_0 + \left[\rho_{B1}\sigma_1\frac{n_1 f(\delta_1) + n_0 f(\delta_2)}{n_1\Phi(\delta_1) + n_0[1-\Phi(\delta_2)]} + \rho_{B0}\sigma_0\frac{n_0 f(\delta_2) + n_1 f(\delta_1)}{n_0\Phi(\delta_2) + n_1[1-\Phi(\delta_1)]}\right]$$

MBd–MBf results correspond to those MBa–MBc, respectively, but with a change of sign after $\mu_1 - \mu_0$.

A.4. ITT

Under MB models

$$\text{MBa: } E[\text{ITT}] = \Phi(\delta)E[Y(1)|B>-\delta] + [1-\Phi(\delta)]E[Y(0)|B\leqslant-\delta] - \mu_0$$

$$= \Phi(\delta)(\mu_1 - \mu_0) - (\rho_{B0}\sigma_0 - \rho_{B1}\sigma_1)f(\delta)$$

$$\text{MBb: } E[\text{ITT}] = \mu_1 - [\Phi(\delta)E[Y(0)|B<+\delta] + [1-\Phi(\delta)]E[Y(1)|B>+\delta]]$$

$$= \Phi(\delta)(\mu_1 - \mu_0) + (\rho_{B0}\sigma_0 - \rho_{B1}\sigma_1)f(\delta)$$

$$\text{MBc: } E[\text{ITT}] = \Phi(\delta_1)E[Y(1)|B>-\delta_1] + [1-\Phi(\delta_1)]E[Y(0)|B\leqslant-\delta_1]$$

$$-[\Phi(\delta_2)E[Y(0)|B<+\delta_2] + [1-\Phi(\delta_2)]E[Y(1)|B>+\delta_2]]$$

$$= [\Phi(\delta_2) + \Phi(\delta_1) - 1](\mu_1 - \mu_0) - (\rho_{B0}\sigma_0 - \rho_{B1}\sigma_1)[f(\delta_1) - f(\delta_2)]$$

The results for MBd and MBe are the same as for MBb and MBa, respectively; for MBf we have

$$\text{MBf: } E[\text{ITT}] = [\Phi(\delta_2) + \Phi(\delta_1) - 1](\mu_1 - \mu_0) + (\rho_{B0}\sigma_0 - \rho_{B1}\sigma_1)[f(\delta_1) - f(\delta_2)]$$

ACKNOWLEDGEMENTS

Thanks to Graham Dunn and Richard Elmsley for access to the ODIN data and comments on the manuscript.

REFERENCES

1. Sheiner LB, Rubin DB. Intention to treat and the goals of clinical trials. *Clinical Pharmacology and Therapeutics* 1995; **57**:6–15.
2. Special Issue: Analysing non-compliance in clinical trials. *Statistics in Medicine* 1998; **17**(3):247–393.
3. White I. Uses and limitations of randomisation-based efficacy estimators. *Statistical Methods in Medical Research* 2005; **14**:327–347.

DOI: 10.1002/sim

4. Angrist J, Imbens G, Rubin D. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996; **91**:444–455.
5. Robins J, Greenland S. Identification of causal effects using instrumental variables: comment. *Journal of the American Statistical Association* 1996; **91**:456–458.
6. Angrist J, Imbens G, Rubin D. Identification of causal effects using instrumental variables: rejoinder. *Journal of the American Statistical Association* 1996; **91**:468–472.
7. Bang H, Davis C. On estimating treatment effects under non-compliance in clinical trials: are Intention to Treat or instrumental variable analyses perfect solutions? *Statistics in Medicine* 2007; **26**:954–964.
8. Lui KJ. Comments on 'On estimating treatment effects under non-compliance in clinical trials: are Intention to Treat or instrumental variable analyses perfect solutions?'. *Statistics in Medicine* 2009; **28**:531–534.
9. Manski C. Nonparametric bounds on treatment effects. *American Economic Reviews*, *Papers and Proceedings* 1990; **80**:319–323.
10. Balke A, Pearl J. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association* 1997; **92**:1171–1176.
11. Robins J. The analysis of randomized and non-randomised AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Services Research Methodology*: *A Focus on AIDS*, Sechrest L, Freeman H, Mulley A (eds). US Public Health Service: Washington, DC, 1989; 113–159.
12. Harris K, Remler D. Who is the marginal patient? Understanding instrumental variables estimates of treatment effects. *Health Services Research* 1998; **33**(5):1337–1360.
13. Lachin J. Statistical considerations in the intention-to-treat principle. *Controlled Clinical Trials* 2000; **21**:167–189.
14. Freedman D. Statistical models for causation. What inferential leverage do they provide? *Evaluation Review* 2006; **30**(6):691–713.
15. European Medicines Agency ICH Topic E9 statistical principles for clinical trials, step 5. Note for guidance on statistical principles for clinical trials. *CPMP/ICH/363/96*, September 1998. Available from: http://www.emea.europa.eu/pdfs/human/ich/036396en.pdf. Date of access 17/12/2008.
16. Jones B, Jarvis P, Jewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *British Medical Journal* 1996; **313**:36–39.
17. Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *Journal of Clinical Epidemiology* 2007; **60**:663–669.
18. Wright C, Sim J. Intention-to-treat to data from randomised controlled trials: a sensitivity analysis. *Journal of Clinical Epidemiology* 2003; **56**:833–842.
19. Little RJ, Long Q, Lin X. A comparison of methods for estimating the causal effect of a treatment in randomised clinical trials subject to non-compliance. *Biometrics* 2008; published online. DOI:10.1111/j.1541-0420.2008.01066.x.
20. Rubin D. Estimating causal effects of treatments in randomised and non-randomised studies. *Journal of Clinical Epidemiology* 1974; **66**:688–701.
21. Heckman J, Urzua S, Vytlacil E. Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics* 2006; **88**(3):389–432.
22. Dunn G, Maracy M, Tomenson B. Estimating treatment effects from randomised clinical trials with non-compliance and loss to follow-up: the role of instrumental variables. *Statistical Methods in Medical Research* 2005; **14**:369–395.
23. Joffe M, Have T, Brensinger C. The compliance score as a regressor in randomised clinical trials. *Biostatistics* 2003; **4**(3):327–340.
24. Follmann D. On the effect of treatment among would-be compliers: an analysis of the Multiple Risk Factor Intervention Trial. *Journal of the American Statistical Association* 2000; **95**:1101–1109.
25. Jo B. Statistical power in randomised intervention studies with non-compliance. *Psychological Methods* 2002; **7**:178–193.
26. Senn S. Controversies concerning randomisation and additivity in clinical trials. *Statistics in Medicine* 2004; **23**:3729–3753.
27. Heckman J. Identification of causal effects using instrumental variables: comment. *Journal of the American Statistical Association* 1996; **91**:459–462.

DOI: 10.1002/sim