Estimating causal effects using prior information on nontrial treatments

Simon J Bond[a,b,c] and Ian R White[a]

Background   Departures from randomized treatments complicate the analysis of many randomized controlled trials. Intention-to-treat analysis estimates the effect of *being allocated to* treatment. It is now possible to estimate the effect of *receiving* treatment without assuming comparability of groups defined by actual treatment. However, the methodology is largely confined to trials where the only treatment changes were switches to other trial treatments.

Purpose   To propose a method for comparing the effects of receiving trial treatments in an active-controlled clinical trial where some participants received nontrial treatments and others received no treatment at all, and to illustrate the method in the PENTA 5 trial in HIV-infected children.

Methods   We combine the instrumental variables approach, which forms unbiased estimating equations based on the randomization but does not fully identify the contrasts of trial treatment effects, with prior information about the distribution of possible effects of nontrial treatments and of one trial treatment; we do not need to use prior information about the comparisons of trial treatments. Prior information in PENTA 5 was elicited from the investigators.

Results   In PENTA 5, the prior information suggested that all treatments were beneficial, but there was uncertainty about the degree of benefit. Allowing for this prior information changed point estimates and increased standard errors compared with ignoring nontrial treatments.

Limitations   The method depends on the correct specification of the causal effect of treatment: in PENTA 5, this assumed a linear effect of dose and no interactions between treatments. This specification is hard to check from the data but can be explored in sensitivity analyses. Prior information would be better derived from the literature whenever possible.

Conclusions   The use of partial prior information gives a way to adjust for complex patterns of departures from randomized treatments. It should be useful in all trials where nontrial treatments are used and in active-controlled trials where trial treatments are not universally used.   *Clinical Trials* 2010; **7**: 664–676. http://ctj.sagepub.com

Introduction

Many forms of departure from randomized treatment occur in clinical trials: nonreceipt of randomized treatment (sometimes termed non-compliance), receipt of the treatment allocated to a different trial arm (sometimes termed contamination), or receipt of a nontrial treatment, defined as a treatment not randomly allocated in the trial. Intention-to-treat (ITT) analysis is accepted as a valid way to explore the effect of *allocating* treatment [1]. However, there has long been interest in estimating the causal effect of *receiving* treatment [2]. Modern statistical methods

---

[a]MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK, [b]Mundipharma Research, Cambridge, UK, [c]Cambridge Cancer Trials Centre and Cambridge Biomedical Research, Cambridge, UK
Author for correspondence: Ian White, MRC Biostatistics Unit, Forvie Site, Robinson Way, Cambridge, CB2 0SR, UK.
E-mail: ian.white@mrc-bsu.cam.ac.uk

are able to base such estimation on comparisons of randomized groups [3], unlike popular methods such as per-protocol analysis which invalidly compare subgroups with different receipt of treatment [4].

Much statistical literature assumes that participants can only switch to receive the treatment allocated to a different trial arm, a situation with the convenient property that the ITT analysis also tests the null hypothesis of equivalence of the randomized treatments. This article considers the more difficult, but common, case where participants can receive nontrial treatments. In this case, the ITT analysis does not test the equivalence of the randomized treatments: for example, one treatment could appear better simply because participants receiving that treatment were more likely to change to an effective nontrial treatment.

The difficulty introduced by participants receiving nontrial treatments is that the effects of these treatments must be included in a statistical model, but cannot easily be estimated from the trial data. Past work has addressed this problem by assuming that there are no unmeasured confounders [5,6] or by making strong distributional assumptions [7,8]. Our approach avoids such assumptions and instead uses information external to the trial to place informative prior distributions on the effects on nontrial treatments. To do this, we introduce a hybrid of Bayesian inference [9,10] and instrumental variables methods [11].

A particular application of our proposed methodology is to equivalence and noninferiority trials in which some participants receive no treatment, since receipt of no treatment introduces the same issues as receipt of nontrial treatments. ITT analysis is generally held to be anti-conservative for equivalence and noninferiority trials [12]. Per-protocol analysis is commonly done [13], but better methods are needed [14].

In this article, we first present the hybrid approach in general. We then show using a simple example how comparing the causal effects of trial treatments depend on the effects of the nontrial treatments, and we illustrate the hybrid approach. We next present an application to PENTA 5, a three-arm comparative trial in pediatric HIV patients in which some children received no treatment and others received nontrial treatments. We describe how we elicited and used expert priors, the results with the various prior distributions that were considered, and the implications. Finally, the discussion places this work in the wider context of the noncompliance literature.

Proposed method

Basic assumptions

In Rubin's causal model [15], participant i has a set of counterfactual outcomes $Y_i(\mathcal{D})$, where $\mathcal{D}$ represents a potential set of treatments, or drug dosages. For each participant only one of this set of outcomes is observed, $Y_i(D_i)$. Two assumptions are implicit in this notation. The Stable Unit Treatment Value Assumption (SUTVA) [15] states that the set of counterfactual outcomes for a particular participant is a function of *that* participant's potential treatments, and is not in any way influenced by *other* participants' potential treatments. The *exclusion restriction* assumption states that randomization has no *direct* effect on the outcome, and may only have an indirect effect through its effect on treatment actually received. Thus if a participant *would* receive an identical treatment regardless of their randomized treatment, then their counterfactual outcomes would be identical in all such arms. All potential outcomes are assumed to be independent of randomization.

Model

For the i-th participant, let R_i be their randomized group, taking values $1, 2, \ldots, g$; let D_i be a p-dimensional vector of their actual amount of the trial and nontrial treatments (or dosages, for drug treatments); let Y_i be their observed outcome; and let $Y_i(0)$ be their counterfactual untreated outcome which would have been observed if no treatment had been received.

We define a causal model relating the observed outcome for the i-th participant to the same participant's counterfactual untreated outcome [16]:

$$Y_i = Y_i(0) + D_i'\alpha + \epsilon_i, \qquad (1)$$

where α is a p-dimensional set of causal parameters for the trial and nontrial treatments. This model assumes that the treatments have an additive, linear effect proportional to the dose taken. For any individuals with $D_i = 0$, consistency between Y_i and $Y_i(0)$ requires $\epsilon_i = 0$. We also assume

$$Y_i(0) = X_i'\gamma + \delta_i, \qquad (2)$$

where X_i is a set of baseline covariates. Random variation is captured by the within-individual error ϵ_i with $\mathbf{E}[\epsilon_i|X_i, R_i] = 0$ and the between-individual error δ_i with $\mathbf{E}[\delta_i|X_i, R_i] = 0$.

The parameter α requires careful interpretation. Making the stronger assumption $\mathrm{E}[\epsilon_i|D_i, X_i, R_i] = 0$ allows us to interpret $d'\alpha$ as the average causal effect of a dose d *among those who receive dose* d. It cannot be interpreted as the average causal effect of dose d *among all individuals*, unless we also assume that treatment assignment is independent of treatment benefit [17].

Ordinary least-squares estimation

One way to estimate α is to regress Y_i on D_i. Unfortunately, this is only valid if $Y_i(0)$ is uncorrelated with D_i, which is usually unrealistic: participants with a good prognosis usually show different patterns of D to those with a poor prognosis. Sometimes, confounding variables X can be measured and included in the analysis in order to motivate a 'no unmeasured confounders' assumption that $Y_i(0)$ is uncorrelated with D_i given X, but this is rarely plausible. We do not consider this method further.

Instrumental variables estimation

A better approach to estimating the parameters uses the idea of instrumental variables (IV). Combining Equations (1) and (2) and taking the expectation conditional on X and R gives

$$\mathrm{E}[Y_i|X_i, R_i] = X_i'\gamma + \mathrm{E}[D_i|X_i, R_i]'\alpha. \tag{3}$$

We estimate α in two stages. First, we fit a linear regression for the d-th treatment dose,

$$D_{di} = X_i'\beta_{d1} + R_i'\beta_{d2} + e_{Ddi} \tag{4}$$

for each $d = 1$ to p, where R_i is a vector of dummy variables for randomized group. We assemble the p fitted values into a vector $\widehat{D}_i = \widehat{D}_i(X_i, R_i)$. Then we estimate α in the linear regression

$$Y_i = X_i'\gamma + \widehat{D}_i'\alpha + e_{Yi}. \tag{5}$$

Identification

Regression (5) is only identified if the number of unknown treatment effects p is less than the number of arms g. For example, in a two-arm trial comparing two treatments, in which some individuals receive no treatment, Equation (5) does not identify the two unknown treatment effects, since there is only one contrast between trial arms. We will consider this example further below.

There are various ways to deal with nonidentification. First, we could expand the causal model to a *full probability model*, modeling the association between $Y(0)$ and D_i, and estimating the model using maximum likelihood [18]. However, this approach can be sensitive to model mis-specification [19].

Second, we could include interactions between X_i and R_i in model (4), as is done in structural mean models [20,21]. Including such interactions typically identifies the treatment effects α, but such identification is highly dependent on the assumption that the causal effect does not vary with X [16]. In this article we do not include such interactions, but they are explored separately [22].

Third, we could perform a sensitivity analysis in which we separate the model parameters into two groups: the 'protocol effects' α_p which would have been estimated if there had been perfect compliance, and the 'nonprotocol effects' α_n which would have been irrelevant if there had been perfect compliance. In trials comparing active treatments, the protocol effects would be contrasts of active treatments, while the nonprotocol effects would be the effect of one trial treatment and the effects of any nontrial treatments used (so that they are not uniquely defined: see 'Obtaining prior information') later. The sensitivity analysis would then estimate the protocol effects over a range of plausible values of the nonprotocol effects.

Hybrid estimation

Building on the idea of sensitivity analysis, our proposal in this article is to combine IV estimation with prior information about the nonprotocol effects in order to identify the model. The next section will discuss how the prior information may be obtained. To implement the method, we use a Bayesian approach. We first fit the linear regression (4) and sample from the posterior distribution of parameters β. For each sampled value, we compute $\widehat{D}_i(X_i, R_i)$. We then consider the linear regression (5) with prior $p(\alpha)$ and sample from the posterior distribution of α and γ.

This procedure is conveniently implemented in WinBUGS [10]. It could be considered a Bayesian procedure using a partial likelihood that does not consider the relationship between D, δ, and ϵ. Alternatively, it could be considered a probabilistic sensitivity analysis for IV estimation where any unidentified parameters may be imputed following an informative prior distribution [23]. Its advantages are enabling the use of prior identifying

information and allowing for uncertainty in the parameters of $E[D|X, R]$.

Obtaining prior information

The nonprotocol effects are commonly effects of widely used treatments, so results of relevant randomized trials or meta-analyses of such trials would ideally be available as prior information. In some cases, however, trials or meta-analyses are inadequate, and it is preferable to elicit prior information from experts [24].

In practice, one may elicit a prior distribution for all of α rather than just for the nonprotocol parameters; this was done in the case study below. A fully Bayesian analysis would use this full prior $p(\alpha)$. However, we argue against using prior beliefs about the protocol effects, because we would not usually use this information in a clinical trial without noncompliance, and because it is not needed in order to adjust for noncompliance. We prefer to use prior information only where it is needed to achieve identifiability.

There is usually more than one way to achieve identifiability, because there is more than one reasonable way to define the nonprotocol parameters. We propose using the full expert prior to define the nonprotocol parameters α_n as a linear combination of parameters in α that are uncorrelated with α_p in the prior distribution. We then modify the full prior distribution by giving a large value to the prior variance of the protocol effects: that is, we use the expert prior for α_n and a vague prior for α_p. This method has the desirable property that the same prior is obtained, however, α_p is defined. In the case of complete compliance, or just switching between trial treatments, the prior will become redundant, as would be desired. Details of how to compute such a set of constrasts are given in appendix D.

Simple example

We now consider a two-arm trial comparing experimental with standard treatment, where participants may receive either treatment, both treatments or neither treatment, but receipt of any one treatment is all-or-nothing. As before, R_i indicates whether participant i was randomized to treatment 1 or 2, Y_i is their actual (quantitative) outcome, and $Y_i(0)$ is their counterfactual untreated outcome. In this setting, treatment for participant i is summarized by two binary variables D_{di} indicating whether or not they took

treatment $d = 1$, 2. The causal model (1) then becomes

$$Y_i = Y_i(0) + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \epsilon_i \qquad (6)$$

where, as before, ϵ_i has mean zero, is uncorrelated with randomized group and actual treatment and equals 0 if $D_{1i} = D_{2i} = 0$.

The target of inference is the 'protocol effect' $\alpha_p = \alpha_1 - \alpha_2$. Strictly, this is the difference between the average effect of treatment 1 in those who receive treatment 1 and the average effect of treatment 2 in those who receive treatment 2. It seems reasonable to believe that the average effect of treatment 1 in those who received it is unaffected by whether they *would have received* treatment 2 if allocated to it. Therefore the protocol effect may be interpreted as the ITT difference in perfect compliers [22].

We initially consider the case where treatment 1 is new and treatment 2 is a standard treatment, so we are likely to have more prior knowledge about α_2 than about α_1, and we define the nonprotocol effect as $\alpha_n = \alpha_2$.

Estimation by instrumental variables

Randomization ensures that the distributions of the error term ϵ_i and the untreated outcome $Y_i(0)$ are identical in both arms, suggesting the unbiased estimating equation

$$\frac{1}{n_1} \sum_{i:R_i=1} (Y_i - \alpha_1 D_{1i} - \alpha_2 D_{2i})$$
$$= \frac{1}{n_2} \sum_{i:R_i=2} (Y_i - \alpha_1 D_{1i} - \alpha_2 D_{2i}) \qquad (7)$$

where n_1 and n_2 are the sample sizes in the two arms. We write this as

$$\bar{Y}_1 - \alpha_1 \bar{D}_{11} - \alpha_2 \bar{D}_{21} = \bar{Y}_2 - \alpha_1 \bar{D}_{12} - \alpha_2 \bar{D}_{22}, \qquad (8)$$

where $\bar{D}_{dr}$ and $\bar{Y}_r$ are the mean of D_{di} and Y_i in arm r, and hence

$$\alpha_p = \alpha_1 - \alpha_2 = \frac{\bar{Y}_1 - \bar{Y}_2 + \alpha_2\{\bar{D}_{22} + \bar{D}_{12} - \bar{D}_{11} - \bar{D}_{21}\}}{\bar{D}_{11} - \bar{D}_{12}} \qquad (9)$$

Equation (9) shows that we cannot in general estimate the protocol effect $\alpha_1 - \alpha_2$ without knowing the nonprotocol effect α_2. The only exception is if $\bar{D}_{22} + \bar{D}_{12} - \bar{D}_{11} - \bar{D}_{21} = 0$, which occurs in the special case when the participants can only switch treatments. In this case the estimate of $\alpha_1 - \alpha_2$ is well defined and is the standard IV estimate

$(\bar{Y}_1 - \bar{Y}_2)/(\bar{D}_{11} - \bar{D}_{12})$. In the case of perfect compliance, $\bar{D}_{11} - \bar{D}_{12} = 1$, so the estimate equals the standard ITT estimate $\bar{Y}_1 - \bar{Y}_2$.

Now suppose we have external knowledge about α_2, the causal effect of the standard treatment. Expressing this knowledge as a probability distribution with mean m_{α_2} and variance $s_{\alpha_2}^2$, our best estimate is

$$\hat{\alpha}_p = \frac{\bar{Y}_1 - \bar{Y}_2 + m_{\alpha_2}\{\bar{D}_{22} + \bar{D}_{12} - \bar{D}_{11} - \bar{D}_{21}\}}{\bar{D}_{11} - \bar{D}_{12}} \quad (10)$$

with approximate variance

$$\text{var}(\hat{\alpha}_p)$$
$$= \frac{\sigma^2(1/n_1 + 1/n_2) + s_{\alpha_2}^2\{\bar{D}_{22} + \bar{D}_{12} - \bar{D}_{11} - \bar{D}_{21}\}^2}{(\bar{D}_{11} - \bar{D}_{12})^2}$$
$$(11)$$

where σ is the standard deviation of Y in each arm. The two terms in the numerator of (11) allow for uncertainty in the data about $\bar{Y}_1 - \bar{Y}_2$ and prior uncertainty about α_2, respectively. (We ignore for illustrative purposes the uncertainty in the $\bar{D}_{dr}$, which is typically much less than the uncertainty in the $\bar{Y}_r$.) The role of the external knowledge is typically limited, because the term $\{\bar{D}_{22} + \bar{D}_{12} - \bar{D}_{11} - \bar{D}_{21}\}$ is typically small: it represents the difference in treatment in arm 2 compared with arm 1, summed over treatments.

We have so far assumed that treatment 1 is new. If instead the trial is a head-to-head comparison of two existing treatments, one might use prior knowledge about $\bar{\alpha} = (\alpha_1 + \alpha_2)/2$. This suggests replacing Equation (9) with

$$\alpha_p = \alpha_1 - \alpha_2$$
$$= \frac{\bar{Y}_1 - \bar{Y}_2 + \bar{\alpha}\{\bar{D}_{22} + \bar{D}_{12} - \bar{D}_{11} - \bar{D}_{21}\}}{\frac{1}{2}(\bar{D}_{22} - \bar{D}_{12} + \bar{D}_{11} - \bar{D}_{21})}. \quad (12)$$

Given a full expert prior $p(\alpha_1, \alpha_2)$, we could also define α_n to be uncorrelated with α_p in the prior, as described above.

Numerical example

We illustrate these ideas using a fictitious data set with imperfect compliance (Table 1). If α_2 were known then we could use Equation (9) to show $\hat{\alpha}_p = \frac{1 - 0.2\alpha_2}{0.8}$. Instead, we use various priors for $\alpha_n = \alpha_2$, using Equations (10) and (11). The first four rows of Table 2 have priors centered at $\alpha_2 = 0$. They all have point estimate $\hat{\alpha}_p = 1.25$, but the standard error of $\hat{\alpha}_p$ increases with uncertainty

about α_2 (although this standard error is much smaller than the prior uncertainty about α_2). The next four rows of Table 2 change the prior mean from 0 to 1, which reduces the posterior mean from 1.25 to 1 – a smaller change, but one that could still be practically important.

Example: PENTA 5 trial

Trial design

The PENTA 5 trial [25] compared the effectiveness of three combination treatments for pediatric HIV-1 infection: lamivudine (3TC) + abacavir (ABC), zidovudine (ZDV) + 3TC, and ZDV + ABC. Of the 128 children randomized, 55 who were initially symptom-free were additionally randomized to receive nelfinavir, a protease inhibitor (PI), or placebo in an incomplete factorial design: symptomatic children all received PI. In this analysis we include all 128 children but consider PI as a nontrial drug, ignoring the fact that it was partly randomized.

The children returned to the clinic at 4-week intervals up to 24 weeks and at 12-week intervals thereafter up to 224 weeks. The primary endpoint was the log concentration of HIV RNA in plasma at 24 and 48 weeks. Here we focus on the 24-week outcome, defined as the outcome observed closest to 24 weeks within the 22–30 week window.

Table 1 Data for simple example

	Arm 1	Arm 2
Mean outcome, $\bar{Y}_r$	3	2
Outcome SD, σ	1	1
Sample size, n	100	100
Mean use of treatment 1, $\bar{D}_{r1}$	0.8	0
Mean use of treatment 2, $\bar{D}_{r2}$	0	0.6

Table 2 Results for simple example

Prior $p(\alpha_2)$	Posterior for $\alpha_p = \alpha_1 - \alpha_2$	
	Mean	Standard error
$N(0, 0^2)$	1.25	0.18
$N(0, 0.5^2)$	1.25	0.22
$N(0, 1^2)$	1.25	0.31
$N(0, 2^2)$	1.25	0.53
$N(1, 0^2)$	1.00	0.18
$N(1, 0.5^2)$	1.00	0.22
$N(1, 1^2)$	1.00	0.31
$N(1, 2^2)$	1.00	0.53

Baseline covariates included sex, ethnic origin, age, CDC disease stage, CD4 cell count, and height-for-age Z-scores.

Observed treatment

The trial drugs ZDV, 3TC and ABC, and the nontrial PI were taken at many different doses. Table 3 shows a 'recommended dose' for each drug (the protocol dose for the trial drugs and current best practice for the nontrial drug), dependent on a child's weight or estimated body surface area. For each drug, we defined the 'dose fraction' for each child at each clinic visit as the fraction of the 'recommended dose' taken, except that doses above the maximum adult dose had a dose fraction of 1 [16]. The 'standardized dose' was defined as the dose fraction if a drug had been taken without interruption; after interruption, the effect of the drug may be diminished by possible acquisition of viral resistance, and the dose fraction was multiplied by the 're-start factor' given in Table 3.

Table 3 PENTA 5 trial: recommended and maximum drug doses and re-start factors

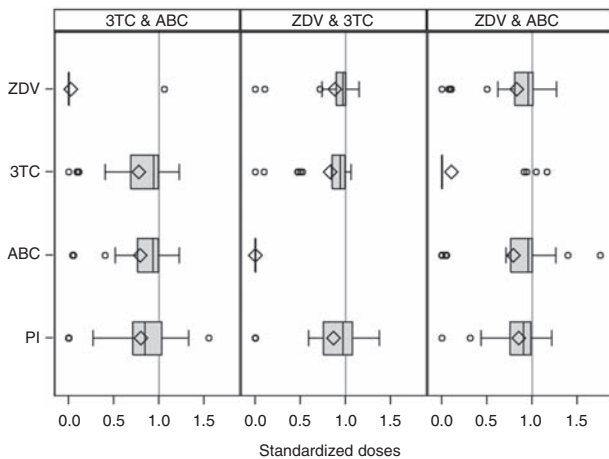| Type	Drug	Recommended dose (mg)	Maximum (mg)	Re-start factor (%)
Trial	ZDV	360 /m^2/day	600	75
	3TC	8 /kg/day	300	50
	ABC	16 /kg/day	600	50
Nontrial	PI	90 /kg/day	2500	75

Figure 1 PENTA 5 trial: standardized drug dosages at 24 weeks

D_{1i}, D_{2i}, D_{3i}, and D_{4i} were defined as the standardized doses for ZDV, 3TC, ABC, and PI, respectively for child i at 24 weeks. A box and whisker plot summarizing the standardized doses, by randomized arm, is given in Figure 1.

The focus of the trial is to estimate differences between 3TC & ABC, ZDV & 3TC, and ZDV & ABC [25]. Assuming the drugs have an additive effect, the contrasts of interest are then represented by $\alpha_2 - \alpha_1$ (3TC & ABC vs. ZDV & ABC) and $\alpha_2 - \alpha_3$ (ZDV & 3TC vs. ZDV & ABC). Thus the protocol effects are $\alpha_p = (\alpha_2 - \alpha_1, \alpha_2 - \alpha_3)$. We consider how to assess sensitivity to the modeling assumptions in the discussion. The nonprotocol effects are taken as $\alpha_n = (\alpha_1 + \alpha_3, \alpha_4)$, representing the causal effect of ZDV & ABC versus no drugs and the causal effect of the nontrial PI drugs. We emphasize that α_p is of primary clinical interest; α_n is only important because it is needed in order to estimate the protocol parameters.

Expert prior

To obtain expert opinion on drug effects, five clinicians active in the field of HIV medicine and HIV clinical trials were sent a questionnaire describing 22 hypothetical randomized controlled trials (Appendices A and B). The trials compared the two- and three-drug combinations of all trial and nontrial drugs actually used in PENTA 5, although several of these (the classes NRTI and nNRTI in the appendices) were not in use at 24 weeks. For each trial, the questionnaire elicited the median, 75th and 87.5th quantile of the expert's prior distribution for the true difference in log HIV RNA between the trial arms. The elicited data were converted into a pooled multivariate Normal prior distribution with the parameters given in Table 4.

Table 4 PENTA 5 trial: means, standard deviations and correlation matrix of the expert prior distribution for the four treatment effects

	ZDV (α_1)	3TC (α_2)	ABC (α_3)	PI (α_4)
Mean	−0.408	−0.576	−0.457	−1.032
SD	0.418	0.478	0.443	0.525
Correlation				
ZDV	1	0.41	0.32	0.45
3TC		1	0.38	0.42
ABC			1	0.30
PI				1

Analysis model

We modeled the dose of the d-th drug for child i at 24 weeks using

$$D_{di} = \beta_{d0} + \beta_{d1}R_{i1} + \beta_{d2}R_{i2} + \beta_{d3}Y_{i0} + \epsilon_{Ddi} \quad (13)$$

where R_{i1} and R_{i2} are indicator variables for the 3TC & ABC and ZDV & 3TC arms, and Y_{i0} is the baseline level of log HIV RNA. The choice of covariates was based on statistical significance.

Our model for the primary outcome was

$$Y_i = \sum_{d=1}^{4} \alpha_d \hat{D}_{dij} + \gamma_0 + \gamma_1 Y_{i0} + \epsilon_{Yi}. \quad (14)$$

The error terms, ϵ_{Ddi} and ϵ_{Yi} were assumed to be independent, normally distributed, with zero mean and variance σ_d^2 and σ_Y^2, respectively.

As described more fully in [16], some outcome values were only reported as being below a cutoff. The MCMC simulation approach easily copes with such censored data by imputing the censored values under the assumption that the underlying distribution is Normal. WinBUGS code is given in Appendix C.

For comparative purposes, we fitted an intention-to-treat (ITT) model which made no use of the observed doses. The model fitted was similar to Equation (14) but with the four $\hat{D}_{dij}$ terms replaced with two indicator variables for the 3TC & ABC and ZDV & 3TC arms of the trial; the baseline log HIV RNA count term was retained. A classical Tobit regression analysis was used to account for the censored observations, replicating the approach used in [25].

Results

We considered five different choices of priors, the last three of which were described above under 'Obtaining prior information':

(1) An uninformative prior, $N(0, 2^2)$, on the two nonprotocol effects. The value of 2 was chosen as a very large value from the clinical context;
(2) A 'naïve' prior that assumes the value of the nonprotocol parameters is zero;
(3) Partial expert (method 1): the experts' full prior distribution, with the marginal variance on the two protocol effects changed to a large value, 1000, and with the prior for the two nonprotocol effects unchanged;
(4) Partial expert (method 2): as method 1, but with the two nonprotocol parameters redefined

to be independent of the protocol parameters in the prior;
(5) The experts' full prior distribution on both protocol and nonprotocol parameters.

Apart from these different specifications of the informative prior, broadly noninformative priors were used for the remaining parameters: the coefficients in the D models were $N(0, 100)$; precisions for all the error terms were $\Gamma(0.01, 0.01)$; α_p (except in the model using the full prior) and γ were $N(0, 1000)$.

The resulting confidence intervals for the two protocol parameters are shown in Figure 2. With the uninformative prior, the data are fairly uninformative on the causal effect of the trial drugs. The naïve prior gives much more certain results, but makes the implausible assumption that nontrial treatments have no effect. The two partial expert prior methods agree to a reasonable extent in both location and width; we prefer method 2, since it is invariant to re-parameterization and has a clear theoretical interpretation. The full expert prior gives results that are similar to the partial expert prior but closer to the null, reflecting the experts' prior beliefs being centered near the null. Recall that these parameters require careful interpretation. The ITT analysis, also shown in Figure 2, estimates the treatment effect of 3TC & ABC to be closer to the null than all the other models, with a similar size confidence interval to the expert prior analysis; the ITT treatment effect of ZDV & 3TC is roughly comparable to the two expert prior analysis.
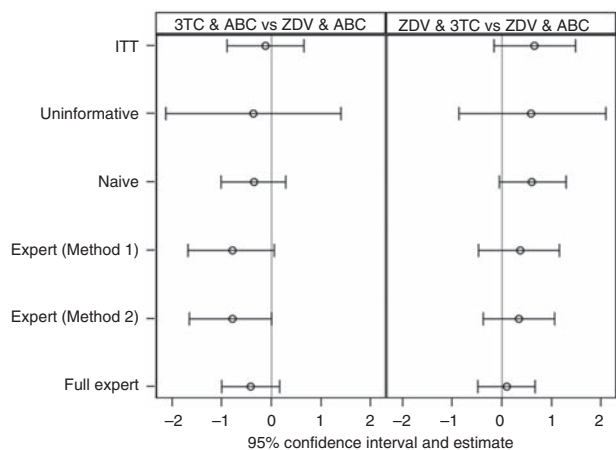

Figure 2 PENTA 5 trial: estimated parameters under intention-to-treat (ITT) analysis and under causal analysis with different priors

The qualitative conclusion is that 3TC & ABC is the superior drug combination, but the degree of its superiority depends on our opinion of the effect of the other drugs.

Discussion

We have proposed a way to estimate causal differences between treatments in a randomized trial with a complex pattern of nonreceipt of randomized treatments and receipt of nontrial treatments. Here we compare our approach with some alternative methods that have been proposed.

The largest body of literature concerns causal analyses when the only treatment change is switching between the trial treatments. The intuitive presentation of Sommer and Zeger [26] has been formalized by Angrist *et al.* [27] and by the notion of principal strata [28], and widely extended [18,29–31]. These techniques estimate a single treatment contrast from a single randomization and do not extend to cases where some participants take nontrial treatments, or to comparative trials where some participants take no treatments, because in these settings it is necessary to estimate more parameters without having more randomizations.

Some attempts have been made to handle nontrial treatments. Robins and Greenland considered a trial of low versus high dose of AZT in AIDS patients where the ITT analysis unexpectedly found that the low dose improved survival [5], possibly due to imbalance in the receipt of a nontrial treatment (prophylaxis against *Pneumocystis carinii* pneumonia). The authors estimated a structural nested failure time model involving the causal effects of the two treatments in two ways. First, they used two different weighted log-rank tests: since there were more parameters than randomized arms, this method gave very imprecise inferences, and depended strongly on distributional assumptions. Second, they estimated the effect of the nontrial treatment under the assumption that conditioning on recorded clinical variables was sufficient to make the decision to switch treatments independent of potential future outcomes. This gave more precise inferences, but the 'no unmeasured confounders' assumption is highly questionable in this and other settings. Robins introduced a wide range of methods for equivalence trials based on the assumption of no unmeasured confounders [6].

An alternative approach by Walter *et al.* [7] extended the approach of Sommer and Zeger [26] to a comparative trial with a binary outcome where participants might receive no treatment. They defined five principal strata [28] by cross-classifying the treatment a participant would receive if randomized to A with the treatment they would receive if randomized to B. To identify the model, they made limited assumptions about comparability of principal strata, either constraining three relative risks to be equal or constraining two pairs of probabilities to be equal: the estimates may be sensitive to such assumptions, and external information is needed to justify them.

Roy *et al.* [8] identified causal effects in different principal strata without assuming either comparability between groups as treated or no unmeasured confounders [8]. However, their method assumed that potential outcomes were independent of covariates within principal strata, and this questionable assumption may have been the key to model identification.

Our approach avoids making strong and typically unjustified assumptions, either comparability between groups as treated, possibly after covariate adjustment, or modeling assumptions. It accepts the likely nonidentifiability of the contrasts of interest from the observed data, and instead uses contextual or prior knowledge about the treatments.

Is such prior knowledge likely to be available and reliable? In a comparative trial with some individuals receiving no treatment, we require prior knowledge about one of the treatment effects, or perhaps about some linear combination of the two treatment effects. This is very likely to be available from the literature, so we would often be able to use published trials or (ideally) meta-analyses to provide our prior knowledge. Even rather imprecise prior information is likely to be useful, and prior imprecision is unlikely to carry over substantially into the estimates of treatment contrasts unless nonreceipt of treatment is widespread and unbalanced.

Greater difficulties may arise in the case of nontrial treatments, which may be new, untested or inadequately tested treatments. In this case expert opinions may be required. The analysis of the PENTA 5 trial presented here was part of a larger investigation of treatment changes throughout follow-up, which required priors for seven drugs from four classes. Even assuming common drug effects within classes, this was considered to be more information than could be obtained from the literature, since drug effects are likely to differ between children and adults, and few HIV trials have been done in children. Another difficulty with the use of expert prior information is deciding who should provide the prior. Ideally the prior would come from independent experts, but it will often more practically come from the trial

investigators [32]. Clearly the prior must be stated so that it can be assessed by readers.

Our methodology is a hybrid of instrumental variables and Bayesian methods. Purists will not like this, since full Bayesian methods require a full probability model while instrumental variables allows us to avoid this. However, we believe that our method combines the best aspects of both methodologies, using instrumental variables methods to avoid assumptions about the relationship between the dosage D and the untreated outcome Y, and Bayesian methods to incorporate prior beliefs about nonprotocol parameters.

Our results depend on the truth of the causal model. In PENTA 5, this assumed that the causal effect of a dose of drug was proportional to the dose received, and that there were no interactions between drugs. These assumptions are not easily checked from the data, since adding further terms to the causal model would only increase the number of nonidentified parameters on which prior information is needed. Therefore, sensitivity analyses are important [33]. In PENTA 5, they could involve adding to the causal model (1) nonlinear terms with assumed coefficient values. Estimation would then require expectations of these nonlinear terms to be computed as in (13) and added into (14).

Our computational approach for the PENTA 5 data used Monte Carlo Markov Chain methods in WinBUGS. Computationally simpler approaches could be available; one idea might be to draw a small sample from the prior $p(\alpha)$, analyze the data conditionally on each value of α, and combine results as for multiply imputed data [34].

The methods presented here can also been extended to analyze longitudinal data, by replacing models (13) and (14) with time-dependent versions. This involves attention to the longitudinal error structure.

Finally, our method in the simple case presented requires just knowledge of the mean receipt of each drug on each arm. Some information about treatment received is typically reported, but application of methods such as ours would be facilitated if treatment received were reported in a more standard way.

Acknowledgments

Funding

References

1. **Pocock SJ.** *Clinical Trials: A Practical Approach*. Wiley, Chichester, 1983.
2. **Efron B, Feldman D.** Compliance as an explanatory variable in clinical trials. *J Am Stati Assoc* 1991; **86**: 9–17.
3. **White IR.** Uses and limitations of randomization-based efficacy estimators. *Stat Methods Med Res* 2005; **14**: 327–47.
4. **Porta N, Bonet C, Cobo E.** Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol* 2007; **60**: 663–69.
5. **Robins JM, Greenland S.** Adjusting for differential rates of prophylaxis therapy for PCP in high-dose versus low-dose AZT treatment arms in an AIDS randomized trial. *J Am Stat Assoc* 1994; **89**: 737–49.
6. **Robins JM.** Correction for noncompliance in equivalence trials. *Stat Med* 1998; **17**: 269–302.
7. **Walter SD, Guyatt G, Montori VM**, *et al*. A new preference-based analysis for randomized trials can estimate treatment acceptability and effect in compliant patients. *J Clin Epidemiol* 2006; **59**: 685–96.
8. **Roy J, Hogan JW, Marcus BH.** Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics* 2007; **9**: 277–89.
9. **Bernardo JM, Smith AFM.** *Bayesian Theory* (2nd edn). Wiley, Chichester, 2001.
10. **Lunn DJ, Thomas A, Best N, Spiegelhalter D.** WinBUGS – a Bayesian modeling framework: concepts, structure, and extensibility. *Stat Comput* 2000; **10**: 325–37.
11. **Bowden RJ, Turkington DA.** *Instrumental Variables*. Cambridge University Press, Cambridge, 1984.
12. **Jones B, Jarvis P, Lewis J, Ebbutt A.** Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996; **313**: 36–39.
13. **Brittain E, Lin D.** A comparison of intent-to-treat and per-protocol results in antibiotic noninferiority trials. *Stat Med* 2005; **24**: 1–10.
14. **Fleming TR.** Current issues in noninferiority trials. *Stat Med* 2008; **27**: 317–32.
15. **Rubin D.** Estimating causal effects of treatments in randomized and non-randomized studies. *J Educ Psychol* 1974; **66**: 688–701.
16. **Bond SJ, White IR, Walker AS.** Instrumental variables and interactions in the causal analysis of a complex clinical trial. *Stat Med* 2006; **26**: 1473–96.
17. **Robins JM.** Correcting for noncompliance in randomized trials using structural nested mean models. *Commun Stat: Theor Meth* 1994; **23**: 2379–2412.
18. **Imbens GW, Rubin DB.** Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann Statist* 1997; **25**: 305–27.
19. **Walker AS, White IR, Babiker AG.** Parametric randomization-based methods for correcting for treatment changes in the assessment of the causal effect of treatment. *Stat Med* 2004; **23**: 571–90.
20. **Goetghebeur E, Lapp K.** The effect of treatment compliance in a placebocontrolled trial: Regression with unpaired data. *J R Stat Soc Ser C* 1997; **46**: 351–64.

21. **Fischer K, Goetghebeur E.** Practical properties of some structural mean analyses of the effect of compliance in randomized trials. *Control Clin Trials* 1999; **20**: 531–46.
22. **Fischer K, Goetghebeur E, Vrijens B, White IR.** A structural mean model to analyze the effect of compliance when comparing two active treatments. *Biostatistics* (in press).
23. **Oakley JE, O'Hagan A.** Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J R Stat Soc Ser B* 2004; **66**: 751–69.
24. **O'Hagan A, Buck CE, Daneshkhah A,** *et al. Uncertain Judgements: Eliciting Expert Probabilities*. Wiley, Chichester, 2006.
25. **Paediatric European Network for Treatment of AIDS.** Comparison of dual nucleoside-analogue reverse-transcriptase inhibitor regimes with and without nelfinavir in children with HIV-1 who have not previously been treated: the PENTA 5 randomized trial. *Lancet* 2002; **359**: 733–40.
26. **Sommer A, Zeger SL.** On estimating efficacy from clinical trials. *Stat Med* 1991; **10**: 45–52.
27. **Angrist JD, Imbens GW, Rubin DB.** Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996; **91**: 444–72.
28. **Frangakis CE, Rubin DB.** Principal stratification in causal inference. *Biometrics* 2002; **58**: 21–29.
29. **Cuzick J, Edwards R, Segnan N.** Adjusting for noncompliance and contamination in randomized clinical trials. *Stat Med* 1997; **16**: 1017–29.
30. **Hirano K, Imbens GW, Rubin DB, Zhou XH.** Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* 2000; **1**: 69–88.
31. **Barnard J, Frangakis CE, Hill JL, Rubin DB.** Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. *J Am Stat Assoc* 2003; **98**: 299–311.
32. **White IR, Carpenter J, Evans S, Schroter S.** Eliciting and using expert opinions about non-response bias in randomised controlled trials. *Clin Trials* 2007; **4**: 125–39.
33. **Vansteelandt S, Goetghebeur E.** Sense and sensitivity when correcting for observed exposures in randomized clinical trials. *Stat Med* 2005; **24**: 191–210.
34. **Rubin DB.** *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons, New York, 1987.
35. **Golub GH, Van Loan CF.** *Matrix Computations* (2nd edn). Johns Hopkins University Press, Baltimore, 1989.

Appendices

(A) Information sheet for elicitees

Randomized control trial prior elicitation

You will shortly be asked for your opinion on the results of a sequence of hypothetical RCTs. These trials all have recruitment criteria that are identical to those used in the PENTA-5 trial (children aged 3 months to 16 years, HIV-1 infection, no previous anti-retroviral treatment and without transmitted drug resistance). They only differ in the combinations of drugs to be compared in the two arms.

For example, in question 1, the trial consists of two arms; arm A (3TC+ABC+PI) versus arm B (ABC+NRTI+PI), and the outcome is change in $\log_{10}$ RNA measurement from baseline to 48 weeks.

The end result of the trial is the mean difference (arm A − arm B) between these changes from baseline in the two randomized groups. So for example, if there was a 2.5 log drop in arm A, and a 3.0 log drop in arm B, the mean difference between arms would be $+0.5 = -2.5 - -3.0$. Call this quantity, X: X is unknown, but in your expert opinion there are some values that are more likely than others.

A potential 'best guess' for X is the point whereby you believe that 50% of the time the X you actually observed in a RCT would be less than this cut-point, and 50% of the time it would be larger: this would be the median X observed if we carried out a large number of trials. This median would be the cut-point whereby you would be ambivalent between accepting a bet, or placing a bet, on the event that the X observed in one trial would be larger than this.

Mark on the line of your expert opinion on this value, the median. If you think B is a better regimen than A, this value should be positive, the right-hand side of the line. If you think A is better it should be negative, the left-hand side of the line. For clarity please annotate the mark with the letter 'M'.

Now, imagine that X has been observed in an RCT and you are told that its value is greater than the value you have marked for the median – that is, a RCT has shown that B is actually better than you thought it was. Conditional on this extra knowledge, mark your 'best guess' at the revised median on the same line. *This mark should always be to the right of your first mark* – B is better than you thought it was.

Finally, you are now told that another trial has observed an X even greater than the value you have just marked , so that B is even better than your second guess: mark your 'best guess' at the revised median. *This mark should always be to the right of your first two marks* – B is even better than you thought it was.

This exercise is now repeated for another 21 hypothetical trials that have the same outcome measure and recruitment procedure, but differ in the treatment assigned in the two arms. Where the randomized comparison is between a class and a specific drug (e.g., 3TC vs. NRTI in the first example), please assume that the clinician has a free choice amongst drugs of that class, and is likely to prescribe according to current clinical practice.

(B) Visual analog scale

Date:...

Name:..

Arm A	Arm B	Arm A - Arm B
3TC ABC PI	ABC NRTI PI	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV ABC PI	ZDV 3TC ABC	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV ABC PI	ABC NRTI PI	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV ABC	ZDV ABC PI	−1.5 −1 −0.5 0 0.5 1 1.5
3TC ABC Pi	ZDV 3TC PI	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV 3TC PI	ZDV 3TC ABC	−1.5 −1 −0.5 0 0.5 1 1.5
3TC ABC NNRTI	ZDV 3TC ABC	−1.5 −1 −0.5 0 0.5 1 1.5
3TC ABC PI	ZDV ABC PI	−1.5 −1 −0.5 0 0.5 1 1.5
3TC ABC	3TC ABC PI	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV ABC NNRTI	ZDV 3TC ABC	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV ABC	ZDV ABC NNRTI	−1.5 −1 −0.5 0 0.5 1 1.5
3TC ABC PI	3TC ABC NNRTI	−1.5 −1 −0.5 0 0.5 1 1.5
3TC ABC	3TC ABC NNRTI	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV 3TC	ZDV 3TC ABC	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV ABC NNRTI	3TC ABC NNRTI	−1.5 −1 −0.5 0 0.5 1 1.5
3TC ABC PI	ZDV 3TC ABC	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV 3TC PI	ZDV ABC PI	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV ABC PI	ZDV ABC NNRTI	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV 3TC	ZDV 3TC PI	−1.5 −1 −0.5 0 0.5 1 1.5
ZDV ABC	ZDV 3TC ABC	−1.5 −1 −0.5 0 0.5 1 1.5
NRTI PI	ABC NRTI PI	−1.5 −1 −0.5 0 0.5 1 1.5
3TC ABC	ZDV 3TC ABC	−1.5 −1 −0.5 0 0.5 1 1.5

(C) WinBUGS code

```
model{

for (i in 1:N){
  # i indexes N subjects

  # D-Model Fitting

  for (j in 1:P){
    # j indexes the P (4) protocol & non-protocol treatments in D
    # columns of D are ZDV, 3TC, ABC, PI

    # Q is the number (4) of baseline variables & arm indicators given in X.
    # Columns of X are arm2, arm3, intercept, baseline log HIV RNA.
    # So for computation, X and R of section 2 are combined into one matrix.

    # eta gives the predicted values for the D models

    eta[j,i] <- inprod(beta[j,1:Q], X[i,1:Q])
    D[j,i] ~ dnorm(eta[j,i], prec[j])
  }
  # Now convert the fitted D to the desired treatment contrasts
  prot[1,i] <- eta[2,i]-eta[1,i]        # 3TC-ZDV
```

```
      prot[2,i] <- eta[2,i]-eta[3,i]      # 3TC-ABC
      nonprot[1,i] <- eta[1,i]+eta[3,i]   # ZDV+ABC
      nonprot[2,i] <- eta[4,i]            # PI

      # Use the WinBugs cut function to ensure the D-models are not influenced
      # by the Y-model estimation
      Dhat[1,i] <- cut(prot[1,i])
      Dhat[2,i] <- cut(prot[2,i])
      Dhat[3,i] <- cut(nonprot[1,i])
      Dhat[4,i] <- cut(nonprot[2,i])

      # Y-Model Fitting

      Yhat[i] <- inprod(alpha.protocol[1:2], Dhat[1:2,i])
                 + inprod(alpha.nonprotocol[1:2], Dhat[3:P,i])
                 + inprod(gamma[1:2], X[i,3:Q])
      # Use the WinBugs I(,) syntax  for censored observations
      Y[i] ~ dnorm(Yhat[i],precY) I(, 1.699)
    }

    # Prior Distributions

    # D-model

    for (j in 1:P) {
      for (k in 1:Q) {

        beta[j,k] ~ dnorm(0,0.01)
      }
      prec[j] ~ dgamma(0.01,0.01)
      sd[j] <- 1/sqrt(prec[j])
    }

    # Y-model

    # Computational efficiency suggests to start with a set of independent
    # normals (delta) and then to linearly transform to obtain the desired
    # multivariate normal for the priors on alpha.protocol and alpha.nonprotocol.

    # To achieve a Multivariate-Normal(mu, Sigma),
    # find Chol such that Sigma= Chol * transpose(Chol)
    #   e.g. Cholesky decomposition,
    # find delta.mean=inverse(Chol)* mu,
    # then Chol*N(delta.mean,I)= MvN(mu, Sigma)

    for(j in 1:P){
      delta[j] ~ dnorm(delta.mean[j],1)
    }

    for(j in 1:2){
      alpha.protocol[j] <- inprod(Chol[j,1:P], delta[1:P])
      alpha.nonprotocol[j] <- inprod(Chol[j+2,1:P], delta[1:P])
      gamma[1:2] ~ dnorm(0,0.001)
```

```
}
precY ~ dgamma(0.01,0.01)
sdY <- 1/sqrt(precY)

# Need to specify in the data:
# N
# P
# Q
# Y[]
# X[,]
# D[,]
# Chol[,]
# delta.mean[]

}
```

(D) Independent prior

Let the $g-1$ protocol contrasts and the $p-g+1$ nonprotocol contrasts be expressed as $\boldsymbol{\alpha}_p = \boldsymbol{L}_p\,\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_n = \boldsymbol{L}_n\boldsymbol{\alpha}$, respectively, where $\boldsymbol{L}_p$ is a $(g-1) \times p$ matrix and $\boldsymbol{L}_n$ is a $(p-g+1) \times p$ matrix. Let the full prior for p-dimensional $\boldsymbol{\alpha}$ be multivariate normal with covariance matrix Σ.

We argued in the section 'Obtaining prior information' that $\boldsymbol{\alpha}_n$ should be independent of $\boldsymbol{\alpha}_p$ in the prior, so we must find $\boldsymbol{L}_n$ that solves $\boldsymbol{L}_n\Sigma\boldsymbol{L}_p' = 0$. This can be done by finding the **QR** decomposition [35, section 5.2] of $\Sigma\boldsymbol{L}_p' = \boldsymbol{QR}$, which computes the $p \times p$ orthogonal matrix $\boldsymbol{Q}$, and the $p \times (g-1)$ matrix $\boldsymbol{R}$, with zero elements apart from the first $g-1$ rows that form an upper triangular matrix. Pre-multiplying by $\boldsymbol{Q}'$ shows that the last $p-g+1$ rows of $\boldsymbol{Q}'$ provide a solution.

For the PENTA 5 prior distribution expressed in Table 4, this procedure gives the two components of $\boldsymbol{\alpha}_n$ as $0.806\alpha_1 + 0.413\alpha_2 + 0.687\alpha_3 + 0.012\alpha_4$, roughly representing an average effect of the three trial treatments, and $-0.046\alpha_1 - 0.088\alpha_2 + 0.176\alpha_3 + 0.985\alpha_4$, roughly representing the effect of the nontrial treatment.