

## Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up

G. DUNN, M. MARACY, C. DOWRICK, J. L. AYUSO-MATEOS, O. S. DALGARD, H. PAGE, V. LEHTINEN, P. CASEY, C. WILKINSON, J. L. VÁZQUEZ-BARQUERO, G. WILKINSON on behalf of the ODIN group

**Background** The Outcomes of Depression International Network (ODIN) trial evaluated the effect of two psychological interventions for the treatment of depression in primary care. Only about half of the patients in the treatment arm complied with the offer of treatment, prompting the question: 'what was the effect of treatment in those patients who actually received it?'

**Aims** To illustrate the estimation of the effect of receipt of treatment in a randomised controlled trial subject to non-compliance and loss to follow-up.

**Method** We estimated the complier average causal effect (CACE) of treatment.

**Results** In the ODIN trial the effect of receipt of psychological intervention (an average of about 4 points on the Beck Depression Inventory) is about twice that of offering it.

**Conclusions** The statistical analysis of the results of a clinical trial subject to non-compliance to allocated treatment is now reasonably straightforward through estimation of a CACE and investigators should be encouraged to present the results of analyses of this type as a routine component of a trial report.

**Declaration of interest** None. Funding detailed in Acknowledgements. Paper accepted when G.W. was Editor of the *Journal*.

Mr Jones has come to the general practitioner's clinic complaining of symptoms of depression. An obvious question for the general practitioner to ask herself is: 'If I were to offer psychological treatment to Mr Jones, would he benefit from the receipt of this treatment?' What can the results of a randomised controlled trial, particularly intention-to-treat (ITT) estimates of treatment effects, tell the general practitioner about the answer to this question? A conventional ITT effect estimate, at best, might provide a good answer to a question of the form 'If I were to offer psychological treatment to Mr Jones, by how much would he benefit from the offer?' Even if there are no problems in generalising from the results of a randomised controlled trial, the answer to this question is only the same as that to the former question if there is complete take-up of the offer of treatment. That is, the ITT estimate of a treatment effect is a valid estimate of the receipt of treatment only if there is 100% acceptance of the treatment in the randomly selected group offered it and if none of those who are randomly allocated to be the controls get access to the treatment. But most real trials in this field are not like that. If we are really interested in assessing the size of the benefit of receipt of treatment, as opposed to merely the offer of it, then our statistical analysis needs to proceed beyond ITT. Using the phrase of Heckman *et al* (1998), we need to estimate 'the effect of treatment on the treated'. Newcombe (1988) refers to an 'explanatory' estimate of a treatment effect as opposed to the more familiar 'pragmatic' (or ITT) estimate. The purpose of the present paper is to use the results of the Outcomes of Depression International Network (ODIN) trial (Dowrick *et al*, 2000) to illustrate how this might be done.

### METHOD

The ODIN trial is a European project studying the prevalence and outcomes of

depression in urban and rural communities (Dowrick *et al*, 2000; Ajuso-Mateos *et al* 2001). One objective of the ODIN trial was to assess the efficacy of psychological interventions. We identified two simple, reproducible interventions that could be delivered in the community: problem-solving and group sessions of a course on the prevention of depression (psychoeducation). The outcomes following either of these two treatments or a treatment-as-usual control were compared in a multi-centre randomised controlled trial. The main results from this trial have been described previously (Dowrick *et al*, 2000). Problem-solving appeared to be more acceptable than a course of psychoeducation (as measured by compliance patterns in the two treatment groups) but both led to improved outcomes (in comparison with the controls) when measured 6 months after randomisation. At 12 months, however, the outcomes in all three groups were very similar.

The detailed aims of the present paper are to study in further depth the estimation of selected measures of efficacy of these psychological treatments. These efficacy measures are formally defined in terms of two types of average treatment effect using recently developed theories of causal inference as applied to randomised controlled trials in which there is the possibility of both non-compliance to allocated treatment and subsequent drop-out (i.e. missing outcome data). Our aim is to provide an illustration of an analysis strategy that might be used as an informal model to be applied to the analysis of a wide variety of trials of complex interventions in psychiatry. A further aim of this paper is to illustrate approaches to the assessment of the sensitivity of the estimates of treatment effects to various assumptions concerning the impact of merely offering treatment – the definition of receipt of treatment (compliance) – after adjusting for the influence of non-compliance on loss to follow-up.

The present report, unlike many descriptions of the results of randomised controlled trials, actually emphasises the problems arising from non-compliance and subsequent loss to follow-up. This approach is chosen for two reasons: to obtain valid estimates of average treatment effects of interest and to challenge our assumptions concerning the influence of patient preferences on the outcome of being offered and/or receiving treatment. Both

should lead to the possibility of more informative designs for complex intervention studies. We hope that we might be able to stimulate other investigators to explore the data from their own trials more thoroughly and not simply sweep the problems 'under the carpet'.

## Study design

The ODIN trial involved nine study centres in Finland (2), the Republic of Ireland (2), Norway (2), Spain (1) and the UK (2). The trial was designed to compare the outcomes of problem-solving treatment or a depression prevention course (psychoeducation) with outcome in a control group receiving no intervention. Within each centre patients were allocated randomly to receive either one of the two types of treatment (the treatment group) or no intervention (the control group). Problem-solving (but not psychoeducation) was available in Spain, Finland (both centres) and the UK (one centre). Psychoeducation (but not problem-solving) was available in Ireland (both centres) and Norway (both centres). The second UK centre was the only one in the trial in which patients could be allocated randomly to any of the three treatment arms. The main implication of this complex design is that the formal analysis should involve stratification by centre (to ensure that the treatment groups are being compared with the appropriate controls). Further details of the design, including detailed descriptions of the interventions offered, are provided in Dowrick *et al* (2000). Note that the results from 427 randomised patients are analysed in the present report; Dowrick *et al* (2000) used 426, as one patient had been inadvertently missed from the previous analysis owing to clerical error. Because of the small number of patients in the two centres from Ireland, in the present analysis the two Irish centres are treated as one.

For the purpose of the present paper there were three measured outcomes of treatment allocation (randomisation): how well the patient adhered to (complied with) the allocation treatment; whether or not the patient was lost to follow-up (six months after randomisation); and, if available, a measure of the severity of depression at follow-up. The latter was assessed using the total score of the Beck Depression Inventory (BDI; Beck *et al*, 1961). Adherence to the allocated treatment was

measured on a four-point nominal scale: 'Attended', 'Refused', 'Discontinued' and 'Did not attend'. In order to proceed with further analyses, this scale was dichotomised in one of two ways: for compliance A, 'Attended' was coded 1 and the rest 0; for compliance B, 'Attended' and 'Discontinued' were both coded 1 and the rest 0. A patient was deemed to have received treatment if he or she was in the allocated treatment group and the relevant compliance code was 1 (patients did not have access to treatment if they had been allocated to the control group).

## Analysis strategy

### Initial description of the data

First, the frequencies for each of the patterns of adherence to allocated treatment are examined for each treatment centre (separately for each treatment type for the UK centre offering both treatments). Then the patients who were allocated to the treatment group are classified as compliers or non-compliers, according to compliance A or B. Observed compliance status has three levels: 'Control', 'Yes' or 'No'. Patterns of observed compliance status are examined for each treatment centre, together with the numbers of patients in each category providing depression severity ratings. Finally, means for the depression severity ratings are calculated for each of the compliance categories within each of the treatment centres. These preliminary data descriptions enable us to evaluate the level of adherence to allocated treatment, whether the levels of adherence depend on the nature of the treatment on offer and the amount of variability in adherence from one treatment centre to another. They also enable us to see whether the rate of loss to follow-up is dependent on compliance status and how this varies from one centre to another. Finally, we see how severity of depression varies with compliance status within and across treatment centres. These data then provide the material for the more detailed analyses described below.

### In-depth analysis

**Assumptions concerning non-compliance.** We start by assuming that the patients taking part in the trial belong to one of two potentially latent classes: compliers and non-compliers. In the treatment group the non-compliers are those who fail to receive treatment when they are offered it.

In the control group they are those patients who would have failed to receive treatment had they been offered it. Compliers are those who received treatment in the treatment group and those in the control group who would have received treatment had they been offered it. We can observe compliance status in the treatment group but it is latent or unobservable in the control group.

Randomisation ensures that, on average, the proportion of compliers in the control group is the same as that in the treatment group (Bloom, 1984; Sommer & Zeger, 1991). This means that we can estimate the proportion of unobserved compliers in the control group (or, equivalently, the proportion of compliers in the trial as a whole,  $\pi_c$ ) from the proportion observed in the treatment group ( $P_c$ ).

**Definitions of treatment effects.** We define the average causal effect (ACE) of treatment as the difference between the 6-month average BDI score for the treatment group and that for the control group (regardless of compliance status or whether the outcome is actually observed). An alternative term is the 'average treatment effect' (Angrist *et al*, 1996). This is the treatment effect that we are trying to estimate in a so-called ITT analysis. It is the difference in outcomes between the two treatment groups as randomised, as opposed to treatment actually received.

We define the complier average causal effect (CACE) as the difference between the 6-month average BDI score for the compliers in the treatment group and that for the compliers in the control group (regardless of whether the outcome is actually observed). An alternative term is the 'local average treatment effect' (Angrist *et al*, 1996). For reasons clearly explained by Sheiner & Rubin (1995) and by Frangakis & Rubin (1999), we do not consider effects estimated by methods involving analysis 'per protocol' or 'as treated' (the former compares the compliers in the treatment group with all of the controls, and the latter compares those who receive treatment with those who do not, regardless of random allocation) – neither being estimates of valid treatment effects described in this paper.

**Exclusion restriction.** Given the treatment received, we assume that outcome is independent of random allocation. That is, the

offer of treatment, in itself, does not influence outcome (Bloom, 1984; Sommer & Zeger, 1991). This assumption is often referred to as an ‘exclusion restriction’ (Angrist *et al*, 1996). From this assumption we can assume that the mean BDI score for the non-compliers in the control group is, on average, the same as that for the non-compliers in the treatment group. This enables us to estimate the unobserved mean for non-compliers in the control group by the observed average for the non-compliers in the treatment group.

It is straightforward to show from the exclusion restriction assumption that

$$ACE = \pi_c \times CACE \quad (1)$$

where  $\pi_c$  is the proportion of compliers in the trial (Angrist *et al*, 1996). Typically, we proceed by first estimating the CACE and then, if we require an estimate of the ACE, using this to get an estimate of the ACE from equation (1) together with our estimate of  $\pi_c$  (i.e.  $P_c$ ). In the simple situation where outcome measures are available for all trial participants (i.e. there is no loss to follow-up), then the required estimate for the ACE is the familiar ITT estimate. In this circumstance we can then simply estimate the effect of receiving treatment (equivalent to the CACE) from

$$CACE = ACE/\pi_c \quad (2)$$

Details can be found in Angrist *et al* (1996). When we have both non-compliance and non-ignorable missing follow-up data (see below), however, the naïve but frequently used ITT estimates are likely to be biased and we have to approach the analysis via CACE estimation by taking into account the missing data mechanism (Frangakis & Rubin, 1999).

**Missing data mechanisms and simple methods of CACE estimation.** If, in addition to non-compliance, we also have missing outcome data then we have to make further assumptions concerning the missing data mechanism. The first option is to assume that the missing data mechanism is ignorable. Here the data are either missing completely at random or missing at random, in the sense defined by Little & Rubin (2002). Looking ahead, it is clear from a glance at Table 2 that the outcome data are not missing completely at random (loss to follow-up is clearly related to compliance status). But suppose, for example, in the simple situation where there are

no measured covariates, that the probability of being missing is determined by observed compliance status (complier, non-complier or a member of the control group) and that, conditional on observed compliance status, outcome is statistically independent of whether outcome is actually observed. Here, the outcome data are missing at random (MAR). Under these assumptions it is straightforward to show that

$$CACE_{MAR} = \frac{\pi_c \mu_{11} + (1 - \pi_c) \mu_{10} - \mu_0}{\pi_c} \quad (3)$$

where  $\mu_{11}$  is the mean outcome for the compliers in the treatment group,  $\mu_{10}$  is the corresponding mean for the non-compliers and  $\mu_0$  is the mean for the controls. The  $CACE_{MAR}$  value can be estimated easily by replacing  $\mu_c$  and the three  $\mu$  terms by their corresponding values in equation (3). If there are no missing outcome data, equation (3) simplifies to the estimator first described by Bloom (1984). The standard error of this so-called moments estimate can be obtained using the delta technique or a simple bootstrap (Efron & Tibshirani, 1993).

The alternative missing data option is that they are non-ignorable (Little & Rubin, 2002). That is, whether a patient has a missing outcome is dependent on the value of that outcome, even after conditioning on observed variables such as compliance status and baseline covariates. This is a much more difficult problem to deal with and we refer the interested reader to a recent paper on this topic by Frangakis & Rubin (1999). A less demanding discussion of the work of Frangakis & Rubin is provided by Dunn (2002*b*). In order to keep the technical details to a minimum, we do not pursue this option in any detail in the present paper.

**Refinement of CACE estimation: incorporating baseline covariates.** Although technically more difficult, if we have access to baseline covariates (including treatment centre) we can develop more efficient (i.e. precise) CACE estimation methods. We can also get more stable estimates of the average treatment effects within each of the centres. Maximum likelihood methods, based on the joint distribution of the binary compliance status and a normally distributed outcome measure, have been developed by Angrist *et al* (1996), Little & Yau (1998)

and Yau & Little (2001) – the latter incorporating data missing at random. These methods enable the incorporation of covariates in the model to predict jointly both the latent compliance status and the outcome (the outcome is also predicted by compliance status as well as by the covariates).

In the present study, CACE models incorporating the potential use of baseline covariates (initial BDI score and centre membership) to predict both compliance status and outcome were fitted via maximum likelihood estimation using the expectation maximisation algorithm (*Mplus* Version 2.12; Muthén & Muthén, 1998–2002). The use of the latter software package in the application of this methodology on randomised controlled trial data with non-compliance is illustrated in detail by Jo & Muthén (2001), although they do not consider problems arising from missing outcome data.

**Sensitivity analysis.** Rather precise assumptions (e.g. concerning the definition of compliance, the missing data mechanism and exclusion restriction) are vital components of the analytical approaches described above for the estimation of average treatment effects. Having to make these assumptions is both a strength and a weakness of these approaches. If we get the assumptions wrong we risk invalid inferences, but a thorough examination of the implications of the assumptions helps to understand what might be going on in a psychological treatment trial. They force us to think more about the trial process and to clarify what we are really interested in estimating. Another vital component of the analytical approach therefore is to attempt to evaluate the sensitivity of our treatment effect estimates to changes in these assumptions.

All preliminary analyses and checks of the sensitivity of the treatment effects to assumptions concerning the definition of compliance were carried out using Stata Version 7.0 (StataCorp, 2001). An exploration of the sensitivity of the CACE estimates to the validity of the main exclusion restriction assumption (treatment allocation does not influence outcome except through its effect on treatment received), using either of the two definitions of compliance, was carried out as described by Jo (2002*a,b*). Readers are also referred to Heckman *et al* (1998) and Hirano *et al* (2000).

## RESULTS

## Preliminary examination of the data

Table 1 shows the patterns of adherence to the offered treatment (i.e. excluding controls) in each of the nine centres (separately for the two types of treatment offered by centre 7). To illustrate the variation in the patterns of adherence in more detail, we look at patterns of compliance using compliance A (Table 2). Compliance rates vary greatly from one centre to another (ranging from 40% in centre 1 to 74% in centre 3). Some of the variation may be explained by the type of treatment being offered, but we do not stress this aspect of the results because the design of the trial leads to this source of variation being almost completely confounded with the centre effects. These compliance rates may appear to be rather low, but the reader must bear in mind that the participants in the ODIN trial were recruited through a case-finding exercise. They were not patients who had actively sought help.

Loss to follow-up (i.e. missing outcome data) varies from one centre to another but is also markedly dependent on compliance status. Loss to follow-up in the compliers in the treatment group is very infrequent. In four of the nine centres the compliers provide 100% of the required outcome data, with follow-up of those in the other five centres ranging from 79% (centre 7) to 91% (centre 5). However, loss to follow-up is both more variable and more common in the non-compliers of the treatment group; here, follow-up rates range from 22% (centre 1) to 75% (centre 4). In no case is the within-centre follow-up rate for the non-compliers as high as that for the corresponding compliers. As might be expected, the follow-up rates for the controls lies somewhere between those for the compliers and non-compliers in the treatment group.

Moving on to consider the severity of depression at outcome (the mean BDI score at 6 months) we see that, on average, patients offered treatment do better than the controls (bottom three rows of Table 3). However, this difference is not always apparent within each of the centres. On average, the compliers in the treatment group have very similar outcomes to those who do not comply with the offered treatment (last two rows of Table 3) but again there is a considerable amount of variability in this difference from one centre

Table 1 Patterns of adherence to allocated treatment

| Centre   | Treatment | Attended  | Refused  | Discontinued | Did not attend | Total |
|----------|-----------|-----------|----------|--------------|----------------|-------|
| 1        | PE        | 6         | 0        | 3            | 6              | 15    |
| 2        | PS        | 12        | 7        | 0            | 0              | 19    |
| 3        | PS        | 17        | 0        | 5            | 1              | 23    |
| 4        | PS        | 20        | 3        | 4            | 1              | 28    |
| 5        | PE        | 22        | 15       | 5            | 0              | 42    |
| 6        | PE        | 17        | 12       | 5            | 2              | 36    |
| 7        | PE        | 3         | 5        | 1            | 6              | 15    |
| 7        | PS        | 16        | 5        | 7            | 4              | 32    |
| 8        | PS        | 15        | 5        | 6            | 0              | 26    |
| Total    |           | 128 (54%) | 52 (22%) | 36 (15%)     | 20 (8%)        | 236   |
| Controls |           |           |          |              |                | 191   |

PE, psychoeducation; PS, problem-solving.

Table 2 Summaries of compliance status and availability of outcome data (using compliance A)

| Centre                   | Offered treatment | Received treatment | Patients n (%) <sup>1</sup> | Patients with outcome data n (%) <sup>2</sup> |
|--------------------------|-------------------|--------------------|-----------------------------|---|
| 1 (Ireland) <sup>3</sup> | No                | No                 | 23 (100)                    | 12 (52)                                       |
|                          | Yes               | No                 | 9 (60)                      | 2 (22)  |
|                          | Yes               | Yes                | 6 (40)                      | 6 (100)                                       |
| 2 (Spain)                | No                | No                 | 11 (100)                    | 7 (64)  |
|                          | Yes               | No                 | 7 (37)                      | 3 (43)  |
|                          | Yes               | Yes                | 12 (63)                     | 12 (100)                                      |
| 3 (Finland)              | No                | No                 | 24 (100)                    | 17 (71)                                       |
|                          | Yes               | No                 | 6 (26)                      | 2 (33)  |
|                          | Yes               | Yes                | 17 (74)                     | 17 (100)                                      |
| 4 (Finland)              | No                | No                 | 22 (100)                    | 20 (91)                                       |
|                          | Yes               | No                 | 8 (29)                      | 6 (75)  |
|                          | Yes               | Yes                | 20 (71)                     | 18 (90)                                       |
| 5 (Norway)               | No                | No                 | 25 (100)                    | 17 (68)                                       |
|                          | Yes               | No                 | 20 (48)                     | 11 (55)                                       |
|                          | Yes               | Yes                | 22 (52)                     | 20 (91)                                       |
| 6 (Norway)               | No                | No                 | 25 (100)                    | 18 (72)                                       |
|                          | Yes               | No                 | 19 (53)                     | 15 (79)                                       |
|                          | Yes               | Yes                | 17 (47)                     | 17 (100)                                      |
| 7 (UK)                   | No                | No                 | 37 (100)                    | 31 (84)                                       |
|                          | Yes               | No                 | 28 (60)                     | 16 (57)                                       |
|                          | Yes               | Yes                | 19 (40)                     | 15 (79)                                       |
| 8 (UK)                   | No                | No                 | 24 (100)                    | 18 (75)                                       |
|                          | Yes               | No                 | 11 (42)                     | 4 (36)  |
|                          | Yes               | Yes                | 15 (58)                     | 13 (87)                                       |
| Total                    | No                | No                 | 191 (100)                   | 140 (73)                                      |
|                          | Yes               | No                 | 108 (46)                    | 59 (55)                                       |
|                          | Yes               | Yes                | 128 (54)                    | 118 (92)                                      |

1. The percentage of patients who were allocated to the treatment arm.

2. The percentage of patients in the previous column of the table.

3. The two treatment centres from Ireland were combined for the purposes of analyses in the present paper.

**Table 3** Observed Beck Depression Inventory scores at baseline and 6 months (using compliance A)

| Centre      | Offered treatment | Received treatment | Baseline mean (s.d.) <sup>1</sup> | Six-month mean (s.d.) |
|-------------|-------------------|--------------------|-----------------------------------|-----------------------|
| 1 (Ireland) | No                | No                 | 25.74 (9.72)                      | 9.33 (8.98)           |
|             | Yes               | No                 | 26.44 (10.03)                     | 1.00 (1.41)           |
|             | Yes               | Yes                | 19.00 (7.46)                      | 10.67 (3.61)          |
| 2 (Spain)   | No                | No                 | 22.27 (6.59)                      | 16.29 (7.87)          |
|             | Yes               | No                 | 21.86 (5.73)                      | 12.67 (5.13)          |
|             | Yes               | Yes                | 22.42 (8.59)                      | 7.17 (5.13)           |
| 3 (Finland) | No                | No                 | 20.83 (6.31)                      | 13.12 (8.53)          |
|             | Yes               | No                 | 22.17 (6.18)                      | 12.50 (6.36)          |
|             | Yes               | Yes                | 21.76 (5.66)                      | 9.18 (6.61)           |
| 4 (Finland) | No                | No                 | 19.91 (6.53)                      | 8.70 (7.41)           |
|             | Yes               | No                 | 23.13 (10.19)                     | 8.83 (3.06)           |
|             | Yes               | Yes                | 21.90 (7.30)                      | 7.67 (4.30)           |
| 5 (Norway)  | No                | No                 | 20.64 (6.26)                      | 13.12 (9.75)          |
|             | Yes               | No                 | 20.85 (8.99)                      | 14.91 (9.70)          |
|             | Yes               | Yes                | 22.82 (8.81)                      | 12.00 (7.95)          |
| 6 (Norway)  | No                | No                 | 19.32 (7.12)                      | 15.44 (9.41)          |
|             | Yes               | No                 | 18.53 (5.17)                      | 14.20 (8.62)          |
|             | Yes               | Yes                | 21.76 (9.42)                      | 18.29 (11.97)         |
| 7 (UK)      | No                | No                 | 24.70 (8.88)                      | 21.90 (11.21)         |
|             | Yes               | No                 | 24.61 (9.19)                      | 14.81 (11.48)         |
|             | Yes               | Yes                | 25.47 (9.12)                      | 21.00 (12.48)         |
| 8 (UK)      | No                | No                 | 26.04 (8.71)                      | 17.72 (10.63)         |
|             | Yes               | No                 | 22.54 (8.44)                      | 12.00 (12.36)         |
|             | Yes               | Yes                | 27.73 (8.03)                      | 20.15 (11.81)         |
| Total       | No                | No                 | 22.58 (8.06)                      | 15.16 (10.42)         |
|             | Yes               | No                 | 22.36 (8.40)                      | 13.22 (9.35)          |
|             | Yes               | Yes                | 23.15 (8.27)                      | 13.32 (10.14)         |

1. Data from every patient randomised (total  $n=427$ ).

to another. In centres 2–5 the compliers appear to fare better than the non-compliers. In centres 1, 6, 7 and 8, however, the non-compliers fare better. Again, we do not concentrate on the differences in effects for the two types of psychological intervention because these differences are confounded by differences between centres. Returning to the data for the whole trial (bottom three rows of Table 3), the equality of the mean of the BDI scores for the compliers and non-compliers in the treatment group, together with the exclusion restriction (the assumption that the mean BDI score for the non-compliers in the control group is the same as that for those in the treatment group), implies that the compliers in the control group have a worse outcome than the corresponding non-compliers. Attempts to understand why this might be so are detailed in the Discussion.

### The CACE estimation

We now look at simple CACE estimates (i.e. moment estimates based on Equation (3)), ignoring centre membership. These estimates are derived using either of the two definitions of compliance. A negative estimate implies that receipt of treatment works. Using compliance A,  $CACE_{MAR} = -3.47$  (s.e.=2.22). Using compliance B,  $CACE_{MAR} = -2.73$  (s.e.=1.65). The CACE estimates are smaller (i.e. closer to zero) using compliance B than compliance A. For comparison, the ITT effect is just under two units (i.e. the ITT estimate is  $-1.88$ ). None of these differences appears, at this stage, to be statistically significant (the ratio of the estimate to its standard error is  $<2$ ).

We now present the result of a more formal series of analyses (Table 4). We use maximum likelihood estimation (assuming normality of the outcome BDI scores) and

allow for the baseline BDI score as a covariate. All 427 subjects are included in the analysis. They all have data for baseline BDI and centre membership, but 110 of them have a missing 6-month BDI score. Here, we again assume that these missing data are ignorable. All analyses presented in Table 4 are based on the exclusion restriction (allocation to the treatment group has no effect on the non-compliers). Section (a) of Table 4 gives the results of fitting a CACE model in which baseline BDI and centre membership are allowed to predict both compliance and outcome (BDI at 6 months). The model also allows for a treatment  $\times$  centre interaction (i.e. CACE estimates are free to vary from one centre to another). There is variation between centres but note, again, that compliance A leads to greater estimated treatment effects than compliance B.

In section (b) of Table 4 we present the results of separate estimations for problem-solving and psychoeducation. These were obtained by fitting a single model to the complete data-set in which baseline BDI score and centre membership were allowed to predict both compliance and the 6-month BDI score. There were no treatment  $\times$  centre interactions in the model. Fitting a common treatment effect (Table 4, section (c)) indicates that, although problem-solving appears to be slightly more effective than psychoeducation, the difference is nowhere near statistically significant: twice the difference in  $\log L$ , that is  $2 \times (1272.33 - 1272.20)$ , is distributed as  $\chi^2$  with one degree of freedom under the null hypothesis that the two treatments are equally effective. A similar comparison of the  $2\log L$  values for the models in sections (a) and (c) also indicates that the treatment  $\times$  centre interactions are not statistically significant. However, the common treatment effects (using either compliance A or B) in section (c) are statistically significant: by refitting the model after constraining the treatment effects to be zero, the change in  $2\log L$  is 9.32 and 8.06, each with one degree of freedom, for compliances A and B, respectively. Section (d) of Table 4 provides an estimate of the ITT effect obtained by direct estimation in *Mplus*, assuming that missing 6-month BDI scores are ignorable.

### Sensitivity of CACE estimates to assumptions

We now consider the results of our final series of sensitivity analyses. We start by

**Table 4** Maximum likelihood estimation of the complier average causal effect using the expectation maximisation algorithm<sup>1</sup>

|  | Compliance A | Compliance B  |
|--|--------------|---------------|
| <b>(a) Treatment effects by centre<sup>2</sup></b> |              |               |
| logL   | -1268.73     | -1246.15      |
| AIC  | 2593.46      | 2548.31       |
| 1 (estimate (s.e.))                                | 1.02 (3.42)  | 0.24 (3.55)   |
| 2 (estimate (s.e.))                                | -9.96 (3.44) | -10.06 (3.62) |
| 3 (estimate (s.e.))                                | -7.26 (2.68) | -5.80 (2.52)  |
| 4 (estimate (s.e.))                                | -5.22 (2.34) | -3.07 (2.01)  |
| 5 (estimate (s.e.))                                | -6.88 (3.62) | -6.41 (3.62)  |
| 6 (estimate (s.e.))                                | -3.10 (3.94) | -2.42 (3.76)  |
| 7 (estimate (s.e.))                                | -4.25 (3.28) | -4.47 (2.99)  |
| 8 (estimate (s.e.))                                | -2.21 (3.63) | 0.56 (3.02)   |
| <b>(b) Treatment effects by type<sup>3</sup></b>   |              |               |
| logL   | -1272.20     | -1250.58      |
| AIC  | 2588.40      | 2545.17       |
| Problem-solving (estimate (s.e.))                  | -5.48 (2.99) | -3.89 (1.64)  |
| Psychoeducation (estimate (s.e.))                  | -4.45 (2.83) | -3.45 (2.60)  |
| <b>(c) Common treatment effect<sup>4</sup></b>     |              |               |
| logL   | -1272.33     | -1250.61      |
| AIC  | 2586.66      | 2543.21       |
| With covariates (estimate (s.e.))                  | -5.27 (2.05) | -3.79 (1.70)  |
| <b>(d) Intention-to-treat</b>                      |              |               |
| With covariates (estimate (s.e.))                  | -2.34 (0.94) |               |

AIC, Akaike's information criterion.

1. Using baseline Beck Depression Inventory (BDI) and centre as covariates to predict both compliance status (when appropriate) and 6-month BDI score, and assuming an ignorable missing data mechanism.

2. Free parameters: 28.

3. Free parameters: 22.

4. Free parameters: 21.

replacing the exclusion restriction (effect of treatment allocation in the non-compliers is zero) by a series of alternative assumptions: the effect of treatment allocation in the non-compliers varies from -2.5 (beneficial to be allocated to treatment) to +2.5 (beneficial to be allocated to the control group). This procedure was carried out for data using either of the two compliance definitions. In each case the fitted model was equivalent to that in section (c) of Table 4. The rationale for the procedure is explained in detail by Heckman *et al* (1998) and Jo (2002a). Because the overall effect of allocation to treatment is a weighted average of the effect in the compliers (the CACE) and that in the non-compliers, we would expect that fixing the effect of allocation in the non-compliers to a negative value would bring the CACE estimate closer to zero. When the effect in

the non-compliers is -2.5, for example, the modified CACE estimate is -3.18 (s.e.=3.66). On the other hand, when the effect in the non-compliers is fixed at +2.5 the CACE estimate is more marked, at -6.04 (s.e.=1.73). Because we set the fixed values of the effect in the non-compliers between -2.5 and +2.5, the CACE estimates (and their standard errors) move smoothly between these two extremes. Because our working model (section (c) of Table 4) has no treatment  $\times$  centre or treatment  $\times$  baseline BDI interactions, it is possible to relax the exclusion restriction and allow for the effect of treatment allocation to be estimated freely in the non-compliers (Jo, 2002b). For compliance A, the estimated effect for the non-compliers was +1.43 (s.e.=5.83); using compliance B, it was +1.41 (s.e.=13.67). The corresponding CACE estimates were -5.81 (s.e.=3.75)

and -4.13 (s.e.=5.07), respectively. Note that all four of these estimates are quite imprecise. In our final models, we constrained the effects of treatment allocation to be the same for compliers and non-compliers. This might seem strange but it is possible that offering treatment is beneficial but its receipt is not. The resulting joint estimates (-2.51 (s.e.=1.02) and -2.46 (s.e.=1.02) using compliances A and B, respectively) are very similar to the ITT estimate (with similar standard errors) in section (d) of Table 4. We conclude that the CACE estimates are reasonably robust to changes in assumptions and the effect of the receipt of treatment in those who get treated is likely to be somewhere between -5 and -4 points on the BDI scale.

## DISCUSSION

### Technical issues

We have presented methods for the estimation of various average treatment effects in randomised controlled trials in which not everyone complies with the allocated treatment. The trial that we have used to illustrate these methods (ODIN) involved simply allocating patients to be offered psychological treatment or not. The control group were not given access to treatment and therefore the only form of non-compliance possible in this trial was for those offered treatment not to accept the offer or to discontinue treatment once it had been started. In other trials it might be possible for patients allocated to the control group, for example, to get access to treatment outside of the trial. Dunn (2002b) discusses an example like this. The simple methods of CACE estimation such as those involving the use of moments, based on equations (3) and (4) of the present paper, are quite straightforward to apply. The more sophisticated maximum likelihood procedures, however, need more technical expertise and experience. It is straightforward to apply similar statistical methods to binary (depressed/not depressed) outcomes and the simpler approaches are illustrated by Dunn (2002a).

One point that we should stress here is that all analyses, however simple, are vitally dependent on assumptions that might be difficult to justify for a given trial and often can be almost impossible to verify. Some of the assumptions will, however, be much more credible than others. This means that there is no one approach to

the analysis that is obviously the best one. An important component of these estimation methods should be checking wherever possible the sensitivity of the results to various assumptions made. Unfortunately, sensitivity analyses are very rare in practice. In their systematic review of how 89 randomised controlled trials with missing follow-up data dealt with this problem in their estimation of ITT effects, Hollis & Campbell (1999) found that only one report included any attempt at a sensitivity analysis. However, our analysis strategy is presented as an informal suggestion and not a prescription. Our aim is to encourage trial statisticians and others to probe their data in more detail. We emphasise, however, that we are not suggesting that ITT methods be abandoned but that more care should be taken in their use and they should be supplemented by CACE-based methods as described above. The best method of analysis must be dependent on the characteristic of the trial under consideration.

### The challenge of patient preference

One of the major challenges for psychological treatment trials is that the patients cannot be blinded. Therapists need the co-operation and often the active participation of their subjects for the success of the therapy. The preferences and other beliefs of the patients may have an important impact on compliance with an offered treatment and also on the efficacy of the treatment actually received. To date, there are only a few intervention studies that have evaluated whether patient preference for a specific treatment has an effect on treatment outcome (Bedi *et al.*, 2000; Ward *et al.*, 2000). The interpretation of the results of a randomised controlled trial of a psychological intervention is particularly challenging in the presence of these preference effects (Brewin & Bradley, 1989; McPherson & Britton, 2001). A statistical analysis strategy that highlights the effects of preferences, in the present case through concentration on the problems of non-compliance and subsequent loss to follow-up, may rest on challengeable assumptions but the process of making these assumptions and offering them to challenge will lead to a clearer understanding of what we need to concentrate on in interpreting the resulting estimates. It might be particularly helpful to consider the definition

of compliance and what we think the separate effects of an offer of psychological intervention (or failure to offer in the case of the control group) on the compliers and non-compliers might be.

### Does the mere offer of treatment have a therapeutic effect?

One of the key assumptions in the analyses presented in this paper is the exclusion restriction – the assumption that the offer of treatment in itself does not have any effect on outcome. This assumption is necessary to ensure the identifiability of the CACE estimates (i.e. can we get unique estimates from the data?) when we do not have access to baseline covariates. When we have access to covariates, which can be used to predict jointly the compliance and outcome, then when given an appropriate model (Jo, 2002b) we can relax the restriction assumption and actually estimate the effect of offering treatment in the non-compliers. Unfortunately, in the present example the effect was only weakly identified (it was estimated with very large standard errors). Interestingly, however, the estimate of treatment allocation in the non-compliers was positive (i.e. it was slightly harmful to be offered treatment if you were then going to decline the offer). Similar findings were obtained by Jo (2002b) in his reanalysis of the JOBS II trial (Vinokur *et al.*, 1995; Vinokur & Schul, 1997). The JOBS II was a randomised trial to prevent poor mental health and to promote high-quality re-employment among the unemployed. The overall level of compliance with the offered treatment (5 half-day training sessions) was similar to that in the ODIN trial. Jo (2002b) argued that the offer of intervention to the non-compliers is likely to have led to demoralisation arising from their failure to take up the offered treatment. The non-compliers in the control group do not suffer this demoralisation, however, because they have not been offered anything.

In our compliance A, patients who initially accepted the offer of treatment but who subsequently failed to turn up for appointments or discontinued their treatment after having started it were classified along with the refusals as ‘non-compliers’. It could be argued, however, that those who discontinued their treatment were partial compliers who might have received some benefit from the offered intervention. Here

it might be better to think of our complier/non-complier dichotomy as a comparison of patients with high compliance with those of low compliance (Jo, 2002b). If this were indeed the correct interpretation, then we might expect the offer of treatment to have a small beneficial effect in the low compliers and a larger beneficial effect in the high compliers. In our compliance B, however, we put the discontinued patients in with those who attended a full course of treatment. The non-compliers in this case might be labelled accurately as non-compliers, whereas the compliers are a mix of high and low compliers. However, the effect of treatment allocation in the non-compliers was not seen to be beneficial using either compliance A or B, but the CACE estimate was more marked (further from zero) when using compliance A than compliance B. One possible explanation is that the treatment had no more benefit in those who discontinued than in those who refused or failed to turn up for any treatment. In this situation the CACE estimated using compliance A gives us the more realistic treatment effect because that obtained using compliance B will be attenuated towards zero by including the discontinued patients with those who fully complied with the offered therapy.

### Is there evidence of resentful demoralisation

In the ODIN trial the compliers in the control group (i.e. those who would have accepted the treatment if they had been offered it) do worse than the non-compliers. Why? One possible interpretation is that those people who would like help (and would have accepted treatment if offered it) but who are denied access to it because of allocation to the control group suffer from resentful demoralisation (Brewin & Bradley, 1989). They do worse than they would have done if they had never been recruited to the trial. This resentful demoralisation, if present, would lead to the CACE estimate being too optimistic. An alternative interpretation is that the non-compliers are patients who think (on the whole, correctly) that they will get better anyway and therefore do not need the offered treatment (the compliers, on the other hand, are sicker and feel more in need of help). These two interpretations cannot be distinguished from the present data. The design of trials to enable separate estimation of treatment

and preference effects would need a lot of careful thought. A starting point might be the two-stage design proposed by Rücker (1989) – first randomise patients to have a choice or not, and then randomise those without a choice to the competing treatments while, at the same time, allowing those allocated to the choice arm to select their own treatment. Rücker's design, however, is probably impractical because it takes little account of reality (i.e. the proposed analysis assumes complete compliance with the two random allocations and also that there will be complete follow-up data). The so-called patient preference design of Brewin & Bradley (1989), despite its popularity among some clinical researchers, would appear to be a blind alley – it has very little validity from a statistical viewpoint. A useful device might be to seek patient preferences prior to randomisation (Torgerson *et al*, 1996). This would not only provide important information on preference effects but also would lead to better prediction of compliance and more efficient (precise) CACE estimates. Interestingly, investigators in one of the Norwegian centres of the ODIN trial informed us after the above analysis that they had asked patients prior to randomisation about their interest in receiving the treatment (as suggested in Torgerson *et al*, 1996). Those patients who were allocated to the control condition but had expressed an interest in the treatment prior to randomisation appeared to do worse than those who had not (Dalgard & Børve, 2000).

### Concluding remarks

In the interpretation and evaluation of the results of a simple randomised controlled trial such as ODIN one can ask two related and complementary questions: 'What is the effect of offering treatment?' and 'What is the effect of the receipt of treatment?' The former is answered using an ITT estimate of the treatment effect (i.e. the impact of randomisation) and the latter through CACE estimation (i.e. adjusting for non-compliance). The answers to both questions are likely to be interesting and important and it is reasonably straightforward to obtain answers to both. We stress that in promoting the use of CACE estimation we are not advocating that trialists should abandon ITT. This should always be the primary analysis. What we are advocating is that trialists move beyond ITT in order

### CLINICAL IMPLICATIONS

- Estimation of the complier average causal effect (CACE) enables one to evaluate the effect of receipt of treatment in a randomised controlled trial in which a proportion of patients do not comply with their allocated treatment.
- Estimation of CACE should not be seen as an alternative to the pragmatic intention-to-treat (ITT) analysis but as a means of going beyond ITT estimates to seek an explanation for the pragmatic effects.
- The CACE estimates present trial results in a way that is closer to the real world of the practicing clinician than ITT estimates of treatment effects, and therefore may be more clinically relevant.

### LIMITATIONS

- The CACE analysis presented here assumes that the specified intervention is not available to patients outside of the trial condition. However, the methods can be extended easily to cope with more complex situations.
- Estimation of CACE assumes that compliance is a dichotomous (yes/no) condition, whereas patients may have differing degrees of compliance. Again, the methodology can be extended to cope with quantitative compliance–response relationships.
- Estimation of CACE is dependent upon potentially challengeable assumptions that frequently cannot be tested using the data at hand. These challenges should, however, stimulate investigators to come up with more informative designs.

GRAHAM DUNN, PhD, MOHAMMAD MARACY, MSc, Biostatistics Group, School of Epidemiology & Health Sciences, University of Manchester, UK; CHRISTOPHER DOWRICK, FRCGP, Department of Primary Care, University of Liverpool, UK; JOSÉ LUIS AYUSO-MATEOS, MD, Department of Psychiatry, Hospital Universitario de la Princesa, Universidad Autónoma de Madrid, Spain; ODD STEFFEN DALGARD, DrMed, Institute of General Practice and Community Medicine, University of Oslo, Norway; HELEN PAGE, PhD, Department of Psychology, Chester College of Higher Education, UK; VILLE LEHTINEN, MD, National Research and Development Centre for Welfare and Health, Mental Health Unit, Turku, Finland; PATRICIA CASEY, MD, Mater Misericordiae Hospital, University College Dublin, UK; CLARE WILKINSON, FRCGP, Division of General Practice, University of Wales College of Medicine, Wrexham, UK; JOSÉ LUIS VÁZQUEZ-BARQUERO, FRCPsych, Unit for Research into Social Psychiatry, University Hospital 'Marques de Valdecilla', Santander, Spain; GREG WILKINSON, FRCPsych, Department of Psychiatry, University of Liverpool, UK

Correspondence: Graham Dunn, Biostatistics Group, School of Epidemiology & Health Sciences, University of Manchester, Stopford Building, Oxford Road, Manchester M13 9PT, UK. E-mail: g.dunn@man.ac.uk

(First received 28 January 2003, final revision 22 April 2003, accepted 21 May 2003)

to learn more from their data and search for explanations for their primary results.

### ACKNOWLEDGEMENTS

The ODIN project was supported by the EC Biomed 2 Programme (contract no. RDO/18/31), the Spanish Fondo de Investigación Sanitaria (contract no. 96/1978 and 02/10069), the Wales Office of Research and Development (contract no. RC092), the Norwegian Research Council, the Council for Mental Health, the Department of

Health and Social Welfare and the Finnish Pensions Institute of Agricultural Entrepreneurs (contract no. 0339). M.M. is in receipt of a PhD studentship from the Iranian Department of Education and G.D. thanks Booil Jo for generously letting him have manuscript copies of his papers prior to publication.

The ODIN group is composed of academic colleagues and research and administrative staff who have worked on this part of the ODIN project. They include Gail Birkbeck, Trygve Børve, Maura Costello, Pim Cuijpers, Ioana Davies, Nicholas Fenlon, Mette Finne, Fiona Ford, Andres Gomes de Barrio, Claire Hayes, Ann Horgan, Tarja Koffert, Nicola Jones,



Lourdes Lasa, Marja Lehtilä, Catherine McDonough, Erin Michalak, Christine Murphy, Anna Nevra, Teija Nummelin and Britta Sohlman.

## REFERENCES

- Ayuso-Mateos, J. L., Vázquez-Barquero, J. L., Dowrick, C., et al (2001)** Depressive disorders in Europe: prevalence figures from the ODIN study. *British Journal of Psychiatry*, **179**, 308–316.
- Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996)** Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association*, **91**, 444–472.
- Beck, A., Ward, C., Mendelson, M., et al (1961)** An inventory for measuring depression. *Archives of General Psychiatry*, **4**, 561–571.
- Bedi, N., Lee, A., Harrison, G., et al (2000)** Assessing effectiveness of treatment of depression in primary care. *British Journal of Psychiatry*, **177**, 312–318.
- Bloom, H. S. (1984)** Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, **8**, 225–246.
- Brewin, C. & Bradley, C. (1989)** Patient preferences and randomised clinical trials. *BMJ*, **299**, 313–315.
- Dalgard, O. S. & Børve, T. (2000)** Psychoedukativ intervensjon ved depresjon. In *Ny viten om psykisk heise – utbredelse og forebyggende tilka*. Oslo: Norges forskningsråd.
- Dowrick, C., Dunn, G., Ayuso-Mateos, J.-L., et al (2000)** Problem solving treatment and group psychoeducation for depression: multicentre randomised controlled trial. *BMJ*, **321**, 1450–1445.
- Dunn, G. (2002a)** The challenge of patient choice and nonadherence to treatment in randomized controlled trials of counselling or psychotherapy. *Understanding Statistics*, **1**, 19–29.
- (2002b) Estimating the causal effects of treatment. *Epidemiologia e Psichiatria Sociale*, **11**, 206–215.
- Efron, B. & Tibshirani, R. J. (1993)** *An Introduction to the Bootstrap*. London: Chapman & Hall.
- Frangakis, C. E. & Rubin, D. B. (1999)** Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, **86**, 365–379.
- Heckman, J., Smith, J. & Taber, C. (1998)** Accounting for dropouts in evaluations of social programs. *Review of Economics and Statistics*, **80**, 1–14.
- Hirano, K., Imbens, G. W., Rubin, D. B., et al (2000)** Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, **1**, 69–88.
- Hollis, S. & Campbell, F. (1999)** What is meant by intention to treat analysis? Survey of published randomized controlled trials. *BMJ*, **319**, 670–674.
- Jo, B. (2002a)** Model misspecification sensitivity analysis in estimating causal effects of intervention with noncompliance. *Statistics in Medicine*, **21**, 3161–3181.
- (2002b) Estimation of intervention effects with noncompliance: alternative model specifications. *Journal of Educational and Behavioural Statistics*, **27**, 673–709.
- & Muthén, B. (2001) Modeling of intervention effects with noncompliance: a latent variable approach for randomized trials. In *New Developments and Techniques in Structural Equation Modeling* (eds G. A. Marcoulides & R. H. Schumacher), pp. 57–87. Mahwah, NJ: Lawrence Erlbaum Associates.
- Little, R. J. & Yau, L. H. Y. (1998)** Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model. *Psychological Methods*, **3**, 147–159.
- & Rubin, R. B. (2002) *Statistical Analysis with Missing Data* (2nd edn). Hoboken, NJ: John Wiley.
- McPherson, K. & Britton, A. (2001)** Preferences and understanding their effects on health. *Quality in Health Care*, **10** (suppl. 1), i61–i66.
- Muthén, L. K. & Muthén, B. O. (1998–2001)** *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén (see <http://www.statmodel.com/version2.html> for details of Mplus Version 2.12).
- Newcombe, R. G. (1988)** Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur. *Statistics in Medicine*, **7**, 1179–1186.
- Rücker, G. (1989)** A two-stage trial design for testing treatment, self-selection and treatment preference effects. *Statistics in Medicine*, **8**, 477–485.
- Sheiner, L. B. & Rubin, D. B. (1995)** Intention-to-treat and the goals of clinical trials. *Clinical Pharmacology and Therapeutics*, **57**, 6–15.
- Sommer, A. & Zeger, S. L. (1991)** On estimating efficacy from clinical trials. *Statistics in Medicine*, **10**, 45–52.
- StataCorp (2001)** *Stata Statistical Software: Release 7.0*. College Station, TX: Stata Corporation.
- Torgerson, D. J., Klaber-Moffett, J. & Russell, I. T. (1996)** Patient preferences in randomised trials: threat or opportunity? *Journal of Health Services Research and Policy*, **1**, 194–197.
- Vinokur, A. D. & Schul, Y. (1997)** Mastery and inoculation against setbacks as active ingredients in intervention for the unemployed. *Journal of Consulting and Clinical Psychology*, **65**, 867–877.
- , Price, R. H. & Schul, Y. (1995) Impact of the JOBS intervention on unemployed workers varying in risk for depression. *American Journal of Community Psychology*, **23**, 39–74.
- Ward, E., King, M., Lloyd, M., et al (2000)** Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. I: Clinical effectiveness. *BMJ*, **321**, 1383–1388.
- Yau, L. H. Y. & Little, R. J. (2001)** Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association*, **96**, 1232–1243.