



Influence of therapist competence and quantity of cognitive behavioural therapy on suicidal behaviour and inpatient hospitalisation in a randomised controlled trial in borderline personality disorder: Further analyses of treatment effects in the BOSCOT study

John Norrie¹, Kate Davidson^{2*}, Philip Tata³ and Andrew Gumley²

¹Centre for Healthcare Randomised Trials (CHaRT), Health Services Research Unit, Aberdeen University, UK

²Institute of Health & Wellbeing, Gartnavel Royal Hospital, Glasgow, UK

³Adult Psychology Services, Central and North West London NHS Foundation Trust, London, UK

Objectives. We investigated the treatment effects reported from a high-quality randomized controlled trial of cognitive behavioural therapy (CBT) for 106 people with borderline personality disorder attending community-based clinics in the UK National Health Service – the BOSCOT trial. Specifically, we examined whether the amount of therapy and therapist competence had an impact on our primary outcome, the number of suicidal acts[†], using instrumental variables regression modelling.

Design. Randomized controlled trial. Participants from across three sites (London, Glasgow, and Ayrshire/Arran) were randomized equally to CBT for personality disorders (CBTpd) plus Treatment as Usual or to Treatment as Usual. Treatment as Usual varied between sites and individuals, but was consistent with routine treatment in the UK National Health Service at the time. CBTpd comprised an average 16 sessions (range 0–35) over 12 months.

*Correspondence should be addressed to Kate Davidson, Institute of Health & Wellbeing, Admin Building, Gartnavel Royal Hospital, 1055 Great Western Road, Glasgow G12 0XH, UK (e-mail: Kate.Davidson@glasgow.ac.uk).

[†]Suicidal act: A suicidal act meets all three of the following criteria: (1) Deliberate (i.e., not be construed as an accident, planning involved, and the individual accepts ownership of the act); (2) life threatening, in that the individual's life was deemed to be seriously at risk, or he or she thought it to be at risk, as a consequence of the act; and (3) the act resulted in medical intervention or intervention would have been warranted. The individual may have sought or would have warranted medical intervention or medical intervention was sought on their behalf. Medical intervention need not be treatment, but at the minimum a physical examination is implied.

NIH Clinical Trials Gov Identifier: NCT00538135.

Re-use of this article is permitted in accordance with the Terms and Conditions set out at <http://olabout.wiley.com/WileyCDA/Section/id-817008.html>

Method. We used instrumental variable regression modelling to estimate the impact of quantity and quality of therapy received (recording activities and behaviours that took place after randomization) on number of suicidal acts and inpatient psychiatric hospitalization.

Results. A total of 101 participants provided full outcome data at 2 years post randomization. The previously reported intention-to-treat (ITT) results showed on average a reduction of 0.91 (95% confidence interval 0.15–1.67) suicidal acts over 2 years for those randomized to CBT. By incorporating the influence of quantity of therapy and therapist competence, we show that this estimate of the effect of CBTpd could be approximately two to three times greater for those receiving the right amount of therapy from a competent therapist.

Conclusions. Trials should routinely control for and collect data on both quantity of therapy and therapist competence, which can be used, via instrumental variable regression modelling, to estimate treatment effects for optimal delivery of therapy. Such estimates complement rather than replace the ITT results, which are properly the principal analysis results from such trials.

Practitioner points

- Assessing the impact of the quantity and quality of therapy (competence of therapists) is complex.
- More competent therapists, trained in CBTpd, may significantly reduce the number of suicidal acts in patients with borderline personality disorder.

We investigated the estimated treatment effects from the BOScot randomized controlled trial of cognitive behavioural therapy for personality disorders (CBTpd) in addition to treatment as usual (CBTpd plus TAU) compared with TAU alone in 106 people with borderline personality disorder (Davidson, Norrie *et al.*, 2006; Davidson, Tyrer *et al.*, 2006). Those results used intention-to-treat (ITT) principle, recognized as the appropriate methodology for the principal reporting of randomized controlled trials. The additional analyses presented here go beyond these results and utilize information on quantity and quality of therapy received (recording activities and behaviours that took place after randomization), relating this information to two of our primary outcomes, suicidal acts and inpatient hospitalization, using instrumental variable regression modelling. In addition, we investigate any possible delay in the treatment effect manifesting itself, sometimes a feature of complex interventions such as CBT (Medical Research Council, 2000, 2008). These additional analyses are important to patients and clinicians as they provide an estimate of the benefit CBTpd would give under ideal conditions. They also inform comparisons of different psychological therapies when resources are scarce and may help identify subgroups of patients who may benefit most from therapy.

Method

The BOScot study design (Davidson, Tyrer *et al.*, 2006), results (Davidson, Norrie *et al.*, 2006), and economic implications (Palmer *et al.*, 2006) have been reported elsewhere. In brief, patients in both treatment arms showed gradual and sustained improvement in both primary and secondary outcomes, with evidence to show that addition of CBTpd

benefited the positive symptom distress index at 1 year, and state anxiety, dysfunctional beliefs, and the quantity of suicidal acts over the 2-year study period. We subsequently reported on the 6-year follow-up of this randomized cohort (Davidson, Tyrer, Norrie, Palmer, & Tyrer, 2010) – the analyses presented here do not include this longer term follow-up as we do not have correspondingly high quality (in terms of completeness and accuracy) of therapy received in the years 3–6 after the completion of the randomized trial follow-up.

Nearly all 106 participants randomized had primary and secondary outcome data at 6, 12, 18, and 24 months. These additional analyses on how quantity and quality of CBT might influence behaviour and inpatient hospitalization are therefore not complicated by missing outcome data. In addition, BOSCOT had no ‘treatment crossovers’, as the form of CBT for personality disorders specified in the therapy protocol differs considerably from conventional CBT for Axis I disorders (Davidson, 2007), and was not available in the NHS during the study period.

Therapist competence

Any psychological therapy will have some variability in terms of quality of delivery. CBT therapists vary in their degree of skill, and some will become better (or worse) as they become more experienced. Quality of therapy in CBT trials comprises at least two dimensions: the therapeutic alliance and the competence with which therapy is delivered.

Therapeutic alliance is the specific working relationship that develops between a patient and therapist. Competence of the therapist in delivering CBTpd is the focus here. A therapist’s competence involves his/her understanding of the patient’s problems, the appropriateness of an intervention or use of techniques, and the therapist’s aptitude at delivering interventions in a skilled manner.

We measured therapist competence using the BOSCOT Rating Scale (Davidson, 2007) and the Cognitive Therapy Rating Scale (CTRS; Young & Beck, 1990). All five BOSCOT therapists submitted CBTpd session audiotapes. A random selection of audiotapes from 24 of the 38 patients (73% of the 54 patients randomized to CBTpd), who gave written consent to their sessions being recorded, were rated by two experienced clinical psychologists (KD and AG), both blind to treatment outcome, and established good inter-rater agreement on therapist competence (see Davidson, Norrie *et al.*, 2006). Data on how often and when CBTpd sessions were offered, declined, and attended, were collected, allowing characterization of therapy delivery in terms of therapist competence and frequency and intensity of therapy sessions.

Is there a time lag for treatment effect on suicidal behaviour? Any intervention may take time for an effect to manifest itself. After randomization to CBTpd, it might take several weeks for the first appointment to be scheduled, due to the practical constraints of delivering the service. It seems reasonable that a ‘treatment effect’ would not be seen before treatment has been received. In addition, it may also take several sessions to engage patients and develop a collaborative therapeutic relationship, which will permit the implementation of specific cognitive behavioural techniques. Even after therapy commences, it may take several sessions for any therapeutic effect to accumulate as techniques become practised and implemented. We therefore discounted the earliest suicidal acts as being unlikely to have been influenced by therapy, which either was yet to start, or had only recently begun.

Statistical methods

There are three parts to the analyses. (1) time lag of a treatment effect; (2) the inter-relationships between quality (therapist competence) and quantity (number of sessions attended) of therapy, and outcomes; and (3) instrumental variable regression to investigate the influence of quantity and quality of therapy on outcome, for which Stata 10.0SE was used. All other analyses used SAS 9.2 for Windows. No adjustment was made for multiple comparisons as we judged that the risk of making a Type I error was offset by the importance of these post-randomization analyses to the development of improved understanding of the interactions between characteristics of therapy delivered in the context of clinical trials and the primary outcomes of these trials.

- (1) The time lag of treatment effect analyses used standard ITT statistical techniques, with the 'time zero' moved forward incrementally by 30, 60, 90, and 180 days, so deleting all suicidal events that happened before these milestones. For the corresponding analysis on inpatient hospitalizations (IPH), we restricted these to the first 6 months as this was most likely to be accurately reported and case notes were checked.
- (2) For analyses of associations between intervention characteristics (sessions offered, attended, cancelled, and 'did not attend', and delay from randomization to first session, duration of sessions [elapsed time from first to last session attended], and intensity of sessions [sessions attended per 3-month period]), we grouped variates above or below their medians and then performed two sample *t*-tests on the other variates of interest. For linear models of predictors of intensity, we used stepwise regression with *p*-to-enter and *p*-to-stay both .10, and indicator variables to adjust for therapist forced into the model.
- (3) For the analyses adjusting outcome for quantity and quality of therapy, we used Complier Averaged Causal Effects models, as described by Dunn and Bentall (2007), Dunn *et al.* (2003) and Dunn, Maracy, and Tomenson (2005), as implemented in the two-stage least-squares routine 'ivregress' in Stata 10.0SE. We present three estimates: (1) unadjusted; (2) adjusted for four baseline factors strongly associated with outcome, which was the number of suicidal acts – these baseline factors were number of suicidal acts in the 12 months before randomization, being single, age at first deliberate self-harm, and EQ-5D score at baseline; and (3) the interaction of treatment with each of these four baseline predictors to check whether the exclusion restriction (see below) was likely to hold in this data set.

Results

Outcome of time lag for treatment effects on suicidal behaviour analyses

From the main results paper, visual examination suggested that the time to first suicidal act curves are initially coincident and only separated and diverged after about 6 months (figure 1A, Davidson, Norrie *et al.*, 2006). To investigate whether this could indicate a delayed onset of a treatment effect, Figure 1 is the corresponding Kaplan–Meier curve after excluding events before 182 days after randomization. There is no support for a delayed treatment effect. To understand this, we observe that approximately half the participants have no event over 24 months, about a quarter have one event, and the other quarter more than one suicidal act. So, the main effect

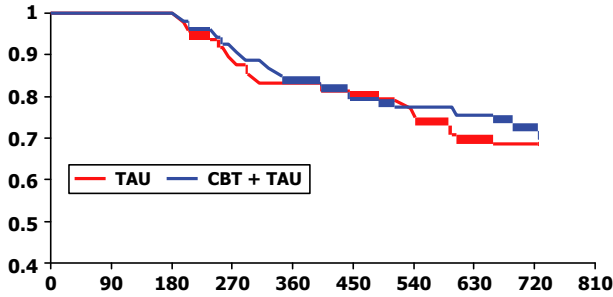


Figure 1. Time to first suicidal act, omitting the first 180 days post randomization.

of excluding the first 6 months of follow-up data is to simply delay the time at which a multiple-suicide-act participant will have a first suicidal act. In the original analysis; 26 TAU against 23 CBTpd had at least one suicidal act, log-rank $p = .29$. Omitting any events in the first 30 days gives 26 versus 22, log-rank $p = .33$; omitting the first 60 days gives 24 versus 21, log-rank $p = .46$; omitting the first 90 days gives 23 versus 20, $p = .44$; and omitting the first 6 months gives 17 versus 16, log-rank $p = .97$. For suicidal acts, the ITT analysis gave a 24-month mean of 1.73 (SD 3.11) on TAU and 0.87 (1.47) on CBTpd + TAU – an adjusted difference of -0.91 (95% confidence interval -1.67 to -0.15 , $p = .020$). Omitting the first 180 days, the corresponding estimates are 1.31 (2.73) and 0.42 (0.87), with an adjusted difference of -0.86 (95% CI -1.51 to -0.20 , $p = .010$). That is, there is very little difference between the two analyses.

A similar analysis for inpatient psychiatric hospitalizations (IPH) excluded events within the first 182 days after randomization. The numbers of participants with at least one IPH falls from the original 23 (47%) for TAU to 18 (34%) compared with CBTpd rates of 12 (24%) and 14 (27%), respectively, over the remaining 18 months. Interestingly, then, the event rate is much higher in the first 6 months (Figure 2). Therefore, for IPH, the reverse may be true – a treatment effect that manifests early and then disappears, which may be a quite common phenomenon seen across a variety of therapeutic interventions.

A *post-hoc* Cox regression analysis for IPH (adjusting for the pre-specified covariates used in the original ITT analyses) shows a significant benefit in favour of CBTpd + TAU over TAU: 18 (37%) on TAU had at least one IPH compared with 12 (24%) on CBTpd + TAU, adjusted hazard ratio 0.41 (95% CI 0.18–0.93, $p = .032$). However, care should be taken not to overinterpret this finding as the numbers are small and this is a data-driven *post-hoc* comparison.

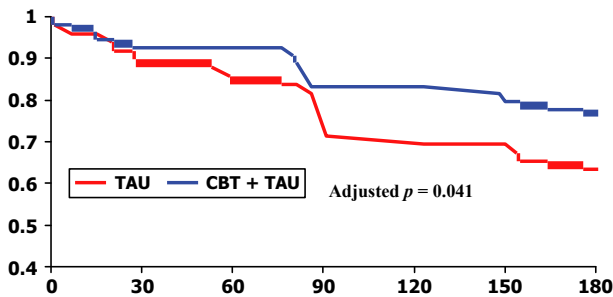


Figure 2. Time to first Inpatient hospitalization – in the first 180 days.

Inter-relationships between quantity and quality of therapy and outcomes

It is important to understand the inter-relationships between the quality and quantity of therapy, and outcomes, and what baseline data might help predict these, to aid interpretation of the instrumental variables regression models that adjust the treatment effects for these characteristics of the therapy.

We have previously reported uptake of CBTpd sessions and therapist competence according to the CTRS and a specific rating scale developed by the author for CBT for personality disorder – the BOSCOT Rating scale. Figure 3 gives the boxplots of descriptive statistics for the CTRS by therapist (very similar results were seen for the BOSCOT Rating Scale). These boxplots indicate that there is substantial variability in the ratings both within and between therapists.

Table 1 gives further information on CBTpd quantity, reporting session characteristics by therapist. There is considerable variability across therapists in all the measures of CBTpd sessions – offered, attended, cancelled, and ‘did not attend’. Not surprisingly, sessions attended is highly significantly associated with number of sessions offered, with an additional eight attended for every 10 offered ($p < .0001$). Sessions cancelled is significantly associated with sessions offered, with approximately every 10 additional sessions offered likely to result in one additional cancellation ($p = .039$).

The number of ‘did not attends’ was not associated with sessions offered ($p = .41$), which may indicate that once patients have low engagement in therapy, this persists despite therapists attempts to re-engage them. In addition, delay in

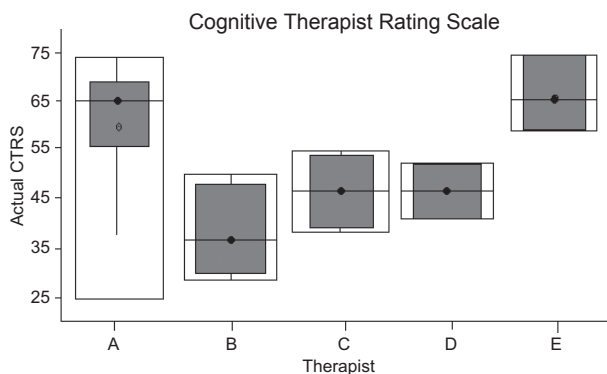


Figure 3. Boxplots of Cognitive Therapist Rating Scale by Therapist. Data shown are minimum and maximum (dotted box), interquartile range (yellow box), median (solid dot), and mean (open dot).

Table 1. Details of cognitive behavioural therapy sessions (mean [standard deviation]) offered, attended, cancelled, and did not attend

Therapist (number of patients)	Offered	Attended	Cancelled	Did not attend
A (20)	24 (11)	14 (12)	3 (3)	8 (7)
B (5)	35 (13)	14 (13)	4 (5)	17 (12)
C (13)	31 (11)	21 (12)	4 (4)	7 (5)
D (3)	47 (3)	20 (1)	2 (3)	24 (2)
E (11)	22 (12)	18 (13)	2 (2)	2 (3)
Overall (52)	27 (13)	17 (12)	3 (3)	8 (8)

initiating treatment (time elapsed from randomization to first session attended, excluding those who attended no sessions) in 40 of 54 subjects in the CBTpd group with data averaged 43 days (SD 33; median 33; range 19–123 days). Duration of treatment (time between first and last CBTpd sessions) was 350 days (SD 173; median 386, range 1–574 days). Intensity of therapy (sessions attended/duration of therapy, per 3 months, for 35 participants with data) averaged 4.81 (SD 3.22; median 5.05; range 0.68–20.1 CBTpd sessions/3 months).

Relationship between quality and quantity of therapy and session characteristics

More 'did not attends' (>5) were associated with a significantly lower scores on the CTRS – 14 points lower than the group with 5 or less DNAs (95% CI 3–25, $p = .014$). Otherwise, there was little association between any other measures – sessions attended and sessions cancelled – albeit with non-significant trends in the expected direction – nor delay to initiation of therapy, duration, or intensity of therapy. Sessions cancelled (three sessions/quarter more, approximately 95% 1–5, $p = .0091$) and DNA (seven extra sessions, approximately 95% CI 4–11, $p = .0004$) are both significantly lower among subjects with a higher intensity of therapy (>5 sessions/quarter). Likewise, a higher intensity of therapy is associated with a shorter delay in initiating therapy (26 days, 95% CI 0–53, $p = .047$).

Relationship of outcome with quantity and quality of therapy

Figure 4 shows the mean number of suicidal acts (in the previous 6-month period) by treatment group across 12 months of treatment and 12 months of follow-up. Figure 5 relates suicidal acts over the 12-month period of therapy to the quantity of CBTpd and likewise Figure 6 shows quality of CBTpd (ranking therapists by their average CTRS score). From Figure 5, although the baseline rate is higher than the treated rate across the board, there is no indication of an obvious, simple relationship between treatment received and treatment effect. Indeed, the lowest rate of suicidal acts post randomization (0.1/year) is among those with fewest CBTpd sessions, and the highest among those who received the most. However, in our design, participants were not randomized to therapists and it might be the case that some therapists had easier, or more difficult patients, or that those who needed the least CBTpd actually took the least number of sessions of CBTpd, while those who needed most received the largest amount of therapy. We have therefore added the expected rate from the stepwise baseline predictive model (Table 2).

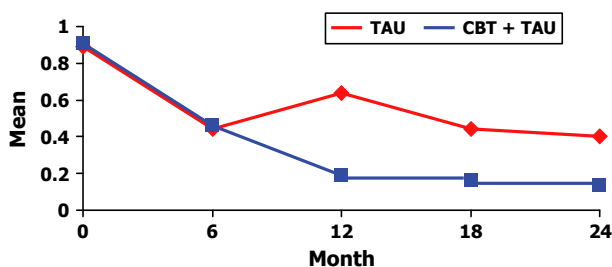


Figure 4. Mean number of suicidal acts (in the previous 6-month period) by randomized treatment group.

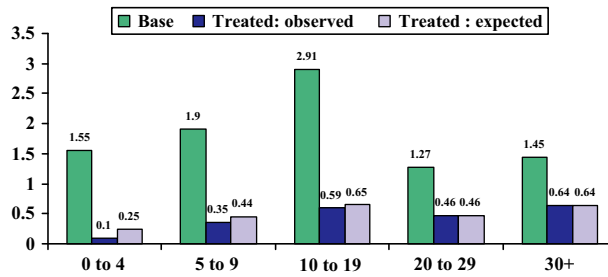


Figure 5. Suicidal acts per year by number of sessions taken.

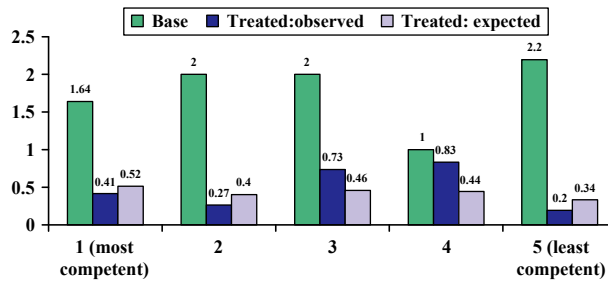


Figure 6. Suicidal acts per year by therapist competence.

Table 2. Baseline predictors of quantity (number of sessions) and quality (Cognitive Therapy Rating Scale), using data only from those randomized to cognitive behavioural therapy for personality disorders plus treatment as usual (TAU)

Baseline predictor	Quality – Cognitive Rating Scale		Quantity – number of sessions	
	Estimate (SE)	p-Value	Estimate (SE)	p-Value
Age at randomization (5 years)			2.79 (0.78)	.0008
Age at first Deliberate Self-Harm (5 years)	7.00 (1.73)	.0007		
Female	14.9 (5.04)	.0087		
EQ-5D (0.1 units)	-2.91 (0.64)	.15		
Young’s Schema Quest (total)	-7.81 (3.02)	.018	-6.83 (2.13)	.0023

Note. Stepwise model only. Data shown are the estimated change in number of suicidal acts (standard error of estimate) with associated p-value.

Although much of the difference between low and high session attendance can be explained by client type (0.25 suicide acts expected for 0–4 sessions attended through to 0.64 for 30+ sessions attended), there is still an indication that those receiving less therapy did proportionately better than would have been expected. For therapist competence, a similar picture emerges: apparently, the least competent therapist gets better results, and that coming from a position of highest average baseline suicidal acts. Considering the

expected number of suicidal acts, the picture is less clear – it is the ‘best’ and the ‘worst’ therapists who exceed expectations, while the middle ranked therapists in terms of competence have higher numbers of ‘observed’ as opposed to ‘expected’ suicidal acts.

Baseline predictors of quality and quantity of therapy

Using similar methodology as we did for the baseline predictors of outcome model, Table 2 shows that lower age at first deliberate self-harm episode and being female predict higher CTRS score, while higher quality of life and higher score on the Young Schema Questionnaire predict lower CTRS score. For number of sessions attended, older age predicts more sessions, while higher scores on the Young Schema Questionnaire predicts fewer sessions. Table 3 gives the rank correlations between all the baseline predictors and quality and quantity of therapy, and also outcome.

Instrumental variable regression modelling of treatment effects

To summarize, we have explored the data on quality and quantity of therapy, and informally associated these data with the outcome of number of suicidal acts. We have seen that linking measures of quality and quantity of therapy to outcomes is difficult as these are not baseline measures equalized by randomization and temporally measured before any outcomes of interest. Fortunately, there has been much development in the statistical methodology in the last decade to robustly incorporate such post-randomization data assessing the ‘success’ in delivering intervention and outcomes. Broadly, the idea is to identify participants who, in some sense, had the best chance of responding, and then the magnitude of their response indicates what treatment benefit may be

Table 3. Spearman rank correlations of baseline covariates with (1) quantity (number of sessions); (2) quality (Cognitive Therapy Rating Scale); and (3) outcome (number of suicidal acts over 2 years post randomization)

Baseline predictor	Quantity (sessions) N = 54	Quality (CT rating) N = 26	Outcome (suicidal acts) N = 54
# Suicidal acts in last 12 months	-.14	-.17	.51 ^a
Age at randomization (5 years)	-.06	-.07	.42 ^c
Age at first Delib Self-Harm (5 years)	.39 ^c	-.04	-.05
High Self-harm	.17	.45 ^d	.14
Female	-.17	.08	-.17
Single	-.23	.17	-.21
Left School < 16	.13	-.01	.28 ^d
Special Needs	-.11	.14	.08
Lives Alone	.20	-.13	-.17
Crime in last 12 months	.16	-.14	.07
Unemployed	-.17	-.12	.02
EQ-5D (0.1 units)	-.18	-.22	-.03
Young's Schema Quest (total)	-.35 ^d	-.41 ^d	.19

Note. The Spearman rank correlation between quality and quantity of therapy is .30, between quality and outcome .18, and between quantity and outcome .16.

^a $p < .0001$; ^b $p < .001$; ^c $p < .01$; ^d $p < .05$.

achieved if we were to optimize the delivery of the intervention to the right recipients. The Complier-Average Causal Effect, or CACE estimate, introduced by Angrist, Imbens, and Rubin (1996) and discussed by Bellamy, Lin, and Ten Have (2007), Dunn and Bentall (2007) and Dunn *et al.* (2003, 2005), is an attractive methodology for this task, identifying a group of 'compliant' participants before randomization, who are then equally distributed by randomization across the groups. The model then compares the observed behaviour of the compliers in the treatment group with what would have been observed if this group had been randomized to the other group. All of these so-called instrumental variable regression models rest heavily on assumptions that cannot be verified objectively.

For CACE models, an important assumption is the 'exclusion restriction' – the assumption that the randomization itself does not influence the outcome: for example, a participant who wanted to get CBTpd and was randomized to TAU does not suffer 'resentful demoralisation', which then worsens their outcomes (Torgerson & Sibbald, 1998a, 1998b). We tested whether the exclusion restriction held by including the interaction of treatment with each of the important baseline factors strongly associated with outcome (i.e., baseline number of suicidal acts in the previous 12 months, being single, age at first deliberate self-harm, and the EQ-5D score prior to randomization).

Table 4 gives the CACE model results for (1) quality of CBTpd therapy (CTRS score ≥ 60 vs. <60); and (2) quantity of therapy (sessions attended ≥ 15 vs. <15). In both cases, the treatment effect approximately doubles, with the ITT of -0.91 saved suicidal acts becoming -1.93 for the high-quality CBTpd and -2.17 for the higher number of sessions. Note that the confidence intervals around the estimates are wider, and note further that it is only after adjusting for baseline suicidal acts (in the previous 12 months before randomization), age at first deliberate self-harm, single marital status, and EQ-5D quality of life scores that these estimates reach statistical significance.

Table 4. Results of Complier-Average Causal Effect modelling

Model	Description	CBT-TAU (95% CI)	p-Value
Full (ITT)	Full (ITT)	-0.91 ($-1.67, -0.15$)	.020
More competent (>60)	Unadjusted	-1.52 ($-3.17, 0.13$)	.070
	Adjusted ^a	-1.93 ($-3.19, -0.67$)	.003
	Fully instrumented ^b	-2.05 ($-3.31, -0.80$)	.001
Quantity (>15)	Unadjusted	-1.63 ($-3.49, 0.23$)	.085
	Adjusted ^a	-2.17 ($-3.79, -0.56$)	.008
	Fully instrumented ^b	-1.43 ($-2.85, -0.01$)	.048
More competent + quantity (<15)	Unadjusted	-2.85 ($-6.04, 0.34$)	.079
	Adjusted ^a	-3.35 ($-5.64, -1.07$)	.004
	Fully instrumented ^b	-3.97 ($-6.12, -1.84$)	$<.001$
More competent + quantity (3–20)	Unadjusted	-4.14 ($-9.09, 0.79$)	.10
	Adjusted ^a	-4.73 ($-8.25, -1.22$)	.008
	Fully instrumented ^b	-5.94 ($-9.28, -2.61$)	$<.001$

Note. CBT, cognitive behavioural therapy; ITT, intention-to-treat.

^aModels adjusted for baseline suicidal acts, singleness, age at first deliberate self-harm, and EQ-5D quality-of-life score.

^bModel is as for adjusted model including baseline suicidal acts, singleness, age at first deliberate self-harm, and EQ-5D quality-of-life score, but now the interaction of treatment with each of these factors is included as additional instruments to check the exclusion restriction assumption.

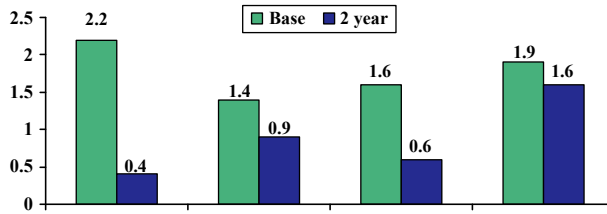


Figure 7. Suicidal acts by therapist competence and sessions attended.

Figure 7 gives a visual depiction of the interesting feature that there may be a qualitative interaction between quality and quantity of therapy, with the largest reductions coming from those who had the more competently delivered therapy, but in limited quantity. To investigate this, we combined these two variables to create the subgroup, which had more competently delivered therapy (a therapist with an average score on CTRS ≥ 60), but attended <15 sessions. The treatment estimate now changes to -3.35 suicidal acts averted every 2 years. We went one stage further and redefined the quantity as being between 3 and 20, so removing those who had so little therapy (including none) that it was difficult to see how benefit could have derived, and adding in some participants with a few more sessions (from 16 to 20). The CBTpd treatment estimate now rose to almost five suicidal acts averted.

For each model, we also give the 'fully instrumented' version, which includes the interaction of quality and quantity of treatment with the four baseline predictors. In all cases, the model holds, indicating that there is little or no evidence that the exclusion restriction is invalid here. By using these additional instruments, the IV estimate of the quantity and quality of therapy increase a bit and are more precisely estimated.

Discussion

These further analyses allow some dismantling of the previously reported ITT analyses (Davidson, Norrie *et al.*, 2006) and suggest a relationship between the quantity and quality of therapy received and suicidal behaviour. Specifically, using CACE models, we find the ITT treatment estimate of approximately one suicidal act averted over 2 years approximately doubles when treated by more competent therapists and when in receipt of over 15 therapy sessions. Intriguingly, there was an indication of a qualitative interaction between quantity and quality, with a treatment effect of approximately three suicidal acts averted (over triple the ITT estimate) for those receiving a limited amount of therapy (<15 sessions) from a competent therapist.

The investigation of a possible time lag in the treatment effect of CBTpd led to discovering an effect for inpatient hospitalization in the first 6 months only. The finding might indicate that because hospitalization often follows severe, acute self-harm behaviour, CBTpd may have a short-term effect in averting such behaviour, but it may not have a longer term effect, in contrast to the hypothesized longer term delayed treatment effect investigated for CBTpd for suicidal behaviour. This finding should be considered cautiously, however, as number of inpatient hospitalizations were low and this was a purely *post-hoc* analysis.

These findings underscore the importance of examining the effect of therapist competence and amount of therapy that may be required to improve adverse outcomes. Therapists competent in CBTpd can deliver change in patient's suicidal behaviour in 20 or fewer sessions over 1 year, and this effect remained throughout the 2-year period. This is

in contrast to the majority of studies of psychological therapy that offer highly intensive and lengthy treatment regimens (e.g., Bateman & Fonagy, 1999; Giesen-Bloo *et al.*, 2006). Our findings suggest that the precise length of therapy offered should be evaluated more rigorously.

These additional analyses raise interesting and complex methodological issues. Intention-to-treat analysis is preferred over other approaches (such as analysing the data by groups defined by what intervention was actually received) for the primary analysis of efficacy outcomes. However, participants do not receive their randomized intervention in a perfectly uniform and prescribed manner, nor do they return perfect data on all outcomes measured at all times and routinely in trials (White, 2005). Such departures from the ideal result in different (statistical) analytic approaches yielding different – and sometimes very different – estimates of treatment effects. The more non-compliance with randomized intervention, the more major protocol deviation, the more loss to follow-up, and so on, the larger the potential discrepancy between the various analyses might become. As the magnitude of these factors increases, the quality of the evidence decreases, and there is no panacea for the analysis of poor-quality data – all the analyses, including ITT, will be potentially misleading given poor data. Fortunately in BOScot, the quality of data was high with minimal loss to follow-up, and in addition, no complication of treatment crossovers. Also, we measured several ‘process variables’ related to the quality and quantity of therapy, and in this study, having explored the associations between these process variables and both baseline and outcome data, we then used formal CACE modelling to investigate what treatment effect appears to have been enjoyed by those that received the more competently delivered therapy in the appropriate quantity.

The ITT analysis addresses the issue of what benefit might accrue from the offer of one type of management over another, whereas the instrumental variables analysis tries to assess what the benefit for a subgroup who actually receive, in some sense, ‘optimized therapy’. Both analyses are therefore of value – the ITT analysis gives unbiased, rigorous scientific evidence of whether the intervention is likely to be of benefit. If the ITT analysis indicates likely benefit, then the instrumental variable regression analysis can estimate this benefit under ideal conditions, which is of interest to clinicians and patients and in addition can be helpful when comparing different psychological therapies when resources are scarce and for identifying subgroups of patients who may benefit most.

There are, however, some limitations. First, despite the high-quality data, at 106 participants, this trial is quite small, albeit large for a study of borderline personality disorder. Furthermore, our estimates of therapist competence were based on only a subset of all the sessions undertaken, and on a random sample of only 24 of the 54 patients randomized to CBTpd. Not all patients agreed to have their CBT sessions recorded, and not all of the sessions taped were sufficiently audible to rate. In addition, participants were not randomly allocated to therapists, which would have made many of the analyses we have undertaken conceptually easier. Although all therapists were trained in CBTpd for borderline personality disorder and received weekly supervision, the robustness of the trial would have been enhanced if they had been trained and then evaluated as being at or above a threshold of competency before the trial began. In a pragmatic trial, we accepted that therapists would vary in their degree of competence in delivering CBTpd. As well as having competence ratings for all the trial sessions (albeit a prohibitive amount of work), it may have been useful to have additional competence ratings for each therapist presented with the same clients (who were not randomized

into the study). Such measures of competence could have then been considered as genuine baseline covariates.

In conclusion, we have supplemented the reporting of the BOSCOT trial with this article, following the design and baseline characteristics (Davidson, Tyrer *et al.*, 2006), the ITT results (Davidson, Norrie *et al.*, 2006), and the economic evaluation (Palmer *et al.*, 2006), and the 6-year follow-up analyses (Davidson *et al.*, 2010). We have shown that if competently delivered therapy is given in the right quantity, patients can derive substantial benefit. Future trials of CBT should explore the effect of quantity and of quality of therapy on outcome.

Acknowledgements

The authors thank the 106 participants who made the study possible, and the other members of the BOSCOT research team (see Davidson, Tyrer *et al.*, 2006 for a complete roll call). The authors declare they have no conflict of interests of any description in publishing these results. BOSCOT was funded by the Wellcome Trust (064027/Z/01/Z). The funder played no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript. Kate Davidson (Chief Investigator) and John Norrie (Study Statistician) take responsibility for the integrity of the data and accuracy of the data analysis, and we confirm that all authors had full access to all the data in the study.

References

- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of American Statistical Association*, *91*, 444–472.
- Bateman, A., & Fonagy, P. (1999). Effectiveness of partial hospitalization in the treatment of borderline personality disorder: A randomized controlled trial. *American Journal of Psychiatry*, *156*, 1563–1569.
- Bellamy, S. L., Lin, J. Y., & Ten Have, T. R. (2007). An introduction to causal modelling in clinical trials. *Clinical Trials*, *4*, 58–73.
- Davidson, K. M. (2007). *Cognitive therapy for personality disorders: A guide for clinicians*. (2nd ed.) Hove, UK: Routledge.
- Davidson, K., Norrie, J., Tyrer, P., Gumley, A., Tata, P., Murray, H., & Palmer, S. (2006a). The effectiveness of Cognitive Behavior Therapy for Borderline Personality Disorder: Results from the BOSCOT Trial. *Journal of Personality Disorders*, *20*, 450–465.
- Davidson, K. M., Tyrer, P., Gumley, A., Tata, P., Norrie, J., & Palmer, S. (2006b). Rationale, description, and sample characteristics of a randomised controlled trial of Cognitive Therapy For Borderline Personality Disorder: the BOSCOT study. *Journal of Personality Disorders*, *20*, 431–449.
- Davidson, K., Tyrer, P., Norrie, J., Palmer, S., & Tyrer, H. (2010). Cognitive therapy v. usual treatment for borderline personality disorder: Prospective 6-year follow-up. *British Journal of Psychiatry*, *197*, 456–462.
- 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, J. L., Dalgard, O. S., Page, H., . . . Wilkinson, G.; ODIN group. (2003). Estimating psychological treatment effects from a randomised controlled trial 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 non-compliance and loss to follow up: The role of instrumental variable methods. *Statistical Methods in Medical Research*, *14*, 369–395.
- Giesen-Bloo, J., Van Dyck, R., Spinhoven, P., Van Tilburg, W., Dirksen, C., . . . Arntz, A. (2006). Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, *63*, 649–658.
- Medical Research Council (2000). *A framework for development and evaluation of RCTs for complex interventions to improve health*. London, UK: Author.
- Medical Research Council (2008). *Developing guidance for complex interventions – Updated guidance*. London, UK: Author.
- Palmer, S., Davidson, K., Tyrer, P., Gumley, A., Tata, P., Norrie, J., . . . Seivewright, H. (2006). The cost-effectiveness of Cognitive Behavior Therapy for Borderline Personality Disorder: Results from the BOScot Trial. *Journal of Personality Disorders*, *20*, 466–481.
- Torgerson, D., & Sibbald, D. B. (1998a). What is a patient preference trial? *British Medical Journal*, *316*, 360.
- Torgerson, D., & Sibbald, D. B. (1998b). Incorporating patient preferences into clinical trials (multiple letters). *British Medical Journal*, *317*, 78–79.
- White, I. (2005). Uses and limitations of randomised based efficacy estimators. *Statistical Methods in Medical Research*, *14*, 327–347.
- Young, J. E., & Beck, A. T. (1990). *Manual of the Cognitive Therapy Rating Scale*. University of Pennsylvania, Philadelphia, PA: Center for Cognitive Therapy.

Received 4 February 2012; revised version received 28 November 2012