Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies in UK primary-care routine practice: replication in a larger sample

William B. Stiles[1]*, Michael Barkham[2], John Mellor-Clark[3] and Janice Connell[4]

[1] *Miami University, Oxford, OH, USA*

[2] *University of Sheffield, Sheffield, UK*

[3] *CORE Information Management Systems, Rugby, UK*

[4] *University of Leeds, Leeds, UK*

Background. Psychotherapy's equivalence paradox is that treatments tend to have equivalently positive outcomes despite non-equivalent theories and techniques. We replicated an earlier comparison of treatment approaches in a sample four times larger and restricted to primary-care mental health.

Method. Patients ($n = 5613$) who received cognitive–behavioural therapy (CBT), person-centred therapy (PCT) or psychodynamic therapy (PDT) at one of 32 NHS primary-care services during a 3-year period (2002–2005) completed the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) at the beginning and end of treatment. Therapists indicated which approaches were used on an End of Therapy form. We compared outcomes of groups treated with CBT ($n = 1045$), PCT ($n = 1709$), or PDT ($n = 261$) only or with one of these plus one additional approach (e.g. *integrative*, *supportive*, *art*), designated CBT+1 ($n = 1035$), PCT+1 ($n = 1033$), or PDT+1 ($n = 530$), respectively.

Results. All six groups began treatment with equivalent CORE-OM scores, and all averaged marked improvement (overall pre/post effect size = 1.39). Neither treatment approach nor degree of purity ('only' *v.* '+1') had a statistically significant effect. Distributions of change scores were all similar.

Conclusions. Replicating the earlier results, the theoretically different approaches tended to have equivalent outcomes. Caution is warranted because of limited treatment specification, non-random assignment, incomplete data, and other issues. Insofar as these routine treatments appear effective for patients who complete them, those who fail to complete (or to begin) treatment deserve attention by researchers and policymakers.

Introduction

We sought to replicate a comparison of the outcomes of cognitive–behavioural therapy (CBT), person-centred therapy (PCT) and psychodynamic or psychoanalytic therapy (PDT) as delivered in routine primary-care mental health practice within the UK National Health Service (Stiles *et al.* 2006). CBT, PCT, and PDT are distinct approaches in terms of their usual repertoires of interventions and their assumptions about the nature and sources of psychopathology. Each encompasses a range of techniques and should be considered a family of treatments rather than a specific treatment protocol. Each is considered

by its practitioners as widely applicable to the varied problems presented for psychotherapeutic treatment (Gabbard *et al.* 2005; Feltham & Horton, 2006).

There is strong evidence for the efficacy and effectiveness of CBT for a wide variety of disorders (Dobson, 1989; Clark *et al.* 1999; Ladouceur *et al.* 2000; Teasdale *et al.* 2000; Hollon *et al.* 2002; Hollon & Beck, 2004; Ma & Teasdale, 2004; Ehlers *et al.* 2005; Westbrook & Kirk, 2005; Whittal *et al.* 2005; Butler *et al.* 2006). Fewer studies have systematically examined outcomes of the other two approaches, but available evidence similarly supports the efficacy and effectiveness of at least some varieties of PCT (Greenberg & Watson, 1998; Ward *et al.* 2000; Elliott *et al.* 2004; Goldman *et al.* 2006) and PDT (Leichsenring, 2001; Leichsenring & Leibing, 2003; Leichsenring *et al.* 2004). Clinical trials comparing alternative approaches

* Address for correspondence : Professor W. B. Stiles, Department of Psychology, Miami University, Oxford, OH 45056, USA.

(Email: stileswb@muohio.edu)

(Elkin *et al.* 1989; Shapiro *et al.* 1994; Barkham *et al.* 1996; Ward *et al.* 2000) and broadly based reviews (Wampold, 2001; Roth & Fonagy, 2004) have concluded that *bone fide* therapies that have been actively researched tend to be similarly effective. This is the equivalence paradox: many psychotherapies appear to have equivalently positive outcomes despite manifestly non-equivalent theories and techniques. The paradox is expressed by the Dodo verdict, 'Everybody has won, and *all* must have prizes' (Carroll, 1865/1946, p. 28; italics in original). This verdict has been quoted by psychotherapy researchers for more than 70 years, although debate about it continues (Rosenzweig, 1936; Luborsky *et al.* 1975; Stiles *et al.* 1986; Beutler, 1991; Norcross, 1995; Seligman, 1995; Hunsley & Di Giulio, 2002).

Despite such indications of equivalent outcomes across many treatments, the predominance of published research on CBT has given CBT a greater credibility than the other approaches. In the USA, the great majority of approaches on the list of empirically supported treatments produced by the American Psychological Association's Division 12 Task force (Chambless & Hollon, 1998; Chambless *et al.* 1998) were in the CBT family, while in the UK, guidelines proposed by the National Institute for Clinical Excellence (NICE 2004*a*, *b*, 2005) and a widely discussed proposal by Layard (2006) for improving access to psychological therapies concentrated on CBT. However, these endorsements were based explicitly on the quantity and quality of research about CBT approaches rather than their demonstrated superiority to alternative *bona fide* treatments (Hunot *et al.* 2007).

Our approach followed the logic of clinically representative, or *effectiveness* research, in which the risks of selection biases associated with lack of randomization and the lack of assurance that the treatments were delivered in a standard way are balanced by the greater realism, or external validity (Seligman, 1995). Results address the effects of treatments as routinely delivered, using practitioners' versions of the treatments and the patients who typically receive them (Shadish *et al.* 2000; Street *et al.* 2000; Stirman *et al.* 2003).

In our earlier study (Stiles *et al.* 2006), we found negligible differences in effectiveness among these alternative approaches (CBT, PCT, and PDT) as delivered in routine NHS practice. The practical importance of the question of differential effectiveness, however, mandates replication. The present comparison drew from a similar population using a later, non-overlapping sample of patients, which was more than four times larger (5613 *v.* 1309). In addition, whereas the previous sample included some patients treated in secondary and tertiary care, all in the present sample were treated in primary-care mental heath services, eliminating a possible source of confounding.

Method

Participants

We studied data from 5613 adult patients who received CBT, PCT, or PDT and completed the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM; Barkham *et al.* 1998, 2001, 2005; Evans *et al.* 2000, 2002, 2006; Cahill *et al.* 2006; Mellor-Clark *et al.* 2006; Connell *et al.* 2007) at the beginning and end of their treatment. These data were collected during a 3-year period (April 2002–September 2005) at 32 National Health Service (NHS) primary-care services delivering counselling and psychological therapy as part of routine evaluation and outcome auditing. The services each contributed from 3 to 669 of these patients (median = 85.5; 11 of the services each contributed 200 or more of the patients and five of the services each contributed fewer than 20 of the patients). These patients were seen by 399 therapists, who each saw 3–153 patients (median = 6; 90 of the therapists each saw 20 or more of these patients and 145 of the therapists each saw 3 or fewer of these patients). Therapist characteristics were not recorded.

Of these 5613 patients, 70.7% (*n* = 3970) were female, and their mean age was 40.7 years (S.D. = 12.7, range 16–99). Patients presented a variety of psychological problems, as described later. Over half of the patients (*n* = 2989, 53.3%) were taking prescribed psychotropic medications at the start of therapy, most commonly antidepressants and anxiolytics/hypnotics.

CORE system measures

Self-report outcome measure

The CORE-OM comprises 34 items addressing domains of subjective well-being, symptoms (anxiety, depression, physical problems, trauma), functioning (general functioning, close relationships, social relationships) and risk (risk to self, risk to others), half low intensity (e.g. 'I feel anxious/nervous') and half high intensity (e.g. 'I feel panic/terror'). Items are scored on a 0–4 scale, anchored *Not at all*, *Only occasionally*, *Sometimes*, *Often*, and *All or most of the time*. CORE clinical scores, computed as the mean of all completed items multiplied by 10, can range from 0 to 40. A recommended cut-off between clinical and normal populations is 10 (Connell *et al.* 2007). Forms are considered valid if up to three items are omitted

(Evans *et al.* 2002). Internal consistency reliability was $\alpha = 0.93$ based on pre-treatment forms in the present study ($n = 5613$). Test–retest reliability was $r = 0.88$ over 1 month in a clinical sample ($n = 89$, Barkham *et al.* 2007).

Therapist assessments

The CORE Assessment (Mellor-Clark *et al.* 1999; Mellor-Clark & Barkham, 2006) comprises the Therapist Assessment form, completed at intake, and the End of Therapy form. On the Therapist Assessment form, therapists gave referral information, patient demographics, and data on the nature, severity, and duration of presenting problems using 14 categories: depression, anxiety, psychosis, personality problems, cognitive/learning difficulties, eating disorder, physical problems, addictions, trauma/abuse, bereavement, self-esteem, interpersonal problems, living/welfare and work/academic.

On the End of Therapy form, therapists indicated which type(s) of therapy was (were) undertaken with the patient – as many as appropriate. Categories were psychodynamic, psychoanalytic, cognitive, behavioural, cognitive/behavioural, structured/brief, person-centred, integrative, systemic, supportive, art, and other. Therapists also reported the number of sessions attended and other aspects of the treatments.

Procedure

Data collection

Patients attending for psychological assessment or therapy at services using the Personal Computer format of the CORE System (CORE-PC; Mellor-Clark *et al.* 2006) were asked to complete a CORE-OM before treatment began – during screening or assessment or immediately before the first therapy session. Services were asked to administer CORE system measures to all such patients, although compliance was not monitored. Patients were allocated to treatments and to therapists following services' normal procedures. Services were instructed to give the post-treatment CORE-OM at the last session; the timing and specific procedures were determined by what worked best for each service administratively and were not recorded. Therapists completed the Therapist Assessment form after the intake session and the End of Therapy form when the patient was discharged or stopped attending for therapy. Patients and therapists completed forms by hand, and the data were then entered onto proprietary CORE-PC software to prepare standardized reports for the services. Each patient was allocated a unique code number by the service. With the agreement of the NHS services, the anonymized data were stored by an independent IT management system for use in research. Data collection complied with data protection procedures for the use of routinely collected clinical data. Ethics approval for this study was covered by NHS COREC application 05/Q1206/128.

Selection of patients

The 5613 patients studied were selected from the CORE National Research Database-2005 (CORE NRD-2005; Mellor-Clark *et al.* 2006), which includes information on 33 587 patients (69.4% female; mean age = 38.5, S.D. = 13.1) whose therapist returned a CORE Assessment form at one of 34 NHS primary-care counselling services during the 3-year period we considered. These services had been using CORE-PC for at least 2 years and represented approximately 50% of the NHS primary-care psychological therapy services using the CORE common methodology (Evans *et al.* 2006). The CORE NRD-2005 comprises data donated by services that were engaged in ongoing data collection, rather than a time-limited study.

Of the 33 587 patients in the CORE NRD-2005, 5327 did not return any valid CORE-OM forms; 569 returned post-therapy but not pre-therapy forms; and 14 945 returned pre-therapy but not post-therapy forms. The latter, largest category included patients who did not attend any sessions, patients who attended sessions but left without completing the final form, and patients who had not ended their treatment by the closing date of data collection. The therapists of 584 of the remaining 12 746 patients failed to complete the End of Therapy form (which included information on treatment approaches used).

From the 12 162 patients, who completed reliable pre- and post-treatment CORE-OM forms and whose therapist completed End of Therapy forms, we selected six treatment groups based on therapists' reports on the End of Therapy form regarding the type(s) of therapy undertaken. A majority of therapists indicated more than one of the 12 categories provided (mean = 1.85, range = 1–9). Following the same criteria as previously (Stiles *et al.* 2006), we classified the three targeted approaches as follows:

- CBT, cognitive, behavioural, and/or cognitive/behavioural;
- PCT, person-centred;
- PDT, psychodynamic and/or psychoanalytic.

Using these targeted approaches, we defined six groups of patients. Three groups included those whose therapists specified therapies belonging to one and only one of the targeted approaches – CBT, PCT,

Table 1. *CORE-OM clinical scores for treatment groups: pre- and post-therapy means, differences and confidence intervals for the differences*

Treatment group	n	Pre-therapy		Post-therapy		Pre/post difference		Effect size
		Mean	S.D.	Mean	S.D.	Mean	S.D.	
CBT	1045	17.3	6.7	8.6	6.5	8.7	6.8	1.38
PCT	1709	17.7	6.4	8.9	6.9	8.8	7.0	1.39
PDT	261	17.7	6.7	9.5	6.9	8.2	7.1	1.29
CBT+1	1035	17.3	6.0	8.4	5.7	8.9	6.2	1.40
PCT+1	1033	17.9	6.4	8.9	6.4	9.0	6.4	1.43
PDT+1	530	17.7	5.6	8.8	6.2	9.0	6.2	1.42

CBT, Cognitive, behavioural, or cognitive/behavioural therapy; PCT, person-centred therapy; PDT, psychodynamic/psychoanalytic therapy; CBT+1, CBT combined with one other therapy; PCT+1, PCT combined with one other therapy; PDT+1, PDT combined with one other therapy; Effect size, calculated as the mean difference divided by the pooled pre-therapy standard deviation.

or PDT. The other three groups were those whose therapists specified one of the targeted approaches plus one treatment not included in the targeted approaches (i.e. one of the following: *structured/brief, integrative, systemic, supportive, art*, or *other*), abbreviated CBT+1, PCT+1, and PDT+1, respectively. We reasoned that the latter three groups offered comparisons among the targeted approaches that were parallel to, but, depending on one's perspective, somewhat diluted or somewhat enhanced.

Of the 12 162 patients for whom the required information was available, 5613 adult patients met specifications for one of the six groups (see Table 1). In this study, we did not give further consideration to patients whose therapists indicated none of the targeted approaches ($n = 3800$) or more than one of the targeted approaches ($n = 1849$) or more than one treatment in addition to one of the targeted approaches ($n = 894$). Each of these residual categories comprised a large variety of the approaches listed on the CORE End of Therapy form. We also excluded six patients whose recorded age was below 16.

Results

Outcomes of treatment in NHS settings

Patients in these treatments showed very substantial gains, with patients improving, on average, from 17.60 (S.D. = 6.33) to 8.77 (S.D. = 6.43) on the CORE-OM, a difference of 8.83 (S.D. = 6.64). The overall treatment effect size, calculated as the mean pre/post difference divided by the pre-therapy S.D. was 1.39.

Table 1 shows the mean pre-treatment and post-treatment CORE-OM clinical scores for each of the six groups, mean differences across treatment, and effect sizes. A one-way analysis of variance (ANOVA) comparing the pre-therapy means across the six groups was not significant [$F(5, 5607) = 1.90$, $p = 0.091$, partial $\eta^2 = 0.002$ (partial η^2, calculated as effect variance divided by effect plus error variance)], indicating that all groups began treatment with approximately equivalent levels of disturbance.

To assess treatment outcomes, we conducted a repeated-measures (pre-treatment *v.* post-treatment) ANOVA, with treatment approach (CBT *v.* PCT *v.* PDT) and degree of purity (pure *v.* '+1') as fixed factors. Results showed a very large overall within-patients main effect of treatment [$F(1, 5607) = 6805.63$, $p < 0.001$, partial $\eta^2 = 0.548$], indicating that improvement across treatment accounted for a large proportion of the variation in the obtained CORE-OM scores. In this analysis, a differential treatment effect appears as a treatment by occasion of assessment (pre/post) interaction. This effect was not significant [$F(2, 5607) = 0.81$, $p = 0.446$, partial $\eta^2 < 0.001$]. The comparative effectiveness of the pure *versus* '+1' forms of treatments (the purity by occasion of assessment interaction) also failed to reach significance [$F(1, 5607) = 3.23$, $p = 0.073$, partial $\eta^2 = 0.001$], as did the three-way treatment by purity by occasion interaction, which would have indicated that purity was differentially important for the therapies [$F(2, 5607) = 0.58$, $p = 0.561$, partial $\eta^2 < 0.001$].

Fig. 1 depicts the distributions of the pre/post differences in CORE-OM clinical scores for each group in the form of notched box plots, which indicate the median, middle 50%, and range. Although these change scores ranged widely (from -29 to 40 out of the possible CORE-OM range of -40 to 40), the medians and distributions of all six groups were very similar.
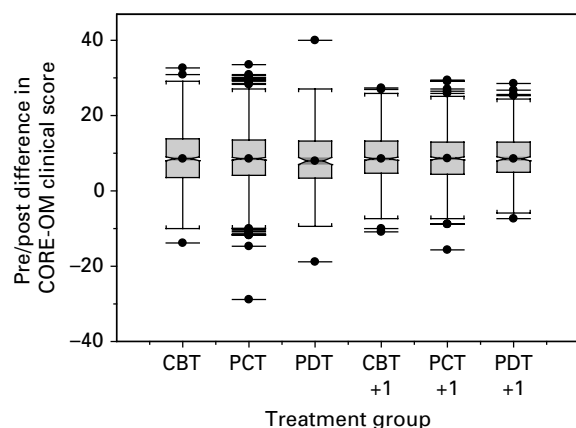
Fig. 1. Notched box plots showing pre/post differences in CORE-OM clinical scores. The notch shows the 95% confidence interval around the median. The boxes show the middle 50% of the distribution. The whiskers show the range, except that observations falling 1.5 times the interquartile range or more away from the top or bottom of the box are considered outliers and are shown separately. CBT, Cognitive, behavioural, or cognitive/behavioural therapy ($n=1045$); PCT, person-centred therapy ($n=1709$); PDT, psychodynamic/psychoanalytic therapy ($n=261$); CBT+1, CBT combined with one other therapy ($n=1035$); PCT+1, PCT combined with one other therapy ($n=1033$); PDT+1, PDT combined with one other therapy ($n=530$).

As an additional approach to the same question, following Jacobson & Truax (1991), we distinguished patients who had achieved reliable and clinically significant improvement (RCSI) as those who met two criteria. First, patients must show *reliable improvement*, defined as a pre/post difference that, when divided by the standard error of the difference, is equal to 1.96. Calculating the standard error of the difference using our S.D.$_{\text{diff}}$=6.64 and pre-treatment internal consistency reliability, $\alpha=0.93$, yielded a reliable change index of 4.9 (see Jacobson & Truax, 1991 for formulae). Second, the patient must show *clinically significant improvement* – entering treatment scoring within the clinical population and leaving treatment in the normal population – defined as moving from above to below the recommended CORE-OM clinical cut-off score of 10 (Connell *et al.* 2007).

Table 2 shows the number and percentage of patients in each treatment group who (*a*) achieved RCSI (CORE-OM scores decreasing from ⩾10 to <10, having dropped by 4.9 or more), (*b*) achieved reliable improvement only (decrease of 4.9 or more that did not fall below the cut-off of 10), (*c*) showed no reliable change, and (*d*) showed reliable deterioration, defined as an *increase* of 4.9 or more points on the CORE-OM. We restricted this analysis to the 4954 patients whose pre-therapy CORE-OM scores were ⩾10, insofar

as patients whose initial scores were below the clinical cut-off could not, by definition, achieve clinically significant improvement (Barkham *et al.* 2006). The results showed substantial RCSI rates in all groups. A 2×6 χ^2 comparing the rates of RCSI *versus* non-RCSI across the six treatment groups was not significant [$\chi^2(5)=4.14$, $p=0.530$, $n=4954$], reflecting the very similar improvement rates across the six groups.

Characteristics of patients allocated to treatment groups

Table 3 shows the distributions of presenting problems within treatment groups, as reported by the therapists just following their first contact with the patients. Note that therapists often indicated multiple problems. Broadly similar distributions of problems were treated within each group, but there appeared to be a few small differences. For example, patients seen in CBT were less likely to be reported as presenting with interpersonal problems.

Table 4 shows the gender, age, mean number of problems indicated, and mean number of sessions attended for patients in each treatment group. Gender was somewhat unevenly distributed across groups [$\chi^2(5)=25.14$, $p<0.001$, $n=5613$], with lower percentages of female patients in the CBT and PDT groups and higher percentages in the PCT+1 and PDT+1 groups. Mean age was similar in the three groups. A one-way ANOVA was nominally significant [$F(5,5607)=2.39$, $p=0.036$, partial $\eta^2=0.002$], however, none of the groups was significantly different from the others in *post-hoc* tests. The numbers of presenting problems across treatment groups varied significantly across the six groups [$F(5,5607)=31.42$, $p<0.001$, partial $\eta^2=0.027$], reflecting somewhat larger numbers of problems being attributed to patients in the PDT and PDT+1 groups. There was also significant variation in numbers of sessions attended [$F(5,5447)=25.10$, $p<.001$, partial $\eta^2=0.023$], as patients in PDT averaged a somewhat higher number of sessions than patients in other groups.

Discussion

The design improvements in this replication (larger sample and restriction to primary-care treatments) in comparison to the previous study (Stiles *et al.* 2006) yielded, if anything, a closer approximation to equivalent outcomes. For patients who completed the pre- and post-treatment forms, these therapies appeared effective, with mean pre/post improvements that approached those observed in efficacy trials

Table 2. *Reliable and clinically significant improvement (RCSI) by treatment group*

Treatment group	RCSI		Reliable improvement only		No reliable change		Deterioration	
	Frequency	Row percent	Frequency	Row percent	Frequency	Row percent	Frequency	Row percent
CBT ($n=897$)	526	58.6	156	17.4	202	22.5	13	1.4
PCT ($n=1511$)	886	58.6	272	18.0	330	21.8	23	1.5
PDT ($n=228$)	124	54.4	45	19.7	55	24.1	4	1.8
CBT+1 ($n=911$)	539	59.2	182	19.9	182	20.0	8	0.9
PCT+1 ($n=918$)	517	56.3	214	23.3	182	19.8	5	0.5
PDT+1 ($n=489$)	295	60.3	91	18.6	96	19.6	7	1.4
Total ($n=4954$)	2887	58.3	960	19.4	1047	21.1	60	1.2

$n=4954$ patients whose pre-therapy CORE-OM scores were at or above the recommended cut-off of 10 (Connell *et al.* 2007).

RCSI, Reliable and clinically significant improvement; Reliable improvement only, decrease of 4.9 or more that did not fall below the cut-off of 10 on the CORE-OM; CBT, cognitive, behavioural, or cognitive/behavioural therapy; PCT, person-centred therapy; PDT, psychodynamic/psychoanalytic therapy; CBT+1, CBT combined with one other therapy; PCT+1, PCT combined with one other therapy; PDT+1, PDT combined with one other therapy.

Table 3. *Percentage of patients in each treatment group with indicated presenting problems*

Presenting problem	Treatment group					
	CBT	PCT	PDT	CBT+1	PCT+1	PDT+1
Anxiety	79.2	74.7	71.6	77.3	78.3	77.5
Depression	63.3	67.7	72.4	67.6	73.0	75.3
Interpersonal problems	35.2	52.9	59.0	42.2	52.5	54.2
Self-esteem	37.9	48.3	45.6	42.6	52.5	51.7
Bereavement/Loss	18.3	38.9	42.9	25.3	35.6	42.1
Work/Academic	21.5	17.4	23.0	21.0	18.2	25.1
Trauma/Abuse	15.3	16.3	22.2	15.7	17.1	22.1
Living/Welfare	9.8	11.4	9.6	12.8	11.1	29.1
Addictions	3.2	2.5	8.0	3.9	2.8	4.3
Eating disorder	3.6	2.1	5.0	3.7	1.8	3.6
Personality problems	2.7	1.2	2.3	3.9	1.5	2.8
Psychosis	0.6	0.2	0.8	0.7	0.7	0.2
No. of patients in group	1045	1709	261	1035	1033	530

Based on $n=5613$ patients.

CBT, Cognitive, behavioural, or cognitive/behavioural therapy; PCT, person-centred therapy; PDT, psychodynamic/psychoanalytic therapy; CBT+1, CBT combined with one other therapy; PCT+1, PCT combined with one other therapy; PDT+1, PDT combined with one other therapy.

Columns in the presenting problems section add to more than 100% because therapists often indicated multiple problems for some patients.

(M. Barkham *et al.* unpublished observations; Hunsley & Lee, 2007). The mean differences among the three targeted approaches, CBT, PCT, and PDT, did not approach significance, despite the high statistical power of this test, nor did the differences between the pure and the diluted/enhanced ('+1') treatment groups (Table 1). The RCSI rates too were closely similar across these treatments (Table 2).

In replicating the earlier results, the present results uphold the 70-year-old Dodo verdict on this issue (cf. Rosenzweig, 1936). More immediately, they extend findings from randomized trials conducted within the NHS (Barkham *et al.* 1996; Ward *et al.* 2000), confirming that the equivalence of these treatments may also be observed in routine practice. These replicated results may be of particular interest to

Table 4. *Demographics and mean numbers of presenting problems and sessions attended by treatment groups*

Treatment group	Percent female	Mean age (years)	Mean of number of presenting problems	Mean number of sessions attended
CBT ($n = 1045$)	66.0	39.8	2.91	7.30
PCT ($n = 1709$)	71.5	41.2	3.34	6.82
PDT ($n = 261$)	67.4	40.5	3.62	8.06
CBT + 1 ($n = 1035$)	69.3	40.2	3.17	6.10
PCT + 1 ($n = 1033$)	74.2	41.2	3.45	5.80
PDT + 1 ($n = 530$)	75.1	40.8	3.87	5.97

Based on $n = 5613$ patients.

CBT, Cognitive, behavioural, or cognitive/behavioural therapy; PCT, person-centred therapy; PDT, psychodynamic/psychoanalytic therapy; CBT + 1, CBT combined with one other therapy; PCT + 1, PCT combined with one other therapy; PDT + 1, PDT combined with one other therapy.

practitioners of PCT and PDT, insofar as these approaches' comparable effectiveness to CBT in routine practice may have been unappreciated (cf. Holmes, 2002).

As Fig. 1 shows, the distributions of change scores were mostly overlapping. There was a great deal of variation in outcomes within each group, however, contrasting with the lack of differences between groups. Thus, much scope remains for research on sources of variation in psychotherapy outcome, including differences among therapists (Lutz *et al.* 2007), patients, problems, and contexts.

In addition to the noted design improvements, there were incremental improvements in data gathering in the CORE NRD-2005, compared with the database used in the previous study (Stiles *et al.* 2006). A larger percentage of patients returned valid pre-treatment and post-treatment CORE-OM forms than previously (38% or 12746/33587 *v.* 33% or 3424/10351), and the larger sample was drawn from a smaller set of services (34 *v.* 58), representing more therapists from each site and larger percentages of therapists' caseloads.

Limitations and alternative accounts

Despite this study's improvements, many of the limitations and caveats of the previous study (Stiles *et al.* 2006) apply to this one as well. All are common in research on routine treatments (Shadish *et al.* 2000; Westbrook & Kirk, 2005).

Limited specification of treatments and therapist responsiveness

As previously (Stiles *et al.* 2006), we had no independent check on whether or how the therapists delivered the treatments they indicated, no precise descriptions of what the treatments comprised, and no details of

the therapists' qualifications. The therapists received no special training for this study. The observed equivalent outcomes could, in principle, reflect a lack of differences in how the treatments were conducted. Or, systematic failure to implement a treatment correctly could account for any particular treatment having failed to prove superior.

On the other hand, we know of no reason why the therapists would have misrepresented the theoretical approach from which they worked. The most plausible assumption, we suggest, is that they sincerely sought to implement the approach they specified. To put it another way, for better or worse, these treatments represent the way CBT, PCT, and PDT are currently practised in these settings.

Presumably, any concern about deviations from the indicated approach(es) applies to all of the treatments, although proponents may be differentially sensitive to possible deviations within their favoured approach. The observation that patients in the diluted/enhanced treatments improved as much as those in the pure treatments suggests that, across the range of purity represented by our comparisons and given the limitations of self-report, greater purity did not yield better outcomes.

Even psychotherapies that specify component techniques are far from standardized. More than most medical treatments, psychotherapies must be adapted to the emerging needs of varied patients in ways that are not specified in a protocol but depend on the skill and interpersonal responsiveness of the therapist (Stiles *et al.* 1998). Therapists responsively vary their interventions depending on patients' backgrounds, circumstances, capacities, and personalities as well as on the nature and severity of patients' presenting problems (Hardy *et al.* 1998; M. Barkham *et al.* unpublished observations).

Non-random assignment and equivalence of treatment groups

Because patients were not randomly assigned to treatments, we cannot rule out biased assignment as an explanation for the observed equivalence. In principle, assigning the most disturbed patients to the most effective therapies and vice versa might mask outcome differences that would emerge in a randomized trial. Alternatives to the Dodo verdict could be built on scenarios that explain how a selective bias differentially penalized an approach that would otherwise have proved superior, so that it appeared equivalent to the other approaches.

Most importantly, however, the groups had equivalent initial scores on the CORE-OM – more equivalent than is typically the case in small-sample randomized trials – indicating that the groups began with similar overall levels of disturbance (Table 1). There were some other modest differences in the patient mix across groups. Replicating the previous results (Stiles *et al.* 2006), the largest proportion of males was assigned to the CBT group, the largest average numbers of problems were in the PDT and PDT + 1 groups, and patients in the PDT group tended to receive more sessions than patients in the other groups (Table 4). Distributions of presenting problems were broadly similar across groups (Table 3) and in any case were based on therapists' reports, so they could reflect differences in interviewing styles or case conceptualization (e.g. PCT and PDT therapists might be more inclined than CBT therapists to consider interpersonal problems as central). To us, the nature and magnitude of these differences did not suggest compelling scenarios in which otherwise superior treatments were differentially penalized.

Lack of experimental control

Because there was no control group of untreated patients, we cannot attribute patients' improvement to the treatments with certainty. Of course, this is usual in naturalistic studies (cf. Corney & Simpson, 2005). In another NHS primary-care sample, however, CORE-OM scores remained reasonably stable over periods of up to a year among patients waiting to be assessed for therapy (high test–retest correlations and negligible mean change from screening scores among patients who returned for treatment; Barkham *et al.* 2007).

In the same vein, we had little control over procedures used to collect data, which were gathered by participating services for other purposes as part of an ongoing evaluation process. Although we had no indication of biased procedures that would have produced a spurious equivalence of outcomes, we can not rule out this possibility.

Incomplete data, the possibility of selective reporting, and the case for improved access

Whereas 38% of patients represented in the CORE NRD-2005 (12 746/33 587) returned valid pre-treatment and post-treatment CORE-OM forms, 44% (14 945/33 587) returned valid pre-therapy but not post-therapy forms. Incomplete data are common in routine practice settings (Stiles *et al.* 2003; Gilbert *et al.* 2005; Greasley & Small, 2005), and we emphasize that our conclusions are limited to patients who completed post-treatment forms. Among patients who do receive treatment, those who complete post-treatment measures are more likely to have agreed with their therapist about when treatment should end (Barkham *et al.* 2006) and more likely to have improved more during treatment (Stiles *et al.* 2003) than are patients who fail to complete them.

The low percentage of patients with post-treatment CORE-OM scores raised concerns about possible selective reporting. In an analysis not yet published, we addressed one concern by examining data from the 343 therapists who saw 15 or more of the patients; these included 31 966 (95%) of the patients in the CORE NRD-2005. If therapists were selectively influencing their good-outcome patients to complete post-treatment forms, improvement rates would be negatively correlated with reporting rates, i.e. the more selective therapists would tend to have relatively better improvement rates. Results showed that the therapists varied widely in their reporting rates (proportions of patients with post-treatment forms) and in their patients' rates of improvement (mean CORE-OM change scores and proportions of patients who achieved RCSI). The return rates were essentially uncorrelated with improvement rates across therapists, however, suggesting little or no selective reporting of their good-outcome cases after all; or perhaps therapists tried but failed. Successful selective reporting would require therapists to know which of their cases would do well or poorly on their post-treatment CORE-OM. However, other research suggests that most therapists think most of their cases did very well and have no accurate idea of which of their patients did well or poorly (Hunsley *et al.* 1999; Hannan *et al.* 2005). If therapists can not discriminate successful from unsuccessful patients, as these results suggest, they could not select the successful ones.

Patients who drop out before completing treatment or who never begin treatment in the first place are of great policy interest and deserve attention by researchers. The substantial mean improvement we observed among those who did complete treatment (and post-treatment measures) is a strong argument

for improving access to psychotherapy (cf. Layard, 2006), working to overcome economic, social, and personal barriers to treatment.

Restriction to one self-report measure

In principle, self-report instruments are vulnerable to distortions. Patients may exaggerate their distress before treatment or exaggerate their improvement following treatment. The CORE-OM is a broad spectrum measure and, in our use of it, did not focus on the specific problems. Qualitatively different results of the different treatments might have shown up on more targeted measures.

Arguably, however, subjective symptoms and distress define the need for psychotherapeutic treatment in most cases. The CORE-OM is highly correlated with other self-report and clinician-rated measures of outcome (Evans *et al.* 2002; Leach *et al.* 2005, 2006; Cahill *et al.* 2006). Additional measures would be informative but, the gain must be balanced against the burden they would impose on the patients and the system.

Investigator allegiance

In a review of psychotherapy outcome studies, Luborsky *et al.* (1999) reported that the investigator's allegiance, assessed by ratings of previous publications, ratings by colleagues, and self-ratings, was strongly correlated with the outcomes of the treatments in published reports ($r = 0.85$, $p < 0.001$, $n = 29$ studies). The present study's first author has published papers dealing with the equivalence paradox, so in a sense this report, like the previous one (Stiles *et al.* 2006), fits the pattern. He has a particular interest in the PCT approach. The second author is an accredited CBT therapist and has delivered manualized versions of both CBT and PDT in randomized trials. The third and fourth authors are not qualified therapists, although the third belongs to an organization whose members have predominantly PCT orientations.

Self-regulation as an account of equivalent outcomes

Our results underline the call for more research on non-CBT approaches (Hunot *et al.* 2007). Equivalent outcomes, such as those replicated in this study, are often ascribed to common factors in the relationship, such as the alliance, empathy, and collaborative involvement (Norcross, 2002). To us, however, the outcomes of these treatments seemed so remarkably similar that we wondered if active self-regulatory processes could be responsible (Stiles *et al.* 1998).

Perhaps the equivalent results reflect therapists and patients responsively optimizing gains. In analyses not yet published, we found that average patient outcomes in the CORE NRD-2005 were equivalent or declined across increasing numbers of sessions. That is, the average outcomes of patients who had one or two sessions were at least as positive as those of patients who had 15 or 16 sessions. This replicated result (cf. Barkham *et al.* 2006) may seem paradoxical and surprising if treatment is considered as an independent variable in an experimental manipulation, but it seems clinically sensible if patients and therapists are considered as responsively ending treatment when they have reached a satisfactory balance between gains achieved and further effort required. Insofar as different patients begin at different points and change at different rates, patients achieve satisfactory gains at different treatment durations.

If active, responsive self-regulation determines the level of gains achieved, then the type of approach (CBT, PCT, PDT, or other) might be incidental to the gross degree of improvement. Yet, the logic of self-regulation does not contradict the treatment theories. Alternative approaches may offer equally effective but different solutions to psychological problems (Stiles, 1983).

References

Barkham M, Connell J, Stiles WB, Miles JNV, Margison F, Evans C, Mellor-Clark J (2006). Dose-effect relations and responsive regulation of treatment duration: the good enough level. *Journal of Consulting and Clinical Psychology* **74**, 160–167.

Barkham M, Evans C, Margison F, McGrath G, Mellor-Clark J, Milne D, Connell J (1998). The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health* **7**, 35–47.

Barkham M, Gilbert N, Connell J, Marshall C, Twigg E (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in

psychological therapy services based in primary and secondary care settings. *British Journal of Psychiatry* **186**, 239–246.

Barkham M, Margison F, Leach C, Lucock M, Mellor-Clark J, Evans C, Benson L, Connell J, Audin K, McGrath G (2001). Service profiling and outcomes benchmarking using the CORE-OM: towards practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology* **69**, 184–196.

Barkham M, Mullin T, Leach C, Stiles WB, Lucock M (2007). Stability of the CORE-OM and BDI-I prior to therapy: evidence from routine practice. *Psychology and Psychotherapy: Theory, Research and Practice* **80**, 269–278.

Barkham M, Rees A, Shapiro DA, Stiles WB, Agnew RM, Halstead J, Culverwell A, Harrington VMG (1996). Outcomes of time-limited psychotherapy in applied settings: replicating the Second Sheffield Psychotherapy Project. *Journal of Consulting and Clinical Psychology* **64**, 1079–1085.

Beutler LE (1991). Have all won and must all have prizes? Revisiting Luborsky *et al.*'s verdict. *Journal of Consulting and Clinical Psychology* **59**, 226–232.

Butler AC, Chapman JE, Forman EM, Beck AT (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review* **26**, 17–31.

Cahill J, Barkham M, Stiles WB, Twigg E, Hardy GE, Rees A, Evans C (2006). Convergent validity of the CORE Measures with measures of depression for clients in cognitive therapy for depression. *Journal of Counseling Psychology* **53**, 253–259.

Carroll L (1946). *Alice's Adventures in Wonderland*. Random House: New York. (Original work published 1865.)

Chambless DL, Baker MJ, Baucom DH, Beutler LE, Calhoun KS, Crits-Christoph P, Daiuto A, DeRubeis R, Detweiler J, Haaga DAF, Johnson SB, McCurry S, Mueser KT, Pope KS, Sanderson WC, Shoham V, Stickle T, Williams DA, Woody SR (1998). Update on empirically validated therapies, II. *The Clinical Psychologist* **51**, 3–16.

Chambless DL, Hollon SD (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology* **66**, 7–18.

Clark DM, Salkovskis PM, Hackmann A, Wells A, Ludgate J, Gelder M (1999). Brief cognitive therapy for panic disorder: a randomized controlled trial. *Journal of Consulting and Clinical Psychology* **67**, 583–589.

Connell J, Barkham M, Stiles WB, Twigg E, Singleton N, Evans O, Miles JNV (2007). Distribution of CORE-OM scores in a general population, clinical cut-off points, and comparison with the CIS-R. *British Journal of Psychiatry* **190**, 69–74.

Corney R, Simpson S (2005). Thirty-six month outcome data from a trial of counselling with chronically depressed patients in a general practice setting. *Psychology and Psychotherapy: Theory, Research and Practice* **78**, 127–138.

Dobson K (1989). A meta-analysis of the efficacy of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology* **57**, 414–420.

Ehlers A, Clark DM, Hackmann A, McManus F, Fennell M (2005). Cognitive therapy for post-traumatic stress disorder: development and evaluation. *Behaviour Research and Therapy* **43**, 413–431.

Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry* **46**, 971–982.

Elliott R, Greenberg LS, Lietaer G (2004). Research on experiential psychotherapies. In *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*, 5th edn (ed. M. J. Lambert), pp. 493–539. Wiley: New York.

Evans C, Connell J, Barkham M, Margison F, Mellor-Clark J, McGrath G, Audin K (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry* **180**, 51–60.

Evans C, Mellor-Clark J, Margison F, Barkham M, Audin K, Connell J, McGrath G (2000). CORE: Clinical outcomes in routine evaluation. *Journal of Mental Health* **9**, 247–255.

Evans R, Mellor-Clark J, Barkham M, Mothersole G (2006). Developing the resources and management support for routine evaluation in counselling and psychological therapy service provision: reflections on a decade of CORE development. *European Journal of Psychotherapy and Counselling* **8**, 141–161.

Feltham C, Horton IE (eds) (2006). *Handbook of Counselling and Psychotherapy*, 2nd edn. Sage: London.

Gabbard GO, Beck JS, Holmes J (eds) (2005). *Oxford Textbook of Psychotherapy*. Oxford University Press: Oxford, UK.

Gilbert N, Barkham M, Richards A, Cameron I (2005). The effectiveness of a primary care mental health service delivering brief psychological interventions: a benchmarking study using the CORE System. *Primary Care Mental Health* **3**, 241–251.

Goldman RN, Greenberg LS, Angus L (2006). The effects of adding emotion-focused interventions to the therapeutic relationship in the treatment of depression. *Psychotherapy Research* **16**, 537–549.

Greasley P, Small N (2005). Evaluating a primary care counselling service: outcomes and issues. *Primary Care Research and Development* **6**, 124–136.

Greenberg LS, Watson JC (1998). Experiential therapy of depression: differential effects of client-centered relationship conditions and process experiential interventions. *Psychotherapy-Research* **8**, 210–224.

Hannan C, Lambert MJ, Harmon C, Nielsen SL, Smart DW, Shimokawa K, Sutton SW (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology* **61**, 155–163.

Hardy GE, Stiles WB, Barkham M, Startup M (1998). Therapist responsiveness to client interpersonal styles during time-limited treatments for depression. *Journal of Consulting and Clinical Psychology* **66**, 304–312.

Hollon SD, Beck AT (2004). Cognitive and cognitive behavioral therapies. In *Garfield and Bergin's Handbook of*

Psychotherapy and Behavior Change, 5th edn (ed. M. J. Lambert), pp. 447–492. Wiley : New York.

Hollon SD, Thase ME, Markowitz JC (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest* **3**, 39–77.

Holmes J (2002). All you need is cognitive-behaviour therapy? *British Medical Journal* **324**, 288–290.

Hunot V, Churchill R, Silva de Lima M, Teixeira V (2007). Psychological therapies for generalised anxiety disorder. *Cochrane Database of Systematic Reviews*. Art. No.: CD001848. doi:10.1002/14651858.CD001848.pub4.

Hunsley J, Aubry TD, Verstervelt CM, Vito D (1999). Comparing therapist and client perspectives on reasons for psychotherapy termination. *Psychotherapy* **36**, 380–388.

Hunsley J, Di Giulio G (2002). Dodo Bird, Phoenix, or Urban Legend? The question of psychotherapy equivalence. *Scientific Review of Mental Health Practice* **1**, 11–22.

Hunsley J, Lee CM (2007). Research-informed benchmarks for psychological treatments : efficacy studies, effectiveness studies, and beyond. *Professional Psychology : Research and Practice* **38**, 21–33.

Jacobson NS, Truax P (1991). Clinical significance : a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* **59**, 12–19.

Ladouceur R, Dugas MJ, Freeston MH, Leger E, Gagnon F, Thibodeau N (2000). Efficacy of a cognitive-behavioural treatment for generalized anxiety disorder : evaluation in a controlled clinical trial. *Journal of Consulting and Clinical Psychology* **68**, 957–964.

Layard R (2006). The case for psychological treatment centres. *British Medical Journal* **332**, 1030–1032.

Leach C, Lucock M, Barkham M, Noble R, Clarke L, Iveson S (2005). Assessing risk and emotional disturbance using the CORE-OM & HoNOS outcome measures. *Psychiatric Bulletin* **29**, 419–422.

Leach C, Lucock M, Barkham M, Stiles WB, Noble R, Iveson S (2006). Transforming between Beck Depression Inventory and CORE-OM scores in routine clinical practice. *British Journal of Clinical Psychology* **45**, 153–166.

Leichsenring F (2001). Comparative effects of short-term psychodynamic psychotherapy and cognitive-behavioral therapy in depression : a meta-analytic approach. *Clinical Psychology Review* **21**, 401–419.

Leichsenring F, Leibing E (2003). The effectiveness of psychodynamic therapy and cognitive behavior therapy in the treatment of personality disorders : a meta-analysis. *American Journal of Psychiatry* **160**, 1223–1232.

Leichsenring F, Rabung S, Leibing E (2004). The efficacy of short-term psychodynamic psychotherapy in specific psychiatric disorders : a meta-analysis. *Archives of General Psychiatry* **6**, 1208–1216.

Luborsky L, Diguer L, Seligman DA, Rosenthal R, Krause ED, Johnson S, Halperin G, Bishop M, Berman JS, Schweizer E (1999). The researcher's own therapy allegiances : a 'wild card' in comparisons of treatment efficacy. *Clinical Psychology : Science and Practice* **6**, 95–106.

Luborsky L, Singer B, Luborsky L (1975). Comparative studies of psychotherapies : is it true that 'Everyone has won and all must have prizes'? *Archives of General Psychiatry* **32**, 995–1008.

Lutz W, Leon SC, Martinovich Z, Lyons JS, Stiles WB (2007). Therapist effects and therapist performance clusters in outpatient psychotherapy. *Journal of Counseling Psychology* **54**, 32–39.

Ma SH, Teasdale JD (2004). Mindfulness-based cognitive therapy for depression : replication and exploration of differential relapse prevention effects. *Journal of Consulting and Clinical Psychology* **72**, 31–40.

Mellor-Clark J, Barkham M (2006) Quality evaluation : methods, measures and meaning. In *Handbook of Counselling and Psychotherapy*, 2nd edn (ed. C. Feltham and I. Horton), pp. 207–224. Sage Publications : London.

Mellor-Clark J, Barkham M, Connell J, Evans C (1999). Practice-based evidence and standardized evaluation : informing the design of the CORE system. *European Journal of Psychotherapy, Counselling and Health* **2**, 357–374.

Mellor-Clark J, Curtis Jenkins A, Evans R, Mothersole G, McInnes B (2006). Resourcing a CORE network to develop a National Research Database to help enhance psychological therapy and counselling service provision. *Counselling and Psychotherapy Research* **6**, 16–22.

NICE (2004*a*). Anxiety : Management of anxiety (panic disorder with our without agoraphobia, and generalised anxiety disorder) in adults in primary, secondary and community care (Clinical guideline 23). National Institute for Clinical Excellence : London.

NICE (2004*b*). Depression : Management of depression in primary and secondary care (Clinical guideline 22). National Institute for Clinical Excellence : London.

NICE (2005). Post-traumatic stress disorder (PTSD) : the management of PTSD in adults and children in primary and secondary care (Clinical guideline 26). National Institute for Clinical Excellence : London.

Norcross JC (1995). Dispelling the Dodo bird verdict and the exclusivity myth in psychotherapy. *Psychotherapy* **32**, 500–504.

Norcross JC (ed.) (2002). *Psychotherapy Relationships that Work : Therapist Contributions and Responsiveness to Patient Needs*. Oxford University Press : New York.

Roth A, Fonagy P (2004). *What Works For Whom? A Critical Review of Psychotherapy Research*, 2nd edn. Guilford Publications : New York.

Rosenzweig S (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry* **6**, 412–415.

Seligman M (1995). The effectiveness of psychotherapy. *American Psychologist* **50**, 965–974.

Shadish WR, Navarro AM, Matt GE, Phillips G (2000). The effects of psychological therapies under clinically representative conditions : a meta-analysis. *Psychological Bulletin* **126**, 512–529.

Shapiro DA, Barkham M, Rees A, Hardy GE, Reynolds S, Startup M (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioural and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology* **62**, 522–534.

Stiles WB (1983). Normality, diversity, and psychotherapy. *Psychotherapy: Theory, Research, and Practice* **20**, 183–189.

Stiles WB, Barkham M, Twigg E, Mellor-Clark J, Cooper M (2006). Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies as practiced in UK National Health Service settings. *Psychological Medicine* **36**, 555–566.

Stiles WB, Honos-Webb L, Surko M (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice* **5**, 439–458.

Stiles WB, Leach C, Barkham M, Lucock M, Iveson S, Shapiro DA, Iveson M, Hardy G (2003). Early sudden gains in psychotherapy under routine clinic conditions: practice-based evidence. *Journal of Consulting and Clinical Psychology* **71**, 14–21.

Stiles WB, Shapiro DA, Elliott R (1986). 'Are all psychotherapies equivalent?' *American Psychologist* **41**, 165–180.

Stirman SW, DeRubeis RJ, Crits-Christoph P, Brody PE (2003). Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *Journal of Consulting and Clinical Psychology* **71**, 963–972.

Street L, Niederehe G, Lebowitz B (2000). Toward greater public health relevance for psychotherapeutic intervention research: an NIMH workshop report. *Clinical Psychology: Science and Practice* **7**, 127–137.

Teasdale JD, Segal ZV, Williams JM, Ridgeway VA, Soulsby JM, Lau MA (2000). Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. *Journal of Consulting and Clinical Psychology* **68**, 615–623.

Wampold BE (2001). *The Great Psychotherapy Debate: Models, Methods, and Findings.* Lawrence Erlbaum Associates: Mahwah, NJ.

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**, 1383–1388.

Westbrook D, Kirk J (2005). The clinical effectiveness of cognitive behaviour therapy: outcome for a large sample of adults treated in routine practice. *Behaviour Research and Therapy* **43**, 1243–1261.

Whittal ML, Thordarson DS, McLean PD (2005). Treatment of obsessive-compulsive disorder: cognitive behavior therapy vs. exposure and response prevention. *Behavior Research and Therapy* **43**, 1559–1576.