

The statistical pitfalls of the partially randomized preference design in non-blinded trials of psychological interventions

ISLA GEMMELL¹ & GRAHAM DUNN²

1 School of Community Based Medicine, The University of Manchester, Manchester, UK

2 Health Methodology Research Group, School of Community Based Medicine, The University of Manchester, Manchester, UK

Key words

partially randomized preference trials, simulation, depression trials, confounders

Correspondence

Isla Gemmell, School of Community Based Medicine, Simon Building, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK.
Telephone +44 (0) 161-306-0524
Email:
isla.gemmell@manchester.ac.uk

Received 5 September 2008;
revised 6 February 2009;
accepted 15 May 2009

Abstract

In a partially randomized preference trial (PRPT) patients with no treatment preference are allocated to groups at random, but those who express a preference receive the treatment of their choice. It has been suggested that the design can improve the external and internal validity of trials. We used computer simulation to illustrate the impact that an unmeasured confounder could have on the results and conclusions drawn from a PRPT. We generated 4000 observations (“patients”) that reflected the distribution of the Beck Depression Index (DBI) in trials of depression. Half were randomly assigned to a randomized controlled trial (RCT) design and half were assigned to a PRPT design. In the RCT, “patients” were evenly split between treatment and control groups; whereas in the preference arm, to reflect patient choice, 87.5% of patients were allocated to the experimental treatment and 12.5% to the control. Unadjusted analyses of the PRPT data consistently overestimated the treatment effect and its standard error. This led to Type I errors when the true treatment effect was small and Type II errors when the confounder effect was large. The PRPT design is not recommended as a method of establishing an unbiased estimate of treatment effect due to the potential influence of unmeasured confounders. *Copyright © 2011 John Wiley & Sons, Ltd.*

Introduction

Random assignment to experimental groups is recognized as the best way to assemble comparison groups for making causal inference (Kleijnen *et al.*, 1997) and for many years the randomized controlled trial (RCT) has been seen as the optimal design for estimating treatment efficacy in

medical experiments (Cochrane, 1972; Pocock, 1983). Randomization of subjects to treatment groups implies that any known and unknown confounders will be equally distributed between the treatment groups. A confounder is a variable that is related to exposure (treatment or control) and is related to outcome but is not on the causal pathway. Thus in an RCT any differences in outcome

between the groups can be reliably attributed to the treatment effect. In observational (non-randomized) studies where the researcher has no control over the allocation of subjects to treatment groups methods of analysis have been developed to estimate and adjust for common sources of bias such as confounding, selection bias and measurement bias (Rothman and Greenland, 1998). However these methods cannot adjust for the presence of unknown or unmeasured sources of bias.

Frequently RCTs are designed as double blind trials where both the patient and the physician are unaware of which treatment the patient is receiving (Jadad, 1998). This methodology reduces bias that may arise due to psychosocial influences where patients who believe they are receiving an effective treatment will improve regardless of the actual effect of the treatment (Silverman and Altman, 1996). In a double blind RCT the placebo effect is equally distributed among treatment groups. However trials of psychological interventions tend to involve both the patient and the physician being aware of the treatment that the patient is receiving. It has been suggested that if the physician prefers a particular treatment they may impart their enthusiasm to the patients and achieve better compliance than one who does not (Korn, 1991). More commonly it is suggested that patients who are allocated to a treatment that they do not wish to receive may suffer resentful demoralization (Bradley, 1998) and become less motivated to follow the treatment protocol and more likely to drop out before completing the treatment regimen. Although these effects are difficult to evaluate (Torgerson and Sibbald, 1998) it is widely believed that patients who receive their preferred treatment may experience greater improvements in outcome than patients who do not receive their preferred treatment. In these instances alternatives to the RCT design have been proposed.

In 1979, Zelen proposed the randomized consent design whereby patients are randomized to treatment groups before informed consent is sought and informed consent is only sought for those patients who are allocated to the experimental treatment (Zelen, 1979). Thus patients who are allocated to the standard treatment do not suffer resentful demoralization as they are unaware that they had an opportunity to receive a new treatment. Patients who are allocated to receive the new treatment but who would prefer to receive the standard treatment can refuse to consent to inclusion in the trial and continue with the standard treatment. However it has been suggested that this design is unethical and that all patients should give informed consent prior to randomization (Last, 2001).

In 1989, Brewin and Bradley proposed the partially randomized preference trial (PRPT) for situations where

randomization was not suitable (Brewin and Bradley, 1989). In the partially randomized preference design patients are asked about treatment preferences before agreeing to participate in the trial. Patients who have no treatment preference are randomized and patients who express a treatment preference are allocated to the treatment of their choice. Thus there are four treatment groups that allow estimation of the treatment effect and the effect of additional motivational factors. A comparison between the randomized arms alone should replicate the results of an RCT and analysis of the non-randomized arms should be treated as an observational study and analysis should include adjustment for all known confounders.

It has been suggested that participants in a PRPT may be more likely to reflect the real world and that the PRPT design can improve the external validity of the trial (TenHave *et al.*, 2003). An additional benefit of the PRPT is that the researcher may be able to estimate the effects of preference on treatment outcome. However the PRPT is subject to the biases of an observational study and may not provide an unbiased measure of treatment effect.

In 1996 Torgerson *et al.* suggested that all consenting patients were randomized but their preferences were recorded so that they could be taken into account in the final analysis. Thus the trial would not be subject to the biases associated with observational studies yet estimates of the effect of receipt of the preferred treatment could be elicited. A test of a statistical interaction between treatment and preference can determine whether patients fare better with the treatment of their choice.

The PRPT design has been used in several trials of psychological interventions in the last decade. A trial of anti-depressant drug treatment and counselling for major depression demonstrated no significant difference between the two groups and no significant effect of preference at eight weeks (Bedi *et al.*, 2000) and 12 months (Chilvers *et al.*, 2001) follow up. A trial of counselling, cognitive-behaviour therapy and usual general practitioner care demonstrated no significant difference between the therapy groups and usual care after 12 months and demonstrated no effect of preference (King *et al.*, 2000; Ward *et al.*, 2000). All of these trials employed the partially randomized preference design and despite the implication that preference may play an important role in the effectiveness of psychological interventions, failed to demonstrate a treatment effect or a preference effect. A study of three anti-depressants that included a preference arm also demonstrated no significant difference in effectiveness between the three groups (Kendrick *et al.*, 2006).

The aim of this paper is to illustrate how data arising from the use of a partially randomized preference design

might lead to biased estimates of treatment effects when there are unmeasured baseline variables giving rise to confounding of treatment and outcome in the preference arm of the trial.

Methods

Using computer simulation we generated 4000 observations (“patients”) that reflected the distribution of the Beck Depression Index (BDI) in trials of treatment for depression (Beck *et al.*, 1961). We used the inverse Chi-square distribution in STATA to create a skewed distribution of integers within the range zero to 63 with a mean around 13 and a standard deviation of around eight. Each BDI score represents the outcome variable that would constitute a patient observation in the hypothetical trial. The distribution four simulations of BDI score with no treatment effect and no confounder effect are shown in Figure 1. For each combination of treatment effect and confounder effect we generate 4000 such distributions.

Half of these patient observations in each simulation were randomly assigned to a RCT design and half were assigned to a PRPT design. In the RCT, patients were evenly split between treatment and control groups; whereas in the PRPT 400 patients had no preference and were randomized to receive either the treatment or the control and 1600 patients expressed a preference. Of these 1600, we allocated 1400 (87.5%) to the group representing those who chose to receive the treatment and 200 (12.5%) to the group who chose to receive the control.

We created a treatment effect which represented a decrease in BDI of between zero and two points and a confounder effect which represented an increase in BDI of between zero and two points. Thus the active treatment was associated with an improvement in outcome and the confounder was associated with a decline in outcome. Treatment effect sizes of up to two BDI are commonly observed in trials of interventions in depression (Bedi *et al.*, 2000; Bower *et al.*, 2003; Chilvers *et al.*, 2001; Ward *et al.*, 2000). The unmeasured confounder was evenly split

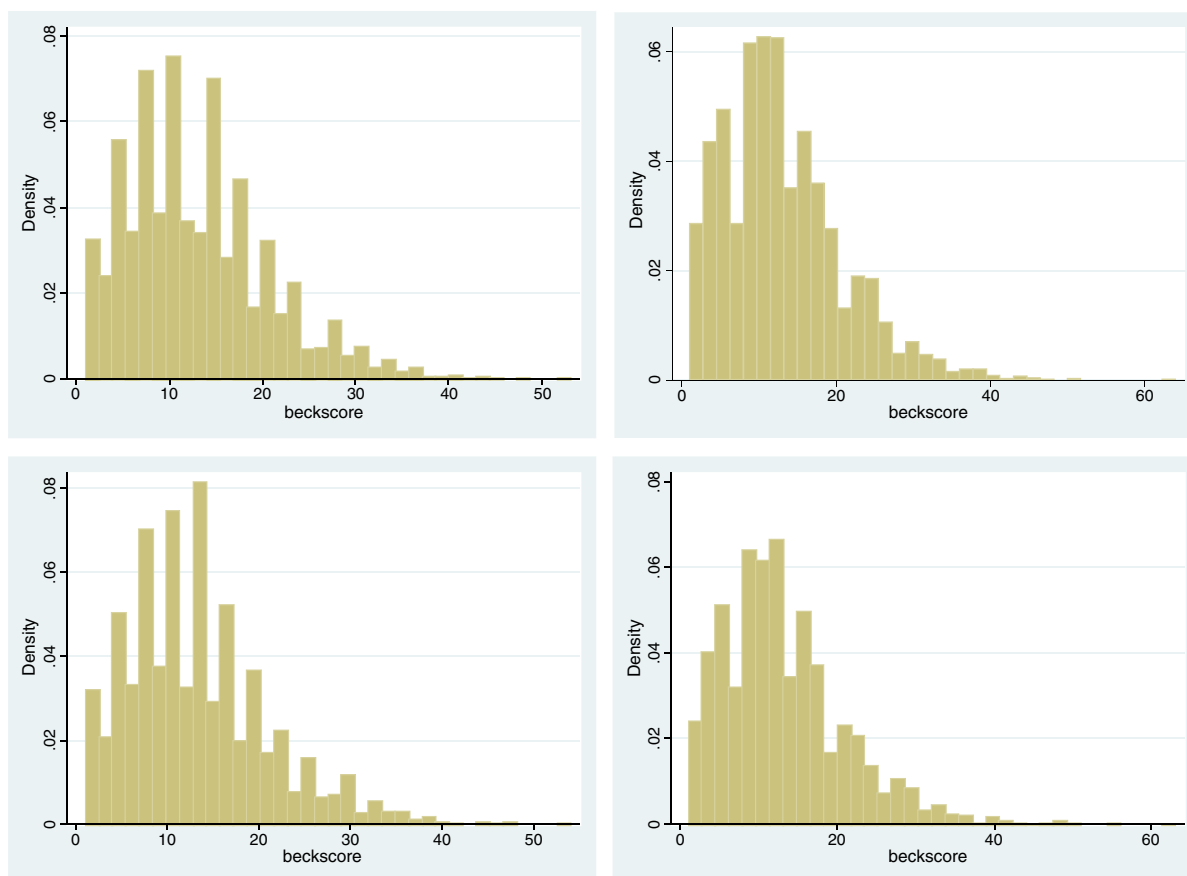


Figure 1 Four simulations of BDI score.

Table 1 Allocation of patients

| | PRPT ($n=2000$) | | | | | |
|---------|-------------------|-------|-----------------------------|-------|----------------------------|-------|
| | RCT ($n=2000$) | | Preference arm ($n=1600$) | | Randomized arm ($n=400$) | |
| | Control | Treat | Control | Treat | Control | Treat |
| Absent | 500 | 500 | 30 | 770 | 100 | 100 |
| Present | 500 | 500 | 170 | 630 | 100 | 100 |
| Total | 1000 | 1000 | 200 | 1400 | 200 | 200 |

between treatment and control groups in the RCT but in the preference arm of the PRPT confounder was more prevalent among patients in the control group. One hundred and seventy out of 200 patients (85%) who received the control had the confounder compared to 630 out of 1400 (45%) of patients who received the active treatment. The data were analysed using linear regression methods with BDI as the outcome variable.

We performed an unadjusted analysis which ignored the unmeasured confounder as this is what would happen in practice if the confounder was unmeasured or unknown. We also conducted an adjusted analysis to demonstrate how the effect of the confounder could be moderated if it had been measured. In the unadjusted analysis the treatment group was the only explanatory variable in the model and in the adjusted analysis both treatment group and the confounder were included in the regression model.

The simulation was repeated 4000 times. This allowed us to record the number of times that the treatment effect was significant and to estimate the mean treatment effect and its standard error for different treatment effect sizes and confounder effect sizes in the RCT and PRPT design.

We carried out a further simulation to determine how a trial using the method of Torgerson *et al.* (1996) would perform. In this analysis we selected only the 2000 patients who were allocated to the PRPT. We allocated the preferences using the same methods as we did for the PRPT trial however we allocated patients randomly to receive either the active treatment or the control regardless of their preference. In this simulation we used a treatment effect of -1 BDI and a confounder effect of $+1$ BDI.

Results

We describe two separate hypothetical trials, one is a standard RCT with 2000 subjects, the other is a partial patient preference trial (PPT) with 2000 patients. In each

trial there is an unmeasured confounder which takes two values, present or absent. The allocation of “patients” is shown in Table 1.

Patients are split evenly between the active treatment or the control in the RCT and unevenly in the PRPT due to patient preference. Thus in the RCT the unmeasured confounder is evenly split between patients who received the control and patients who received the treatment. However in the preference arm of the PRPT the association between patient choice and the unmeasured confounder has resulted in the unmeasured confounder being more prevalent among patients who received the control compared to the treatment. This does not occur in the randomized arm of the PRPT.

Tables 2–4 show the results of analyses for both trials unadjusted and adjusted for the effect of the unmeasured confounder for different sizes of treatment effect and confounder effect.

In the first row of Table 2 there is no effect of either the treatment or the confounder. Using the statistical convention of $t > |1.96|$ or $p < 0.05$ as an indicator of statistical significance we will observe a statistically significant effect of treatment 5% of the time. The average estimate of treatment effect is zero as we would expect and the standard error of the estimate is smallest in the RCT arm with 2000 patients. The randomized arm of the PRPT with 400 patients has a larger standard error associated with the treatment effect due to its small sample size. The mean treatment effect and its standard error are more variable in the randomized arm of the PRPT and this may help to explain why this arm is significant slightly more frequently than the other arms in this example.

When the treatment effect is maintained at zero but the confounder effect is introduced the RCT is significant 5% of the time, the treatment effect remains at zero and the standard error of the treatment effect remains around 0.35. However in an unadjusted analysis of the PPT data the treatment effect is found to be significant in over 10%

Table 2 No treatment effect

| | PRPT ($n=2000$) | | | | | |
|--------------------------------------|-------------------|----------|-----------------------------|----------|----------------------------|----------|
| | RCT ($n=2000$) | | Preference arm ($n=1600$) | | Randomized arm ($n=400$) | |
| | Unadjusted | Adjusted | Unadjusted | Adjusted | Unadjusted | Adjusted |
| <i>Treatment = 0, confounder = 0</i> | | | | | | |
| Percentage significant | 4.72 | 4.70 | 7.88 | 5.13 | 4.80 | 4.78 |
| Treatment effect | 0.004 | 0.004 | -0.000 | -0.003 | 0.011 | 0.011 |
| Standard error | 0.345 | 0.345 | 0.583 | 0.604 | 0.771 | 0.771 |
| <i>Treatment = 0, confounder = 1</i> | | | | | | |
| Percentage significant | 4.53 | 4.55 | 10.73 | 5.22 | 5.50 | 5.63 |
| Treatment effect | 0.006 | 0.006 | -0.409 | -0.010 | 0.008 | 0.008 |
| Standard error | 0.345 | 0.344 | 0.584 | 0.604 | 0.771 | 0.769 |
| <i>Treatment = 0, confounder = 2</i> | | | | | | |
| Percentage significant | 4.45 | 4.70 | 27.68 | 5.15 | 4.80 | 5.05 |
| Treatment effect | 0.006 | 0.006 | -0.795 | 0.007 | -0.008 | -0.008 |
| Standard error | 0.348 | 0.345 | 0.587 | 0.604 | 0.776 | 0.769 |

Table 3 Treatment effect of -1 BDI units

| | PRPT ($n=2000$) | | | | | |
|---------------------------------------|-------------------|----------|-----------------------------|----------|----------------------------|----------|
| | RCT ($n=2000$) | | Preference arm ($n=1600$) | | Randomized arm ($n=400$) | |
| | Unadjusted | Adjusted | Unadjusted | Adjusted | Unadjusted | Adjusted |
| <i>Treatment = -1, confounder = 0</i> | | | | | | |
| Percentage significant | 82.92 | 82.92 | 39.77 | 37.52 | 25.35 | 25.35 |
| Treatment effect | -1.002 | -1.002 | -0.9930 | -0.992 | -0.995 | -0.995 |
| Standard error | 0.345 | 0.345 | 0.583 | 0.604 | 0.770 | 0.770 |
| <i>Treatment = -1, confounder = 1</i> | | | | | | |
| Percentage significant | 82.40 | 82.55 | 66.97 | 37.27 | 25.57 | 25.65 |
| Treatment effect | -0.994 | -0.994 | -1.395 | -0.993 | -1.007 | -1.007 |
| Standard error | 0.345 | 0.345 | 0.584 | 0.604 | 0.771 | 0.769 |
| <i>Treatment = -1, confounder = 2</i> | | | | | | |
| Percentage significant | 82.08 | 82.67 | 86.55 | 38.95 | 25.48 | 25.80 |
| Treatment effect | -0.9933 | -0.9933 | -1.809 | -1.010 | -0.9933 | -0.9933 |
| Standard error | 0.347 | 0.345 | 0.587 | 0.604 | 0.776 | 0.769 |

of the simulations when the confounder effect is one BDI point and in over 27% of the simulations when the confounder effect is two BDI points. The treatment effect is overestimated as -0.4 when the confounder effect is one BDI point and -0.8 when the confounder effect is two BDI points. Thus the patient preference arm of the trial overestimates the treatment effect and as a result is more likely to find a significant effect of treatment when in fact the treatment effect is zero. If the confounder had been

measured and we were able to carry out an adjusted analysis of the preference arm data the treatment effect would be found to be zero and the trial would only give a significant outcome around 5% of the time. The randomized arm of the PRPT also estimates the treatment effect as zero and obtains a significant result around 5% of the time.

Table 3 shows the results of the two separate trials when there is a treatment effect. Patients who receive the treatment experience a one unit decrease in BDI. This

Table 4 Treatment effect of -2 BDI units

| | PRPT ($n=2000$) | | | | | |
|---------------------------------------|-------------------|----------|-----------------------------|----------|----------------------------|----------|
| | RCT ($n=2000$) | | Preference arm ($n=1600$) | | Randomized arm ($n=400$) | |
| | Unadjusted | Adjusted | Unadjusted | Adjusted | Unadjusted | Adjusted |
| <i>Treatment = -2, confounder = 0</i> | | | | | | |
| Percentage significant | 99.97 | 99.97 | 93.22 | 91.38 | 74.42 | 74.45 |
| Treatment effect | -2.010 | -2.010 | -1.997 | -1.993 | -2.010 | -2.010 |
| Standard error | 0.345 | 0.345 | 0.582 | 0.604 | 0.770 | 0.770 |
| <i>Treatment = -2, confounder = 1</i> | | | | | | |
| Percentage significant | 100.0 | 100.0 | 98.33 | 90.53 | 73.58 | 73.60 |
| Treatment effect | -1.996 | -1.996 | -2.390 | -1.986 | -1.995 | -1.995 |
| Standard error | 0.345 | 0.345 | 0.584 | 0.604 | 0.771 | 0.770 |
| <i>Treatment = -2, confounder = 2</i> | | | | | | |
| Percentage significant | 100.0 | 100.0 | 99.75 | 90.75 | 72.65 | 73.20 |
| Treatment effect | -2.003 | -2.003 | -2.793 | -1.992 | -1.994 | -1.994 |
| Standard error | 0.348 | 0.345 | 0.587 | 0.604 | 0.776 | 0.770 |

analysis demonstrates that the RCT and the randomized arm of the PRPT correctly estimate the treatment effect as -1 . The size of the confounder does not affect the outcome of the RCT and the trial is significant between 82% and 83% of the time regardless of whether the confounder effect is zero, one or two BDI points. In the randomized arm of the PRPT the trial is only significant around 25% of the time, this is due to lack of power. The treatment effect is correctly estimated as -1 but the average standard error of the estimates is large and therefore statistical significance is reached less often.

In an unadjusted analysis of the preference arm of the PRPT the treatment effect is overestimated as -1.4 when the confounder effect is one BDI point and as -1.8 when the confounder effect is two BDI points. The standard error of the estimate is larger than for the RCT and combined with the biased estimate of treatment fewer trials achieve statistical significance when the confounder is zero but a greater number of trials achieve statistical significance when the confounder effect is two BDI points.

When the treatment effect is increased to two BDI points (Table 4) the RCT trials are significant around 100% of the time, the treatment effect is correctly estimated as -2 and the standard error of the estimate is around 0.345. In the unadjusted analysis of the patient preference arm of the PRPT the treatment effect is overestimated as -2.4 when the confounder effect is one BDI point and as -2.8 when the confounder effect is two BDI points.

The average standard error of the estimates is greater than that observed in the RCT which results in the

overestimate of the treatment effect being less likely to reach statistical significance. When we adjust the analysis of the patient preference arm for the effect of the unmeasured confounder we correctly estimate the treatment effect as -2 BDI points regardless of the size of the confounder effect but the trials are less likely to attain statistical significance due to the larger standard errors associated with the estimates in the adjusted analysis. In the randomized arm of the PRPT the treatment effect is correctly estimated as -2 BDI points but the trials are less frequently significant (73–74% of the time) due to the larger standard errors associated with the estimates as a result of the smaller numbers in the trials.

In summary, the unadjusted analysis of the patient preference arm of the trial consistently overestimated the effect of treatment in the presence of an unmeasured confounder that was associated with treatment choice. When the effect of the confounder was larger than the effect of treatment this led to Type I errors and when the effect of the confounder was smaller than the effect of treatment this led to Type II errors. Adjusted analysis improved the estimate of treatment effect in the preference arm but led to a slight increase in Type II errors due to inflation of the standard error associated with the treatment effect. The RCT arm of the PRPT correctly estimated the size of the treatment effect but was prone to Type II errors due to the small sample size.

In a further simulation we randomly allocated subjects in the PRPT to either treatment or control but we recorded their preference which remained the same as for

Table 5 RCT with preference effects recorded

| | Model 1 | Model 2 | Model 3 |
|------------------------|---------|---------|---------|
| <i>Treatment = -1</i> | | | |
| Percentage significant | 81.78 | 81.85 | 64.50 |
| Treatment effect | -0.9904 | -0.9904 | -0.9906 |
| Standard error | 0.3454 | 0.3447 | 0.4220 |
| <i>Confounder = 1</i> | | | |
| Percentage significant | | 83.67 | 67.10 |
| Confounder effect | | 1.0107 | 1.0079 |
| Standard error | | 0.3447 | 0.4220 |
| <i>Preference = 0</i> | | | |
| Percentage significant | | | 4.97 |
| Preference effect | | | -0.0006 |
| Standard error | | | 0.6892 |

the PRPT. The results of different methods of analysis of these data are shown in Table 5. In Model 1 only treatment effect is estimated, Model 2 adjusts for the effect of the unmeasured confounder and Model 3 adjusts for the unmeasured confounder and preference. This model would be used to determine whether there was a true preference effect and whether people who received the treatment of their choice had a difference outcome to those who did not receive the treatment of their choice over and above the actual treatment effect.

In the first model the size of the treatment effect is correctly estimated and a significant effect of treatment was found in 82% of trials. In the second model which adjusts for the effect of the confounder, despite the fact that randomization will have adjusted for the confounder anyway, the treatment effect is again correctly estimated as -1 and the effect of the confounder is correctly estimated as one. The treatment effect is significant in 82% of trials and the confounder effect is significant in 84% of the trials.

In Model 3 the effect of preference is estimated as zero and it is significant 6% of the time. We have not included a true effect of preference in the simulation and this is established by the results from Model 3. However by including an additional term for preference in the model we have increased the standard errors and thus made it less likely that we find a significant effect of treatment.

Discussion

The analyses presented here demonstrate how the PRPT can be subject to bias due to the presence of unmeasured or unknown confounders. However this design has been suggested for use in trials of complex interventions in

psychiatry where it is difficult to conduct an RCT as both patients and clinicians are aware of treatment allocation. Our analysis focused only on two parameters, the treatment effect size and the effect of an unmeasured dichotomous confounder and demonstrated that by failing to randomize patients to treatment groups we may arrive at an invalid conclusion about the effects of a particular treatment option. In our example the treatment effect was overestimated due to the association between treatment choice and the unmeasured confounder. Other parameters such as the distribution of patients among the treatment and confounder arms in the PRPT, modelling the treatment and confounder effects as random variables and modelling the confounder as a continuous variable may have identified further biases associated with the PRPT design.

The PRPT design is infrequently used and at present the trials that have been conducted for psychological interventions in depression have failed to demonstrate a treatment effect in either the randomized arm or the preference arm of the trial (Bedi *et al.*, 2000; Kendrick *et al.*, 2006; Ward *et al.*, 2000). However, PRPTs have been recommended as they may improve both the internal validity and the external validity of clinical trials. It is suggested that they improve internal validity because the distribution of patients among treatment groups more closely reflects the real life situation. Yet other authors have suggested that after accounting for baseline differences there is little evidence that patient preference has any effect on internal validity (King *et al.*, 2005). Yet there is evidence that treatment preference can lead to a substantial proportion of people refusing to enter a RCT (King *et al.*, 2005; Torgerson and Sibbald, 1998; van Schaik *et al.*, 2004) and this may affect the external validity of the trial. This may be exaggerated in trials of psychological interventions due to the nature of the intervention.

In order to measure both internal and external validity, RCTs in psychiatry could include measurement of preferences and could record details of those who refuse to take part in the study due to preferences. These data could be used to determine the degree to which the trial participants represent the general population. Furthermore, if patient preference was recorded at the start of a RCT, preference could be taken into account in the final analysis. The patients in the trial would be randomized to treatment groups yet the analysis would enable us to identify preference effects by testing whether those who were randomized to the treatment that they preferred achieved a greater improvement in outcome compared to those who were randomized to the treatment they did not prefer. This design was first suggested by Torgerson *et al.*

(1996) and has since been used in several trials. However, as our analysis has demonstrated, including more terms in the model can lead to higher standard errors and therefore reduce the chance of finding treatment effect statistically significant. In a PRPT of the anti-depressant medication or psychotherapy for patient with depression in the United States (Lin *et al.*, 2005) there was slight evidence of a treatment effect after nine months but there was no evidence of an interaction between treatment and preference suggesting that those patients who received their preferred treatment achieved no additional improvement in depression score. This design was also used in a RCT of exercise and cognitive behavioural therapy for back pain. Treatment preference was recorded prior to randomization and the researchers found some evidence of an interaction between patient preference and outcomes (Johnson *et al.*, 2007). However, care must be taken when implementing this design as the process of asking a patient's preference may make them less likely to consent to randomization.

It has been suggested that RCTs give biased results since they are not representative of the real world as they only include participants who are willing to be randomized. Yet, whilst PRPTs may more closely resemble the real world they are subject to bias due to the non-random allocation of subjects to treatment groups. In fact it is in some ways misleading to call the PRPT a trial as it is simply an observational study and is subject to the potential biases of observational studies.

Instrumental variable methods (Greenland, 2000) have been used to adjust for non-compliance and loss to follow up in RCTs (Dunn and Bentall, 2007; Dunn *et al.*, 2005). These methods could be used to adjust for the presence of unknown or unmeasured confounders in PRPTs. In a PRPT an instrumental variable is defined as a variable that is associated with treatment choice but is not associated with outcome. In a trial of counselling versus anti-depressant treatment an instrumental variable may be proximity to counselling service. Patients who live closer to the counselling may be more likely to choose counselling than those who live further away. Instrumental variable

regression methods have been developed to provide an estimate of treatment effect (StataCorp, 2005) that is adjusted for some of the bias associated with the patient preference design however it is often difficult to identify a suitable instrumental variable.

More recently methods of estimating causal effects and preference effects in doubly randomized preference trials (DRPTs) have been proposed (Long *et al.*, 2008). In a DRPT patients are initially randomized to a randomization arm in which treatments are randomized or to a preference arm in which patients choose which treatment they receive. The authors demonstrated limited evidence of a benefit of patients receiving their preferred treatment however they did demonstrate strong preference effects on treatment adherence.

Conclusion

Partially randomized patient preference designs are subject to the biases associated with observational studies and should be avoided unless techniques to adjust for unmeasured potential biases are employed (Hofler *et al.*, 2007; Long *et al.*, 2008; StataCorp, 2005). However, collecting data on patient preference may be useful in trials of psychiatric interventions (Howard and Thornicroft, 2006) and RCTs in psychiatry could include a measure of preference and should record the characteristics of people who refused to participate due to random allocation of treatment. This will enable preference effects to be measured at the analysis stage and will enable researchers to estimate the external validity of the trial.

Acknowledgement

This work was funded by the UK Mental Health Research Network (MHRN) and at the time that the work was carried out both authors were members of the MHRN.

Declaration of interest statement

The author have no competing interests.

References

- Beck A., Ward C., Mendelson M. (1961) An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571.
- Bedi N., Chilvers C., Churchill R., Dewey M., Duggan C., Fielding K., Gretton V., Miller P., Harrison G., Lee A., Williams I. (2000) Assessing effectiveness of treatment of depression in primary care – partially randomised preference trial. *British Journal of Psychiatry*, 177, 312–318.
- Bower P., Byford S., Barber J., Beecham J., Simpson S., Friedli K., Corney R., King M., Harvey I. (2003) meta-analysis of data on costs from trials of counselling in primary care: using individual patient data to overcome sample size limitations in economic analysis. *British Medical Journal*, 326, 1247–1250.
- Bradley C. (1998) Designing medical and educational intervention studies. *Diabetes Care*, 16, 509–518.
- Brewin C., Bradley C. (1989) Patient preferences and randomised clinical trials. *British Medical Journal*, 299, 313–315.

- Chilvers C., Dewey M., Fielding K., Gretton V., Miller P., Palmer B., Weller D., Churchill R., Williams I., Bedi N., Duggan C., Lee A., Harrison G. (2001) Antidepressant drugs and generic counselling for treatment of major depression in primary care: randomised trial with patient preference arms. *British Medical Journal*, **322**(7289), 772–775.
- Cochrane A. (1972) *Effectiveness and Efficiency: Random Reflections on Health Services*, London, Nuffield Provincial Hospitals Trust.
- Dunn G., Bental R. (2007) Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine*, **26**(26), 4719–4745.
- Dunn G., Maracy M., Tomenson B. (2005) Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Statistical Methods in Medical Research*, **14**(4), 369–395.
- Greenland S. (2000) An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, **29**, 722–729.
- Hofler M., Lieb R., Wittchen H.U. (2007) Estimating causal effects from observational data with a model for multiple bias. *International Journal of Methods in Psychiatric Research*, **16**(2), 77–87.
- Howard L., Thornicroft G. (2006) Patient preference randomised controlled trials in mental health research. *British Journal of Psychiatry*, **188**, 303–304.
- Jadad A. (1998) *Randomised Controlled Trials*, 1st edition, London, BMJ Books.
- Johnson R.E., Jones G.T., Wiles N.J., Chaddock C., Potter R.G., Roberts C., Symmons D.P.M., Watson P.J., Torgerson D.J., Macfarlane G.J. (2007) Active exercise, education, and cognitive behavioral therapy for persistent disabling low back pain – a randomized controlled trial. *Spine*, **32**(15), 1578–1585.
- Kendrick T., Peveler R., Longworth L., Baldwin D., Moore M., Chatwin J., Thornett A., Goddard J., Campbell M., Smith H., Buxton M., Thompson C. (2006) Cost-effectiveness and cost-utility of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine – randomised controlled trial. *British Journal of Psychiatry*, **188**, 337–345.
- King M., Nazareth I., Lampe F., Bower P., Chandler M., Morou M., Sibbald B., Lai R. (2005) Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials. *Health Technology Assessment*, **9**(35), 1.
- King M., Sibbald B., Ward E., Bower P., Lloyd M., Gabbay M., Byford S. (2000) Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. *Health Technology Assessment*, **4**(19), 1–83.
- Kleijnen J., Gotzsche P., Kunz R., Oxman A., Chalmers I. (1997) So what's so special about randomisation? In Maynard A., Chalmers I. (eds) *Non-random Reflections on Health Services Research*, London, BMJ Books.
- Korn E. (1991) Randomised clinical trials with clinician-preferred treatment. *Lancet*, **337**, 149–152.
- Last J.M. (2001) *A Dictionary of Epidemiology*, 4th edition, Oxford, Oxford University Press.
- Lin P., Campbell D.G., Chaney E.E., Liu C.F., Heagerty P., Felker B.L., Hedrick S.C. (2005) The influence of patient preference on depression treatment in primary care. *Annals of Behavioral Medicine*, **30**(2), 164–173.
- Long Q., Little R., Lin X. (2008) Causal inference in hybrid intervention trials involving treatment choice. *Journal of the American Statistical Association*, **103**(482), 474–484.
- Pocock S. (1983) *Clinical Trials: A Practical Approach*, Chichester, John Wiley & Sons.
- Rothman K., Greenland S. (1998) *Modern Epidemiology*, 2nd edition, Philadelphia, PA, Lippincott Williams & Wilkins.
- Silverman W., Altman D. (1996) Patient's preferences and randomised trials. *Lancet*, **347**, 171–174.
- StataCorp (2005) *Stata Statistical Software: Release 9*, StataCorp LP, College Station, TX.
- TenHave T.R., Coyne J., Salzer M., Katz I. (2003) Research to improve the quality of care for depression: alternatives to the simple randomized clinical trial. *General Hospital Psychiatry*, **25**(2), 115–123.
- Torgerson D., Klaber-offett J., Russell I. (1996) Patient preferences in randomised trials: threat or opportunity. *Journal of Health Services Research & Policy*, **1**(4), 194–197.
- Torgerson D., Sibbald B. (1998) What is a patient preference trial. *British Medical Journal*, **316**, 360.
- van Schaik D.J.F., Klijn A.F.J., van Hout H.P.J., van Marwijk H.W.J., Beekman A.T.F., de Haan M., van Dyck R. (2004) Patients' preferences in the treatment of depressive disorder in primary care. *General Hospital Psychiatry*, **26**(3), 184–189.
- Ward E., King M., Lloyd M., Bower P., Sibbald B., Farrelly S., Gabbay M., Tarrier N., Addington-Hall J. (2000) Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. I: Clinical effectiveness. *British Medical Journal*, **321**(7273), 1383–1388.
- Zelen M. (1979) A new design for randomised clinical trials. *The New England Journal of Medicine*, **300**, 1242–1245.