

Effective elements of cognitive behaviour therapy for psychosis: results of a novel type of subgroup analysis based on principal stratification

G. Dunn^{1*}, D. Fowler², R. Rollinson³, D. Freeman⁴, E. Kuipers⁴, B. Smith^{1,5}, C. Steel⁶, J. Onwumere⁴, S. Jolley⁴, P. Garety^{4†} and P. Bebbington^{5†}

¹ Health Sciences Research Group, School of Community-Based Medicine, University of Manchester, UK

² School of Medicine, Health Policy and Practice, University of East Anglia, Norfolk, UK

³ Norfolk and Waveney Mental Health Partnership Trust, UK

⁴ Department of Psychology, Institute of Psychiatry, King's College London, UK

⁵ Department of Mental Health Sciences, UCL, London, UK

⁶ Department of Psychology, University of Reading, UK

Background. Meta-analyses show that cognitive behaviour therapy for psychosis (CBT-P) improves distressing positive symptoms. However, it is a complex intervention involving a range of techniques. No previous study has assessed the delivery of the different elements of treatment and their effect on outcome. Our aim was to assess the differential effect of type of treatment delivered on the effectiveness of CBT-P, using novel statistical methodology.

Method. The Psychological Prevention of Relapse in Psychosis (PRP) trial was a multi-centre randomized controlled trial (RCT) that compared CBT-P with treatment as usual (TAU). Therapy was manualized, and detailed evaluations of therapy delivery and client engagement were made. Follow-up assessments were made at 12 and 24 months. In a planned analysis, we applied principal stratification (involving structural equation modelling with finite mixtures) to estimate intention-to-treat (ITT) effects for subgroups of participants, defined by qualitative and quantitative differences in receipt of therapy, while maintaining the constraints of randomization.

Results. Consistent delivery of full therapy, including specific cognitive and behavioural techniques, was associated with clinically and statistically significant increases in months in remission, and decreases in psychotic and affective symptoms. Delivery of partial therapy involving engagement and assessment was not effective.

Conclusions. Our analyses suggest that CBT-P is of significant benefit on multiple outcomes to patients able to engage in the full range of therapy procedures. The novel statistical methods illustrated in this report have general application to the evaluation of heterogeneity in the effects of treatment.

Received 31 January 2011; Revised 15 August 2011; Accepted 30 August 2011; First published online 23 September 2011

Key words: Cognitive behavioural therapy, principal stratification, psychosis, subgroup analysis, treatment-effect heterogeneity.

Introduction

The pharmacological treatment of positive symptoms of psychosis is only moderately successful (Leucht *et al.* 2009). Cognitive behaviour therapy for psychosis (CBT-P) consistently reduces psychotic symptoms in people with distressing medication-resistant symptoms, although average effect sizes are fairly small (Pilling *et al.* 2002; Jones *et al.* 2004; Zimmerman *et al.* 2005; Wykes *et al.* 2008). National Institute of Clinical

and Health Excellence updated guidelines for schizophrenia recommend CBT-P (NICE, 2009).

The Psychological Prevention of Relapse in Psychosis (PRP) trial was designed to evaluate the effectiveness of CBT-P in reducing relapse and improving symptoms. The PRP trial compared CBT-P and family intervention (FI) with treatment as usual (TAU), and is fully described elsewhere (Garety *et al.* 2008). Intention-to-treat (ITT) analysis found no benefits for the primary outcomes of relapse and days in hospital or for reduction in psychotic symptoms, but did find a significant improvement in depression (Garety *et al.* 2008). The trial manual describes cognitive and behavioural techniques targeting the various symptoms and problems presented by individuals.

* Address for correspondence: Professor G. Dunn, Health Sciences Methodology, 1st Floor, Jean McFarlane Building, University Place, Oxford Road, Manchester M13 9PL, UK.

(Email: graham.dunn@manchester.ac.uk)

† These authors contributed equally to this work.

Therapists shape the techniques to the particular problems that emerge during therapy (Fowler *et al.* 1995). Therapy therefore varies widely, being tailored to the individual needs of a heterogeneous group of clients with differing levels of capacity and willingness to engage.

However, CBT-P has so far only been evaluated as an overall package. We lack information about the delivery of different therapeutic techniques, given that clients may not be ready or able to countenance particular interactions (Durham *et al.* 2003). Single case studies have suggested that it is the more active techniques that lead to symptomatic changes, rather than the necessary but preliminary stage of relationship building and assessment (Fowler & Morley, 1989; Chadwick *et al.* 2003).

The present study investigates how far competent CBT therapists were able to deliver different types of therapy techniques, and the impact this has on efficacy. Reliable methods for monitoring therapeutic delivery in CBT-P allow us to examine its relationship with outcome (Startup *et al.* 2002; Durham *et al.* 2003; Rollinson *et al.* 2007, 2008). Based on our cognitive model (Garety *et al.* 2001), we hypothesized that outcomes would be improved when therapists were able to deliver the more specific cognitive and behavioural techniques.

A key feature of this study is our novel statistical approach for analysing differential efficacy in randomized trials, informed by increasing recognition of the biases and confounding inherent in past attempts at *post-hoc* estimation of outcomes in relation to aspects of therapeutic quality (Dunn & Bentall, 2007). The new approach estimates ITT effects for subgroups by comparing the effects of intervention with putative effects in the control arms ignored in traditional analyses. The evaluation of differential efficacy of CBT-P techniques formed part of the original protocol.

Method

Settings

The trial took place in five mental health services: two in inner-city London, one in suburban outer London, one in a provincial city (Norwich), and one in a rural area (Norfolk).

Study design

The PRP trial comprised two pathways with separate randomization, stratified within the five participating centres, and within in-patient or out-patient status at induction. The first ('individual pathway') included participants without carers randomly allocated to two groups: both received good standard care (treatment

as usual, TAU) whereas the experimental group also received CBT-P. In the second pathway ('carer pathway'), those with carers were allocated to three groups: CBT-P plus TAU, FI plus TAU, or TAU alone. The current analysis is restricted to hypotheses concerning CBT-P only, so the FI participants were excluded.

Participants

We approached consecutive patients with recent relapses, whether or not they had been admitted. They were invited to take part once they could give informed consent. The inclusion criteria were: current clinical diagnosis of non-affective psychosis (F2: ICD-10; WHO, 1992; DSM-IV; APA, 1994); age 18–65 years; a second or subsequent psychotic episode starting not more than 3 months before induction; and a rating of at least 4 (moderate severity) for at least one positive symptom on the Positive and Negative Syndrome Scale (PANSS; Kay, 1991). Exclusion criteria were: a primary diagnosis of alcohol or substance dependency, organic syndrome or learning disability; spoken English inadequate for engaging in psychological therapy; and unstable residential arrangements.

Participants provided informed consent under protocols approved by the appropriate ethics committees. Full details of the trial protocol are provided elsewhere (Garety *et al.* 2008). Participants were assessed at baseline before randomization, and at 3, 6, 12 and 24 months. The CBT-P was completed within 12 months whereas TAU continued throughout.

Treatment

CBT-P was delivered for 9 months, with a planned minimum of 12 and a maximum of 20 sessions. The therapy in our generic CBT-P manual (Fowler *et al.* 1995) was augmented with specific relapse prevention techniques.

Therapy provision

One hundred and thirty-three people were allocated to CBT-P. They received a mean of 14.3 sessions (S.D. = 7.8), each lasting on average 1 h. The number of sessions was very similar in the individual and carer arms.

Trial therapists: training and monitoring of adherence and competence

Five lead trial therapists ('lead therapists'), all doctorate level or equivalent clinical psychologists employed full time on the trial, provided therapy to

96 CBT-P participants. A further 37 CBT-P participants were seen by therapists employed by the National Health Service (NHS) Trusts running the local mental health services ('trust therapists'). The trust therapists were doctoral clinical psychologists and nurses with specialist training in CBT-P. All were fully trained and closely managed and supervised; details of recruitment, training and quality control are provided elsewhere (Garety *et al.* 2008). The Revised Cognitive Therapy for Psychosis Adherence Scale (R-CTPAS; Rollinson *et al.* 2008) is a measure of fidelity, designed to provide precise definitions of the minimum therapeutic delivery of CBT-P activity. It covers 21 different types of CBT-P techniques. Therapist competence was measured by the Cognitive Therapy Scales (CTS; Young & Beck, 1980). All raters were trained to criterion on this scale and met regularly to check rating reliability. A total of 185 tapes from 66 therapy participants (62% of the total treated) were sent for formal monitoring by the lead therapists from other centres. In 90% of the sample, the CBT-P delivered in taped interviews was both adherent and competent. In eight cases (8.3%), the therapy was regarded as supportive work rather than CBT-P. A randomly selected subsample of 36 tapes was sent to external expert raters; their ratings showed excellent agreement with the internal raters (Garety *et al.* 2008).

All therapists also used the R-CTPAS to provide self-report assessments of their therapy sessions. Agreement between tape-rated and self-reported ratings of R-CTPAS across multiple raters was satisfactory, with intraclass correlation coefficients for composite scores ranging from 0.5 to 0.8.

Summary scores of therapy delivery

In the current study, the taped and self-reported adherence ratings from the R-CTPAS (Rollinson *et al.* 2008) were used to create a single summary score for the therapy received by each person treated in the trial. These were derived from the factor analysis of the R-CTPAS, described above. The first factor, which we term 'partial therapy', comprised engagement and assessment techniques: that is, active attempts to engage in therapeutic strategies; the 'Columbo style' (which assesses the degree to which therapists promote guided discovery); and the collaborative assessment of psychotic experience and delusional beliefs. The second factor, termed 'full therapy', comprised active therapy techniques: that is, relapse prevention interventions; enhancing self-regulatory strategies; developing a personal model of relapse; developing a model of psychosis; work on reinterpreting the meaning of delusional beliefs and hallucinations; and schema work. Both factors

described components of active therapy, and the first factor should not be confused with befriending (Sensky *et al.* 2000).

Our intention was to ensure that this overall assessment of therapy techniques was reliable. We used data from 1019 sessions from 102 participants where there were sufficient R-CTPAS data to be fully representative of level of therapy. Thirty-one participants had significant levels of missing individual session data and were coded in the present analysis as 'not known'. There were no significant differences on the baseline variables of interest between the subsample analysed and those with missing data. There was a good spread of sessions across the whole duration of therapy: 30% of the sessions evaluated came from block 1 (sessions 1–4), 26% from block 2 (sessions 5–9), 22% from block 3 (sessions 10–14) and 21% from block 4 (sessions 15 and above). Average item scores from all available sessions were calculated for each participant. The R-CTPAS manual uses a score of 1 (within a range from –7 to +7) to indicate the minimum threshold for highly competent delivery of individual techniques. For an individual technique to be considered present across the course of therapy, the averaged item score needed to be one or above where self-report data were available, or to be judged above the competence threshold in at least three sessions, supported by a tape. This aimed to reflect the definite presence of a therapy technique occurring across the course of therapy, and was deliberately chosen to signify the unequivocal delivery of high quality interventions.

For full or partial therapy to be considered present, at least one of the composite active intervention or engagement and assessment items listed above needed to be present above this threshold across the course of therapy. Clients who received less than five therapy sessions formed a third, no-therapy, group, as this number of sessions was regarded as too small for the delivery of effective CBT-P.

Control condition

TAU consisted of good standard care, delivered according to national and local service protocols and guidelines, including the provision of antipsychotic medication. The frequency and nature of service contacts was monitored, as were medication regimes. TAU did not preclude the provision of psychological interventions by locality teams, although this was unusual.

Primary outcome measures

The primary outcome variable, relapse, was assessed by a blind panel evaluation procedure (Craig *et al.*

2004; Bebbington *et al.* 2006). Consensus remission and relapse ratings were applied to detailed extracts of the clinical case-notes by paired members of the research team, using manualized *a priori* operational definitions (Bebbington *et al.* 2006). The original trial report gives full details (Garety *et al.* 2008). Here we present the data as the total number of months in full remission separately over the first and second years of the trial. Data on hospital admissions were collected through the hospital administration systems.

Secondary outcome measures

Secondary outcomes were rated by research assessors at interview, and again considerable efforts were made to achieve blind ratings (Garety *et al.* 2008). The measures used were the PANSS (APA, 1994) and the Beck Depression Inventory Second Edition (BDI-II; Beck *et al.* 1996). The PANSS is a 30-item, seven-point (1–7) rating instrument assessing psychotic symptoms over the past week. We present results for the PANSS Total (30 items) and PANSS Positive scores (seven items). The BDI-II is a self-report, 21-item, four-point scale (0–3) for the assessment of depression over the past 2 weeks.

Statistical analysis

All analyses reported in the main trial paper (Garety *et al.* 2008) were based on the ITT principle, allowing for potential biases arising from loss to follow-up [under the assumption that missing outcomes were missing at random (MAR) using the terminology of Little & Rubin (2002)].

All analyses presented in the current paper involve estimating ITT effects within three classes of participant. These three classes (principal strata; Angrist *et al.* 1996; Frangakis & Rubin, 2002) are defined by the potential outcome of participants' treatment allocation. Stratum 1 (no therapy) comprises participants who would receive little or no therapy (CBT-P) regardless of their randomized allocation. Stratum 2 (partial therapy) comprises those who received partial therapy in the CBT-P group, together with those controls who would have received partial therapy if they had been allocated to the CBT-P condition. Finally, Stratum 3 (full therapy) comprises those participants who received full therapy in the CBT-P group, together with those controls who would have done so, had they been allocated to CBT-P. These three strata are only partly identified: class membership is known for most participants allocated to CBT-P, but not for the controls, and the model is correspondingly said not to be identified; that is, unique stratum-specific treatment (ITT) effects cannot be derived. However,

it is possible to identify empirically baseline covariates that predict the type of treatment delivered in the randomized-in participants. Because of randomization, these can also be used to predict potential treatment compliance in the randomized group. In the present study, the best predictors were treatment centre (location), presence of a carer, in-patient status and sex of the patient. There was no association between baseline symptomatology and the type of therapy received.

We used the predictors as covariates in a latent class model to predict principal stratum membership. The same covariates were used in the simultaneously fitted analysis of covariance (ANCOVA) model used to estimate the stratum-specific treatment (ITT) effects. Model identification was further improved by assuming that the ITT effect in the no-therapy stratum was zero (i.e. allocation to CBT-P has no average effect when the participant fails to take up the offered therapy). This is a so-called exclusion restriction.

A further refinement is based on the realization that the probability of participants having missing outcome data (i.e. loss to follow-up) is likely to be dependent on stratum membership; the no-therapy group, for example, would seem less likely to provide outcome data than those from the other two strata. The missing data mechanism might still be MAR, but it might equally be latently ignorable (LI) (Frangakis & Rubin, 1999). In the LI model, the probability of loss to follow-up is jointly dependent on stratum membership and the outcome of random allocation, and also on baseline covariates (the structure of the missing data model then being analogous to that for the outcomes). The exclusion restrictions for the missing data indicator were the same as for the final outcomes. Technical details and illustrations using data from psychological treatment trials are provided elsewhere (Dunn *et al.* 2005; Emsley *et al.* 2010).

All analyses reported in the present paper were carried out using Mplus version 5.2 (Muthén & Muthén, 1998–2009). To avoid local maxima (invalid estimates), 1000 randomly perturbed sets of starting values were used. All standard errors were estimated using bootstrapping (250 replications) (Efron & Tibshirani, 1993). Note that in none of the analyses have we allowed for individual therapist effects (clustering by therapist); for technical reasons it would have not been feasible, assuming instead that these are subsumed by the centre effects.

Results

In all participants receiving full therapy, the techniques associated with partial therapy were also

delivered. In every case, at each level of therapy, multiple techniques were present. Forty-two participants had full therapy, and 39 partial therapy. A further 21 participants had less than five sessions of therapy, thus falling into our no-therapy group. It should be emphasized that partial therapy met the definition of highly competent cognitive therapy and was observed to be accompanied by attempts by the therapists to deliver the techniques of full therapy as well. However, in partial therapy such attempts, by definition, fell below the predefined threshold for the identification of full therapy.

Table 1 provides information on the demographic characteristics of the trial participants. Table 2 illustrates the distribution of those in the CBT-P group receiving no therapy, partial therapy and full therapy, cross-classified by various baseline factors. Treatment centre (location) seems to be the best predictor of therapy received (note, in particular, that 19 of the 42 patients receiving full therapy were from Centre 3).

Table 3 provides information on the two main outcomes (time in remission and PANSS Total scores) by treatment arm, separately for the no-therapy, partial therapy and full therapy subgroups. There are no obvious patterns, and the full therapy subgroup did no better than the others. However, using the mean outcomes for these subgroups in this way cannot distinguish between effects arising from the treatment of interest and those deriving from treatment-independent prognosis (confounding or selection effects). Hence the need for more refined analysis. The requirement that is missing for the comparison of subgroup treatment effects is the average outcome in the respective principal strata in the control (TAU) condition.

We now summarize the analyses based on the use of principal stratification. In Table 4, we provide estimates of stratum-specific ITT effects for our four chosen outcomes, displayed separately for follow-up at 12 and 24 months. There were very few missing data for the number of months in remission, and we assumed that such missing data as existed were MAR. Data from research interviews were more likely to be missing, and we therefore used two separate methods for dealing with missing data. In the first, we assumed data were MAR. The second analysis assumes missing outcomes were LI.

Table 4 shows differences between treatment and control groups for each of the two principal strata corresponding to partial and full therapy respectively. Full treatment brings about nearly six additional months in remission between induction and the 12-month follow-up (indicated by an ITT effect with a positive sign) and an additional two months between the 12- and 24-month assessments. The

Table 1. Demographic characteristics of participants

	TAU (<i>n</i> = 140)	CBT (<i>n</i> = 133)
Gender		
Male	97	95
Female	43	38
Ethnicity		
White	105	88
Black Caribbean	8	14
Black African	14	13
Black other	4	3
Indian	0	5
Other	9	10
Employment		
Employed	8	5
Employed part-time	4	2
Voluntary employment	1	7
Unemployed	117	112
Student	8	3
Retired	2	1
Housewife/husband	0	2
Unknown	0	1
Marital status		
Single	107	91
Married	8	16
Divorced/separated	21	24
Widowed	1	2
Unknown	3	0
Diagnosis		
Schizophrenia	117	117
Schizo-affective disorder	23	12
Delusional disorder	0	4
In hospital?		
In-patient	96	92
Out-patient	44	41
History of violence		
No	102	86
Yes	35	42
History of suicide or self-harm		
No	80	81
Yes	55	46
Mean age, years (s.d.)	36.8 (10.9)	39.0 (10.7)
Mean duration of illness, years (s.d.)	10.1 (8.6, <i>n</i> = 135)	10.9 (8.4, <i>n</i> = 131)
Mean no. of admissions (s.d.)	4.4 (4.6, <i>n</i> = 130)	4.7 (5.2, <i>n</i> = 129)
PANSS Total mean (s.d.)	65.8 (15.8)	63.2 (13.7)
PANSS Positive mean (s.d.)	18.5 (5.3)	17.8 (5.5)
BDI-II mean (s.d.)	20.7 (13.5, <i>n</i> = 136)	22.2 (12.2, <i>n</i> = 130)

TAU, Treatment as usual; CBT, cognitive behaviour therapy; PANSS, Positive and Negative Syndrome Scale; BDI-II, Beck Depression Inventory Second Edition; s.d., standard deviation.

Table 2. Number of participants receiving each level of cognitive behaviour therapy (CBT)

By location	Centre 1	Centre 2	Centre 3	Centre 4	Centre 5	Total
Control group	26	19	35	35	25	140
CBT group						
No therapy	2	4	6	2	7	21
Partial therapy	12	8	6	11	2	39
Full therapy	9	1	19	8	5	42
Not known	3	2	3	13	10	31
By carer	Has carer	Does not have carer				
Control group	41	99				
CBT group						
No therapy	10	11				
Partial therapy	11	28				
Full therapy	13	29				
Not known	12	19				
By in-patient status	In-patient	Out-patient				
Control group	96	44				
CBT group						
No therapy	12	9				
Partial therapy	23	16				
Full therapy	32	10				
Not known	25	6				
By sex	Male	Female				
Control group	96	44				
CBT group						
No therapy	16	5				
Partial therapy	25	14				
Full therapy	32	10				
Not known	23	8				

12-month effect is statistically significant (at the 5% level). There is a suggestion that the ITT effect in the partial therapy group may be negative (detrimental) but the effects are not statistically significant.

The results for months in remission are mirrored in the findings for PANSS and BDI scores. At 12 months, the full therapy group had a statistically significant 16-point advantage on PANSS Total score over the control group (an ITT effect with a negative sign). Under the assumption that missing data were LI, the PANSS advantage fell to 12 points and was no longer significant. At 24 months, the advantage was still 11 points (12 under LI assumptions), albeit no longer statistically significant. Again, there is a suggestion that partial therapy might be detrimental rather than beneficial. Stratum-specific ITT effects for the PANSS Positive and BDI scores were consistent with the above findings, although none of the effects were statistically significant.

The results in Table 4 indicate that stratum-specific ITT effects for the 24-month outcomes were very

similar to those at 12 months. We therefore decided to refine our analyses by estimating stratum-specific ITT effects that were assumed to be common (i.e. the same) for the first and second 12-month periods of follow-up (see Appendix). The refined ANCOVA model for the outcomes was now bivariate. This allows for period-specific effects of the baseline covariates and correlations between the residuals of the outcomes at the two periods, and is an example of a Seemingly Unrelated Regression (SUR; Cox & Wermuth, 1996). The rationale was to improve both precision and statistical power, justified in the light of the consistency of effects across measures and periods.

Table 5 records the estimates of the stratum-specific ITT effects common to the two periods covered in the follow-up. Initially, each result is presented three times (three rows of ITT estimates). The first carries no exclusion restrictions (a relaxation of the assumptions in the models fitted above). In the second, the no-therapy group is set to zero (a single pair of exclusion restrictions, one for the 12-month outcome and another for 24 months, corresponding to our initial models). The last analysis includes similar constraints

Table 3. Outcomes by level of therapy (mean, s.d., n)

	Months in remission		PANSS Total		
	0–12 months	12–24 months	Baseline	12 months	24 months
Control group	7.60, 4.47, 137	8.77, 4.56, 132	65.83, 15.76, 140	59.39, 16.61, 113	57.04, 15.96, 109
CBT group					
No therapy	5.10, 5.25, 20	8.45, 4.57, 20	63.00, 16.16, 21	56.40, 14.68, 10	52.27, 15.37, 11
Partial therapy	8.39, 3.78, 38	9.68, 3.91, 38	66.15, 12.68, 39	60.13, 14.41, 38	58.94, 14.57, 36
Full therapy	7.57, 4.52, 42	8.60, 4.43, 42	63.50, 13.09, 42	55.97, 13.98, 37	56.27, 11.59, 37
Not known	7.58, 4.39, 31	8.57, 3.77, 30	59.42, 13.79, 31	53.96, 16.26, 26	53.96, 14.79, 25

CBT, Cognitive behaviour therapy; PANSS, Positive and Negative Syndrome Scale; s.d., standard deviation.

Table 4. ITT estimates within principal strata, separately for 12- and 24-month outcomes (bootstrapped standard errors in parentheses)

	No therapy (non-compliers)	Partial therapy	Full therapy
<i>(a) Negligible amounts of missing data (assumed MAR)</i>			
Months in remission			
0–12 months	0 ^a	–1.7 (1.1)	+5.6 (2.7) ^b
12–24 months	0 ^a	–2.1 (1.4)	+2.1 (2.1)
<i>(b) With missing outcomes assumed to be MAR</i>			
PANSS Total score			
12 months	0 ^a	+6.4 (3.7)	–16.4 (6.8) ^b
24 months	0 ^a	+7.5 (4.6)	–11.3 (5.9)
PANSS Positive score			
12 months	0 ^a	+1.8 (1.6)	–5.5 (2.8)
24 months	0 ^a	+3.4 (2.4)	–2.8 (3.3)
BDI score			
12 months	0 ^a	+2.2 (4.5)	–2.8 (4.5)
24 months	0 ^a	+2.7 (3.5)	–7.4 (4.6)
<i>(c) With missing outcomes assumed to be LI</i>			
PANSS Total score			
12 months	0 ^a	+7.7 (5.0)	–11.8 (8.2)
24 months	0 ^a	+6.5 (5.1)	–12.3 (7.0)
PANSS Positive score			
12 months	0 ^a	+1.7 (2.2)	–3.0 (2.8)
24 months	0 ^a	+3.7 (2.4)	–2.6 (3.2)
BDI score			
12 months	0 ^a	+5.7 (5.1)	–4.9 (4.3)
24 months	0 ^a	+2.5 (3.8)	–8.3 (6.3)

ITT, Intention to treat; MAR, missing at random; PANSS, Positive and Negative Syndrome Scale; BDI, Beck Depression Inventory; LI, latently ignorable.

^a Model constraint (exclusion restriction).

^b Statistically significant ($p < 0.05$): estimate two or more standard errors from zero.

imposed on both the no-therapy and partial therapy groups (two pairs of exclusion restrictions). The introduction of these additional exclusion restrictions prevents the effect of partial therapy from being detrimental and is therefore a stringent test of the effect in the full therapy stratum. This change (and

the relaxation of all restrictions as in the top row) provides a check on the sensitivity of the estimate of the effect of full therapy to a different set of model assumptions. As in Table 4, the results are calculated under different assumptions about the distribution of missing data.

Table 5. Estimated ITT effects within principal strata common to 12- and 24-month follow-up (bootstrapped standard errors in parentheses)

(a) Negligible amounts of missing data (assumed MAR)			
	No therapy (non-compliers)	Partial therapy	Full therapy
Months in remission	-2.8 (1.8)	-1.9 (0.5) ^b	+4.9 (1.5) ^b
	0 ^a	-1.9 (0.5) ^b	+4.7 (1.7) ^b
	0 ^a	0 ^a	+4.4 (2.3) ^b
	No and partial therapy (combined)		Full therapy
	0 ^a		+3.5 (2.6)
(Not known→No or partial therapy)	-2.3 (0.6) ^b		+4.4 (1.4) ^b
(Not known→Full therapy)	-2.3 (0.4)		+5.2 (1.2)
	-2.5 (1.2) ^b		+3.7 (1.6) ^b
(b) Missing data assumed to be MAR			
	No therapy (non-compliers)	Partial therapy	Full therapy
PANSS Total score	+9.0 (5.6)	+3.6 (3.1)	-13.0 (3.6) ^b
	0 ^a	+5.2 (3.2)	-12.7 (4.1) ^b
	0 ^a	0 ^a	-12.2 (4.8) ^b
PANSS Positive score	+3.2 (2.2)	+0.6 (1.4)	-3.9 (1.7) ^b
	0 ^a	+1.2 (1.6)	-4.0 (1.9) ^b
	0 ^a	0 ^a	-3.6 (2.0)
BDI score	-0.5 (4.6)	+5.3 (3.8)	-8.5 (3.5) ^b
	0 ^a	+6.3 (3.4)	-8.5 (3.1) ^b
	0 ^a	0 ^a	-7.0 (3.7)
	No and partial therapy (combined)		Full therapy
PANSS Total score	0 ^a		-10.8 (5.4) ^b
	+6.0 (2.1) ^b		-11.7 (3.5) ^b
(Not known→No or partial therapy)	+5.4 (1.9) ^b		-14.0 (3.2) ^b
(Not known→Full therapy)	+8.2 (4.3)		-9.4 (4.5) ^b
PANNS Positive score	0 ^a		-2.0 (2.8)
	+2.0 (1.3)		-3.4 (2.2)
(Not known→No or partial therapy)	+1.7 (0.9)		-4.2 (2.1) ^b
(Not known→Full therapy)	-3.1 (2.5)		+2.5 (2.3)
BDI score	0 ^a		-6.7 (3.3) ^b
	+3.6 (2.1)		-8.1 (3.1) ^b
(Not known→No or partial therapy)	+2.5 (1.8)		-9.2 (3.2) ^b
(Not known→Full therapy)	+4.0 (2.4)		-6.3 (2.7) ^b

ITT, Intention to treat; MAR, missing at random; PANSS, Positive and Negative Syndrome Scale; BDI, Beck Depression Inventory.

^a Exclusion restriction.

^b Statistically significant ($p < 0.05$): estimate two or more standard errors from zero.

We made two further sensitivity checks. The first involved merging the first two principal strata into one (the ITT estimates in rows four and five). The second involved allocating those with a missing treatment indicator in the CBT-P arm either to the worst option (no therapy) or to the best (full therapy). Overall, the refined analyses based on bivariate outcomes confirmed the findings in Table 4. Whatever the measure used, there was a considerable and significant advantage in the full therapy group, and the suspicion of detriment in those receiving only partial therapy.

Finally, we return to Table 2. There was a centre effect in the delivery of treatment: the delivery of full therapy was more frequently achieved in the rural county of Norfolk (Centre 3). If our conclusions concerning the effects of receiving full therapy are valid, then the direct implication is that the ITT effect of CBT in Centre 3 (rural Norfolk) would be very different to that in the other centres. We therefore carried out a more conventional analysis of centre effects (i.e. testing the Centre 3 by CBT interaction). This is available from the first author. It demonstrated that the treatment (ITT) effects in Centre 3 were

commensurably better in terms of PANSS Total and BDI scores, but not the PANSS Positive score or months in remission.

Discussion

This study used a novel approach to estimate the treatment effects of subgroups of the arm receiving CBT-P in a large randomized controlled treatment trial. The approach provides less biased estimates of the effect size of such subgroups by taking account of the potential outcomes had such cases been randomized to the control group. The study compared three categories of treatment. Our hypothesis was that the subgroup that engaged with and received full CBT-P would have better outcomes than those who received partial therapy or who dropped out. Our results are wholly consistent with this hypothesis. Treatment was effective if, and only if, clients received full therapy. Gains were large, and both clinically and statistically significant. They were also consistent, applying both to the number of months recovered and relapse free (the primary outcome) and to psychotic and depressive symptom outcomes. Participants who received therapy consisting only of engagement and assessment work did not benefit, and neither did those who dropped out. There is a suggestion that therapy had a somewhat deleterious effect on the former group.

This is a novel analysis based on a development of the methods of Complier-Average Causal Effect (CACE) estimation (Angrist *et al.* 1996; Frangakis & Rubin, 2002). CASE estimation has been applied previously to RCTs in psychiatry (Dunn *et al.* 2003; Horvitz-Lennon *et al.* 2005; Bellamy *et al.* 2007; Serfaty *et al.* 2009). The analysis is dependent on modelling that aims *a priori* to circumvent the biased estimates of treatment effects obtained by traditional per protocol approaches to analysis. In applying this technique, we have arrived at an estimate strikingly different from the overall ITT result we reported previously, which showed no effect of CBT-P beyond reducing depression at 24 months. Moreover, the modelled effect is not apparent from simple observation of the mean effects of subgroups within the treated arm alone (Table 3). Without a proper understanding of the assumptions underpinning estimations of treatment effects in randomized trials, this may seem counter-intuitive. We must, however, take account of the fact that simple descriptions of mean effects in subgroups of a single arm of a trial (the treatment arm) are in fact highly biased estimates of treatment effects. Such descriptive statistics do not take account of biases due to dropout, and to the putative effects if those randomized to treatment had instead been randomized to control. The modelling used here has been developed

specifically to overcome such biases, and is described elsewhere in specialist publications (Frangakis & Rubin, 1999, 2002; Dunn *et al.* 2005; Emsley *et al.* 2010). The approach has application to any situation where heterogeneity in treatment response is analysed in terms of subgroups defined by post-randomization explanatory variables. Examples include medication adherence, therapeutic alliance and intermediate biomarkers such as immune response.

The trial was designed *a priori* to study the effects of differing levels of therapy delivery. We used detailed observations of adherence and competence to identify those who received full CBT-P. Only 40% of participants did so. This raises the question of why so few received full therapy. CBT-P is complex, and its effective delivery depends on the interaction between therapist and patient, and hence on two types of factors: those relating to the patient (readiness and willingness to engage, the nature of symptoms, awareness, levels of distress) and those relating to the therapist (ability, training, supervision, adherence and competence). We ensured that the therapists in the trial were trained to the highest standards, and this was supported by our detailed monitoring of therapy sessions. Despite this, they were able to deliver full therapy only to a minority. This might therefore be the result of patient attributes in this sample, although we must emphasize that there were no baseline differences in symptoms. Although people with psychosis have well-known problems with engagement in therapeutic relationships, the techniques of CBT-P have been specifically designed to minimize them. Nevertheless, in a substantial minority in the present study, therapists were not able to move much beyond maintaining engagement and working collaboratively with clients to make sense of their problems. It would have been interesting to relate the characteristics of the CBT-P received with the strength of the therapeutic alliance and to look at their joint relationship with the effects of therapy. However, the statistical methods required to undertake this work are in their infancy (Dunn & Bentall, 2007; Emsley *et al.* 2010).

The superior delivery of therapy and better treatment effects in Norfolk are noteworthy. The difficulties of delivering complex interventions in inner city areas are well known to clinicians, and might be attributed variously to low levels of social support, high levels of deprivation, and relative residential instability. Such contextual disadvantages remain a therapeutic challenge.

Consistent attempts were made to deliver more active cognitive and behavioural techniques to all clients, but with many it was impossible to achieve the level necessary for the *a priori* definition of full

therapy. There are many possible reasons for this, some of which may be especially characteristic of unselected, recently relapsed groups. Despite initial willingness, after a few weeks some patients no longer wanted to receive therapy. Some had symptoms, but were not distressed by them, some had responded to the reinstatement of medication (this was not a medication-resistant sample, indeed many relapses seemed to follow discontinuation of medication) and no longer saw the point of a psychological treatment. Some simply lacked interest in working with a therapist, and others had limited awareness of their problems. Despite such difficulties, our therapists managed to keep these clients engaged in therapy. We had clear observational evidence of therapists establishing a basic working cognitive behavioural relationship, systematically carrying out assessment, and promoting collaborative guided discovery in a highly skilled manner. However, it must be emphasized that, in this trial, persistence was sometimes associated with a worsening of symptoms. This is an important observation, with implications for clinical practice. We conclude that if therapists have not managed to move into the active phase of therapy within a circumscribed period, it may not be worth persisting, although clinical experience backs the option of a later return to therapy. Only clients with whom therapists can deliver a substantial amount of active therapy seem to benefit: future work should aim to identify them.

In summary, this analysis shows clearly that CBT-P has widespread and beneficial effects when delivered as intended in a group of relapse-prone patients. These effects apply to our original primary and secondary outcomes, of relapse prevention and symptomatic improvement. CBT-P is therefore clearly a useful and effective intervention. However, our results also indicate that those clients whom therapists cannot engage in substantial active therapy may not benefit; at best it is not cost-effective to continue therapy under such circumstances.

Acknowledgements

The study was supported by a Wellcome Trust Programme Grant (062452). The developments in statistical methodology (G.D.) were supported by Medical Research Council (MRC) Methodology Research Programme Grants G0600555 and G0900678. [Trial Registration: isrctn.org identifier ISRCTN 83557988.]

Declaration of Interest

None.

References

- Angrist JD, Imbens GW, Rubin DB** (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444–455.
- APA** (1994). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association: Washington, DC.
- Bebbington PE, Craig T, Garety P, Fowler D, Dunn G, Colbert S, Fornells-Ambrojo M, Kuipers E** (2006). Remission and relapse in psychosis: operational definitions based on case-note data. *Psychological Medicine* **36**, 1551–1562.
- Beck AT, Steer RA, Brown GK** (1996). *BDI-II Manual*. The Psychological Corporation: San Antonio, TX.
- Bellamy SL, Lin JY, Have TRT** (2007). An introduction to causal modelling in clinical trials. *Clinical Trials* **4**, 58–73.
- Chadwick P, Williams C, Mackenzie J** (2003). Impact of case formulation in cognitive behaviour therapy for psychosis. *Behaviour Research and Therapy* **41**, 67–80.
- Cox DR, Wermuth N** (1996). *Multivariate Dependencies*. Chapman & Hall: London.
- Craig TJC, Garety P, Power P, Rahaman N, Colbert S, Fornells-Ambrojo M, Dunn G** (2004). The Lambeth Early Onset (LEO) Team: a randomised controlled trial of assertive outreach for early psychosis. *British Medical Journal* **329**, 1067–1070.
- Dunn G, Bentall R** (2007). Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine* **26**, 4719–4745.
- Dunn G, Maracy M, Dowrick C, Ayuso-Mateos JL, Dalgard OS, Page H, Lehtinen V, Casey P, Wilkinson C, Vázquez-Barquero JL, Wilkinson G; The Outcomes of Depression International (ODIN) Group** (2003). Estimating psychological treatment effects from an RCT with both non-compliance and loss to follow-up. *British Journal of Psychiatry* **183**, 323–331.
- 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**, 369–395.
- Durham RC, Guthrie M, Morton V, Reid DA, Treliving LR, Fowler D, Macdonald RR** (2003). Tayside-Fife clinical trial of cognitive behavioural therapy for medication-resistant psychotic symptoms. *British Journal of Psychiatry* **182**, 303–311.
- Efron B, Tibshirani RJ** (1993). *An Introduction to the Bootstrap*. Chapman & Hall: London.
- Emsley R, Dunn G, White IR** (2010). Modelling mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research* **19**, 237–270.
- Fowler D, Garety PA, Kuipers L** (1995). *Cognitive Behaviour Therapy for Psychosis*. Wiley: Chichester.
- Fowler D, Morley S** (1989). The cognitive behavioural treatment of hallucinations and delusions: a preliminary study. *Behavioural Psychotherapy* **17**, 267–282.

- Frangakis CE, Rubin DB** (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.
- Frangakis CE, Rubin DB** (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.
- Garety PA, Fowler D, Freeman D, Bebbington P, Dunn G, Kuipers E** (2008). A randomised controlled trial of cognitive behavioural therapy and family intervention for the prevention of relapse and reduction of symptoms in psychosis. *British Journal of Psychiatry* **192**, 412–423.
- Garety PA, Kuipers E, Fowler D, Freeman D, Bebbington PE** (2001). Theoretical paper: a cognitive model of the positive symptoms of psychosis. *Psychological Medicine* **31**, 189–195.
- Horvitz-Lennon M, O'Malley AJ, Frank RG, Normand SLT** (2005). Improving traditional intention-to-treat analysis: a new approach. *Psychological Medicine* **35**, 961–970.
- Jones C, Cormac I, Silveira Da Mota Neto JI, Campbell C** (2004). Cognitive behaviour therapy for schizophrenia. *Cochrane Database of Systematic Reviews* Issue 4, Art. No. CD000524.
- Kay RS** (1991). *Positive and Negative Syndromes in Schizophrenia: Assessment and Research*. Brunner/Mazel, Inc.: New York.
- Leucht S, Arbter D, Engel RR, Kissling W, Davis JM** (2009). How effective are second generation anti-psychotic drugs? A meta-analysis of placebo controlled trials. *Molecular Psychiatry* **14**, 429–447.
- Little RJA, Rubin DB** (2002). *Statistical Analysis with Missing Data*, 2nd edn. John Wiley & Sons: Hoboken, NJ.
- Muthén LK, Muthén BO** (1998–2009). *Mplus User's Guide*. Muthén & Muthén: Los Angeles, CA.
- NICE** (2009). *Schizophrenia: Core Interventions in the Treatment and Management of Schizophrenia in Primary and Secondary Care (Update)*. National Institute of Clinical and Health Excellence: London.
- Pilling S, Bebbington P, Kuipers E, Garety P, Geddes J, Orbach G, Morgan C** (2002). Psychological treatments in schizophrenia. I: Meta-analysis of family intervention and cognitive behaviour therapy. *Psychological Medicine* **32**, 763–782.
- Rollinson R, Haig C, Warner R, Garety P, Kuipers E, Freeman D, Bebbington P, Dunn G, Fowler D** (2007). The application of cognitive-behavioral therapy for psychosis in clinical and research settings. *Psychiatric Services* **58**, 1297–1302.
- Rollinson R, Smith B, Steel C, Jolley S, Onwumere J, Garety PA, Kuipers E, Freeman D, Bebbington PE, Dunn G, Startup M, Fowler D** (2008). Measuring adherence in CBT for psychosis: a psychometric analysis of an adherence scale. *Behavioural and Cognitive Psychotherapy* **36**, 163–178.
- Sensky T, Turkington D, Kingdon D, Scott JL, Scott J, Siddle R, O'Carroll M, Barnes TR** (2000). A randomized controlled trial of cognitive-behavioral therapy for persistent symptoms in schizophrenia resistant to medication. *Archives of General Psychiatry* **57**, 165–172.
- Serfaty MA, Hawaorth D, Blanchard M, Buszewicz M, Murad S, King M** (2009). Clinical effectiveness of individual cognitive behavioural therapy for depressed older people in primary care. *Archives of General Psychiatry* **66**, 1332–1340.
- Startup M, Jackson M, Pearce E** (2002). Assessing therapist adherence to cognitive-behaviour therapy for psychosis. *Behavioural and Cognitive Psychotherapy* **30**, 329–339.
- WHO** (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Description and Diagnostic Guidelines*. World Health Organization: Geneva.
- Wykes T, Steel C, Everitt B, Tarrrier N** (2008). Cognitive behaviour therapy for schizophrenia: effect sizes, clinical models, and methodological rigor. *Schizophrenia Bulletin* **34**, 523–537.
- Young JE, Beck AT** (1980). *Cognitive Therapy Scale: Rating Manual*. Center for Cognitive Therapy: Philadelphia, PA.
- Zimmerman G, Favrod J, Trieu VH, Pomini V** (2005). The effect of cognitive behavioral treatment on the positive symptoms of schizophrenia spectrum disorders: a meta-analysis. *Schizophrenia Research* **77**, 1–9.

Appendix: Example of Mplus input file (modelling 12- and 24-month PANSS scores)

```

TITLE: PRP Therapy – Principal Stratification
DATA:
  FILE IS 'TherapyPredictors.dat';
VARIABLE:
  NAMES ARE r1 r2 r3 r4 carer sex output treat          ! r1 to r4 are binary dummies indicating treatment centre
  c1 c2 c3 pan0 pan12 resp12 pan24 resp24;            ! resp12 and resp24 are missing value indicators
  CLASSES c(3);
  TRAINING c1 c2 c3;                                  ! binary dummies indicating compliance status
  CATEGORICAL resp12 resp24;
  USEVARIABLES treat sex output r1 r2 r3 r4 r5
  pan0 pan12 resp12 pan24 resp24
  c1 c2 c3 carer resp12 resp24;
  MISSING
  pan12 (999) pan24 (999)
ANALYSIS:
  TYPE=MIXTURE;
  ESTIMATOR=ML;
  STARTS=1000 20;
  BOOTSTRAP=250;
MODEL:
  %OVERALL%
  PAN12 ON treat carer r1 r2 r3 r4 pan0 sex output;
  PAN24 ON treat carer r1 r2 r3 r4 pan0 sex output;
  pan12 WITH pan24;
  resp12 ON treat carer r1 r2 r3 r4 pan0 sex output;
  resp24 ON treat carer r1 r2 r3 r4 pan0 sex output;
  C#1 ON carer r1 r2 r3 r4 PAN0 sex output;
  C#2 ON carer r1 r2 r3 r4 PAN0 sex output;
  %C#1% ! No therapy
  [pan12 pan24 resp12$1 resp24$1];
  pan12 pan24 (3);
  pan12 ON treat@0;
  pan24 ON treat@0;
  resp12 ON treat@0;
  resp24 ON treat@0;
  %C#2% ! Partial therapy
  [pan12 pan24 resp12$1 resp24$1];
  pan12 pan24 (4);
  pan12 ON treat*0 (1);
  pan24 ON treat*0 (1);
  resp12 ON treat*0 (6);
  resp24 ON treat*0 (6);
  %C#3% ! Full therapy
  [pan12 pan24 resp12$1 resp24$1];
  pan12 pan24 (5);
  pan12 ON treat*0 (2);
  pan24 ON treat*0 (2);
  resp12 ON treat*0 (7);
  resp24 ON treat*0 (7);

```
