

## Psychological treatment outcomes in routine NHS services: a commentary on Stiles *et al.* (2007)

D. M. Clark<sup>1\*</sup>, C. G. Fairburn<sup>2</sup> and S. Wessely<sup>1</sup>

<sup>1</sup> Institute of Psychiatry at Kings College London, London, UK

<sup>2</sup> Department of Psychiatry, University of Oxford, Oxford, UK

Received 31 August 2007; Accepted 10 September 2007; First published online 9 October 2007

**Key words:** Cognitive-behaviour therapy, CORE-OM, person-centred therapy, primary care, psychodynamic therapy, treatment outcomes.

Following the Bristol enquiry into the care of children with congenital heart disease, NHS cardiovascular units now make their surgical survival rates available to the public through a website ([www.ccad.org.uk/congenital](http://www.ccad.org.uk/congenital)) with suitable advice about how the data can, and cannot, be interpreted. Sadly, nothing comparable exists for members of the public who are suffering from mental illnesses and wish to know what their chance of recovery is if they take up the offer of treatment X in service Y. This is not simply because NHS mental health services do not make their outcomes available to the public. In many cases, it is because the outcomes are not even monitored. For example, a recent survey of British psychiatrists (Gilbody *et al.* 2002) found that only 11% routinely used standardized measures to assess clinical change in their patients and a majority (58%) had never used such instruments. Clearly, there is a long way to go.

In the present issue, Stiles *et al.* (2007) report a welcome exception. For a number of years, this group have been advocating the use of the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM; Evans *et al.* 2000) to routinely measure outcomes in patients with common mental problems (especially anxiety, depression and interpersonal difficulties) who are receiving treatment in the NHS. Their strenuous efforts to overcome resistance to routine outcome monitoring are exemplary and they deserve enormous credit for the way in which they have moved the field forward. As a direct result of their work, a substantial number of NHS primary-care counselling services, and other psychological treatment services, now aim to give their patients self-report measures of their clinical state at pre- and post-treatment. While this is

a very encouraging development, it is important to realize that the data that have so far been collected are incomplete in key respects and this poses severe limits on their interpretation. In our view, Stiles *et al.*'s study reported in this issue, and the earlier study (Stiles *et al.* 2006) with a smaller sample that it replicates, go well beyond these limits and, as a consequence, conclusions are drawn that are not warranted and risk being misinterpreted.

The aim of this second study was to evaluate the effectiveness, as measured by CORE-OM scores, of three different therapies as they are practised in NHS primary-care counselling services. The design, which was essentially the same as that employed in the earlier study (Stiles *et al.* 2006), is a non-randomized (naturalistic) comparison of patients whose therapy was described by their therapist as falling within the broad categories of: cognitive-behaviour therapy (CBT); person-centred therapy (PCT) or psychodynamic and/or psychoanalytic therapy (PDT) or alternatively one of those categories plus no more than one other therapy approach. The data were collected by encouraging therapists to use CORE-OM with their patients and to anonymously submit the questionnaires to a central database.

No information is given about the proportion of each therapist's caseload that received CORE-OMs and was submitted to the database. However, from the small numbers of cases that were submitted by many therapists (over a 3.5-year data collection period the median number of cases submitted by each therapist was only six), it is clear that not everyone submitted all of their cases. The analyses focused on 5613 submitted patients who had completed CORE-OMs at pre-treatment and post-treatment and whose therapist had completed an End-of-Therapy Form (which identified the type of therapy). We are told this number constitutes 38% of the patients who were submitted to the database. This is because many patients who were

\* Address for correspondence: Professor D. M. Clark, Department of Psychology, Henry Wellcome Building (PO 77), Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK.  
(Email: [d.clark@iop.kcl.ac.uk](mailto:d.clark@iop.kcl.ac.uk))

submitted to the database only completed a CORE-OM at pre-treatment or at post-treatment, but not on both occasions.

The main findings were as follows: (1) the patients who were included in the analyses showed substantial pre-treatment to post-treatment improvement (uncontrolled effect size = 1.39; reliable and clinically significant improvement rate = 58%) and (2) there were no significant differences between the three treatment categories which had very similar improvement rates. The authors conclude from these findings: (1) all three treatments are effective, (2) the treatments do not differ in effectiveness, and (3) the treatments as currently delivered in primary care are doing about as well as they do in tightly controlled clinical trials, when such data is available. The authors frankly acknowledge methodological shortcomings in their Discussion (as they did in the previous report). However, they feel that their results are still interpretable. We disagree. Below we list and review the most serious methodological limitations and indicate why we think they seriously compromise the conclusions drawn from this study and the earlier one.

### Missing cases

The 'unique selling point' of this study is that it represents normal NHS primary-care counselling services' practice of psychological treatments. However, the sample is less than 38% of the cases seen by the therapists in the study and the authors cite one of their own studies (Stiles *et al.* 2003) as demonstrating that patients who complete pre- and post-treatment measures are more likely to have improved than patients who fail to do so. This means that the true pre-treatment to post-treatment improvement in these NHS services must be substantially less than that reported here, even if it is difficult to know just how much less. In randomized controlled trials (RCTs), it is common to carry the last data-point forward to estimate post-treatment clinical status in individuals for whom post-treatment data are missing. If this discipline were applied to the present data, carrying forward pre-treatment scores for the 62% of cases with missing post-treatment data would drop the overall recovery rate (termed 'reliable and clinically significant improvement') from 58% to the strikingly lower figure of 22%.

We note that the authors report there was no significant correlation within a selected set of therapists between the proportion of their patients that were entered into the database with pre- and post-treatment scores and their overall improvement rates. This does not help to deal with the missing data problem because it is just testing a possible prediction from one

very restrictive hypothesis about the mechanism of any selective reporting. There are numerous selection effects whose operation would not imply such a correlation. In addition, the prediction seems to assume that all therapists are similarly effective if one had all of their data, a position that the authors have elsewhere (Stiles *et al.* 2006) argued strongly against.

### No control for other causes of recovery

Whatever the true recovery rates for patients seen in the services that formed part of the project, it is important to realize that they do not only reflect the effect of receiving a psychological treatment. Some of the patients who had psychological treatment would have recovered in the same period of time without that treatment. How many depends on the nature, severity and duration of their problems. The RCTs that led the National Institute for Clinical Excellence (NICE) to advocate the use of CBT for many common mental health problems mainly focused on chronic cases with diagnosed mental disorders, many of whom were seen in secondary care. Within these groups of patients, 'wait-list'/GP treatment as usual' control conditions show little 'natural recovery' and CBT shows clear efficacy. However, in primary-care settings many patients are likely to be seen early in an episode and it is known that: (1) natural recovery may be common in recent-onset cases, and (2) as a consequence, the added benefits of psychological treatment may be modest or non-existent. In part this is regression to the mean or the 'physician's friend'. Three studies illustrate these points. Catalan *et al.* (1984) studied new cases of anxiety or depression in primary care and found that 60% had recovered after 4 weeks and 70% had recovered after 6 months of usual GP care (once-monthly 12-minute consultations). These rates are as high as the improvement rates reported in this study, even before they are corrected for the over-estimate that will have resulted from the neglect of missing data. Fletcher *et al.* (2005) randomized mild to moderate cases of anxiety and/or depression in primary care to self-help bibliotherapy for 12 weeks or a wait-list control condition. The CORE-OM was used to assess outcome. Patients generally rated the bibliotherapy as helpful but the recovery rate for the wait-list control group (40%) was not different from that of the bibliotherapy group (46%). Finally, Kendrick *et al.* (2006) randomized recent-onset cases (4 weeks to 6 months) of anxiety or depression in primary care to treatment as usual by GPs or referral to community mental health nurses who provided NICE (2004b) guidance-advocated problem-solving treatment. The full sample (intention-to-treat) pre-treatment to post-treatment effect size on the main measure (the General Health

Questionnaire) was very large (2.04) but not significantly different to GP treatment as usual, which was also associated with very large improvements.

The end-of-treatment questionnaire that therapists competed in this study and the earlier study (Stiles *et al.* 2006) included an item covering the duration of the current problem but this data is not reported in the papers or utilized in the analyses. However, our own experience of primary-care therapy services is that a substantial proportion of the caseload can be recent-onset cases, so failure to control for 'natural recovery' is a crucial problem. Of course, this does not mean that one should not intervene early in an episode. There are studies that have demonstrated incremental benefits of psychological treatment in recent-onset cases (Ehlers *et al.* 2003) but this needs to be demonstrated, rather than assumed.

In addition to the problem of natural recovery, failure to control for the effects of concurrent medication (53% of patients were also taking psychotropics) is a further reason for suspecting that the CORE-OM data in the two Stiles *et al.* (2006, 2007) studies will have overestimated the effects of each psychological treatment. This is likely to be particularly problematic if patients started medication around the same time that they started psychological therapy, a practice that is common in some primary-care settings.

#### **Lack of randomization to the different treatments**

The great strength of randomizing patients to treatments is the confidence that it gives that any observed equivalence or difference in outcomes between treatments is a function of the treatments rather than confounders such as unbalanced patient characteristics or referral pathways (Wessely, 2007). In the absence of randomization, one has to work very hard to demonstrate that unbalanced patient characteristics or referral practices could not have substantially influenced the treatment outcome comparison. We do not think the authors have gone far enough here. It is good to see that patients in the three treatments had similar initial severity scores. However, as already mentioned, the recent onset *versus* chronic distinction is a major determinant of natural recovery rates so one would want to know the duration distributions in the different treatments and probably use this information in the analyses. Furthermore, there are significant differences in patient characteristics between the treatments. Compared to CBT, patients in PCT and PDT are more likely to have interpersonal problems and bereavement/loss, suggesting that the services are directing different types of patient to the different treatments. This is not surprising, and perhaps desirable from the clinical perspective, but if true it is fatal

for the outcome comparison. The authors point out that the treatments do not differ in the proportion of patients who were categorized as having anxiety or depression by their therapist. However, we know that most patients presenting with mental health problems in primary care have some anxiety or depression. That does not mean that the proportion of patients whose primary problem is an anxiety or depressive disorder is similar in the three treatments. Finally, patients who received PDT had significantly more treatment sessions, a finding that would be expected if the treatment were less effective and therefore needed to be given in a higher dose.

#### **No evidence that the treatments were appropriately delivered**

No information is available on whether the therapists who delivered the various types of treatment had received appropriate training in the treatments. In addition, it is not known whether, within a treatment modality, the procedures that therapists used were appropriate for the problem being treated. These are serious omissions because some therapists are likely to 'label' their treatment as falling within a particular approach even if they do not follow the indicated, evidence-based procedures for treating the patient's problems within that approach. Such therapists may have essentially offered a placebo intervention in which non-specific factors (genuineness, warmth and empathy) were the main ingredients. Alternatively, they may have used procedures that are without a theoretical or empirical basis. For example, a patient with social phobia who was recently referred to one of the author's clinic reported having previously received eight sessions of 'CBT' in a counselling service without any noticeable benefit. On enquiry, the 'CBT' involved teaching the patient 'coping techniques' such as going into a toilet cubicle before a public-speaking event and pushing his arms hard against the walls to get 'psyched up'. The patient was pleased to find that when he subsequently had a course of treatment involving evidence-based CBT procedures he made a complete recovery. In our experience, such misunderstandings of what CBT comprises are by no means unusual.

RCTs of psychotherapy normally devote considerable attention to ensuring that the treatments are delivered appropriately. It is common for investigators to: (a) provide detailed therapist manuals covering the indicated procedures, (b) ensure that therapists receive appropriate training in the procedures, and (c) use audio- or videotapes of the sessions to check on treatment fidelity and competence. Although many of these steps would have been beyond the scope of what

is feasible in research into routine clinical practice, one would have expected the Stiles *et al.* (2006, 2007) studies to include some sort of assessment of the procedures that the therapists actually used.

The authors acknowledge that the quality of the treatments delivered may have been poor but they argue that dilution, if it occurred, would apply equally to the three treatments. We do not think this assumption is sound. PCT is fairly easy to deliver within the constraints of primary care but CBT is not. CBT research has shown that various different 'CBT' procedures vary considerably in effectiveness. For example, in the treatment of anxiety disorders it is known that cognitive restructuring (which is well suited to a primary-care setting) is unlikely to be helpful. It needs to be combined with *in vivo* exposure or behavioural experiments. Practitioners in primary care often report that place or time constraints make within-session exposure/behavioural experiments impossible, especially if they require the therapist and patient to go out into the real world.

Above, we have argued that the Stiles *et al.* (2006, 2007) studies do not provide good evidence that CBT, PCT and PDT are of equivalent effectiveness when given to patients with equivalent problems. The authors argue that the equivalence or 'Dodo bird verdict' conclusion is widely supported by psychotherapy research. It is true that some people read the comparative outcome literature in a way that is consistent with the Dodo bird verdict but it is equally clear that many other people do not. The doubters frequently point to the fact that there are a large number of RCTs that have shown that a particular treatment is more effective than another equally credible treatment when given to patients with the same problem by similar highly trained therapists (for some illustrative examples see Fairburn *et al.* 1993; Clark *et al.* 1994, 1998, 2006; Arntz & Van den Hout, 1996; Deale *et al.* 1997; Agras *et al.* 2000; Stangier *et al.* 2007). Reviews that support the Dodo verdict average over such a wide range of treatment procedures and client problems that they manage lose these effects and/or dismiss them with *post-hoc* arguments such as a presumed therapeutic allegiance. This point is nicely illustrated by Siev & Chambless' (in press) meta-analysis of RCTs that had compared two equally credible treatments (cognitive therapy and relaxation training) in panic disorder and in generalized anxiety disorder. The trials were conducted by researchers with a full range of imputed 'allegiance'. In panic disorder, there was clear-cut evidence that cognitive therapy is superior to relaxation training but in generalized anxiety disorder the two were equally effective. Hopefully, future research will

take this more nuanced approach by asking the question, 'What treatments work for whom in what settings?'

### Where to now?

We started this commentary by praising Stiles and colleagues for their tireless efforts to persuade clinicians to use standardized measures to monitor and report outcomes. Although we disagree with the conclusions they draw from their existing database, we fully agree with their direction of travel. The public has a right to know what outcomes they can expect from a particular treatment delivered by a particular service. In addition, psychotherapy research in general would greatly benefit from the advent of routine outcome monitoring. The current dominance of CBT in NICE Guidelines (2004*a-c*, 2005*a,b*) is bound to be partly a result of the fact that CBT-oriented clinicians have so far been more inclined to monitor outcomes and to conduct RCTs. With the widespread use of routine outcome monitoring, it seems likely that promising pilot data suggesting the effectiveness of other, non-CBT interventions will emerge and will be of sufficient interest to attract the funding and research expertise that are required to rigorously test the new interventions and establish their efficacy, or lack thereof.

However, if routine outcome monitoring is to provide the public with information about how likely they are to recover in a particular service it needs to achieve much higher data completion rates than the 33% in Stiles *et al.* (2006) and the 38% in Stiles *et al.* (2007). This is possible. However, it may require clinicians to give simple outcome measures at every session. In this way, a clinical end-point is achieved even when patients drop-out of therapy or terminate at an unexpected time. Using this system, Gillespie *et al.* (2002) obtained pre- and post-treatment data on 86% of patients who were offered cognitive therapy for PTSD in a walk-in community treatment service that was set up in Omagh following the 1998 car bomb. A similar system was adopted by the three trauma services that treated victims of the July 2005 London bombs with a similarly high data-completeness rate.

Getting, and making publicly available, close to complete data on recovery rates will be an important step forward. However, it is important to remember that such data will always need careful interpretation. The first interpretive issue that arises is: does the observed improvement indicate that the treatment worked? If the problem that is being treated is one for which the no treatment trajectory is well-known, then 'benchmarking' against previously studied wait-list, treatment as usual, or other active treatment control



groups could well be informative. However, this will only be feasible if a benchmark for the same problem, of similar severity and similar duration exists. If such a benchmark is not available, there is no alternative to an initial period of randomization to the treatment of interest or to a control condition (such as a brief wait period/usual GP care followed by the treatment of interest). As psychological treatment services expand to target more recent-onset cases and/or new problems, this type of evaluation will become increasingly important if service providers are not to be misled into being (for example) excessively optimistic about the apparent effectiveness of brief interventions in mild populations with problems of short duration or excessively pessimistic about the apparently modest improvements obtained with some other populations that may have much lower natural recovery rates. The recent work of a clinical team in Omagh provides an illustration of the use of a simple RCT to evaluate a new routine practice clinical initiative. Heartened by the positive results obtained in an audit of cognitive therapy for PTSD with the victims of a recent car bomb, the Northern Ireland Office and local charities provided funding to extend the work to multiply traumatized victims of the 'troubles' over the preceding four decades. As it was not known whether the treatment would be effective for this new population, for the first year the service ran as a RCT with all patients with PTSD being randomized to immediate cognitive therapy or cognitive therapy after a 12-week wait. The results (Duffy *et al.* 2007) showed that the treatment is effective with this new population and, as a consequence, it continues to be made available. However, the RCT could have revealed a very different story, more in line with recent evaluations of hormone replacement therapy (HRT). For more than two decades, women going through the menopause were advised to take oestrogen to reduce aversive symptoms such as hot flashes and to prevent coronary heart disease and osteoporosis. This advice was based on uncontrolled evaluations in which women who agreed to take HRT were compared with women who did not, with the former seeming to do better. However, recent controlled trials in which women are randomized to HRT or placebo showed the opposite, leading to the widespread discontinuation of HRT. The most likely reason for this opposite pattern of results in the uncontrolled *versus* the controlled evaluations is that in the former, women who exercised better self-care of their health in general were more likely to take up the opportunity to have HRT and hence had better outcomes, despite the fact that HRT was overall harmful (see Barlow, 2004; Hollon, 2006). This effect could, of course, equally apply to patients with mental health problems who accept a newly

available treatment in primary care (or any other setting) *versus* those who do not.

### Declaration of Interest

None.

### References

- Agras WS, Walsh BT, Fairburn CG, Wilson GT, Kraemer HC (2000). A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. *Archives of General Psychiatry* 57, 459–466.
- Arntz A, Van den Hout M (1996). Psychological treatments of panic disorder without agoraphobia: cognitive therapy versus applied relaxation. *Behaviour Research and Therapy* 34, 113–121.
- Barlow DH (2004). Psychological treatments. *American Psychologist* 59, 869–878.
- Catalan J, Gath D, Edmonds G, Ennis J (1984). The effects of non-prescribing of anxiolytics in general practice: I. Controlled evaluation of psychiatric and social outcome. *British Journal of Psychiatry* 144, 593–602.
- Clark DM, Ehlers A, Hackmann A, McManus F, Fennell M, Grey N, Waddington L, Wild J (2006). Cognitive therapy and exposure plus applied relaxation in social phobia: a randomised controlled trial. *Journal of Consulting and Clinical Psychology* 74, 568–578.
- Clark DM, Salkovskis PM, Hackmann A, Middleton H, Anastasiades P, Gelder MG (1994). A comparison of cognitive therapy, applied relaxation and imipramine in the treatment of panic disorder. *British Journal of Psychiatry* 164, 759–769.
- Clark DM, Salkovskis PM, Hackmann A, Wells A, Fennell M, Ludgate J, Ahmad S, Richards HC, Gelder M (1998). Two psychological treatments for hypochondriasis: a randomised controlled trial. *British Journal of Psychiatry* 173, 218–225.
- Deale A, Chalder T, Marks I, Wessely S (1997). Cognitive behavior therapy for chronic fatigue syndrome: a randomized controlled trial. *American Journal of Psychiatry* 154, 408–414.
- Duffy M, Gillespie K, Clark DM (2007). Post-traumatic stress disorder in the context of terrorism and other civil conflict in Northern Ireland: randomised controlled trial. *British Medical Journal* 334, 1147–1150.
- Ehlers A, Clark DM, Hackmann A, McManus F, Fennell M, Herbert C, Mayou R (2003). A randomized controlled trial of cognitive therapy, a self-help booklet, and repeated assessments as early interventions for posttraumatic stress disorder. *Archives of General Psychiatry* 60, 1024–1032.
- Evans C, Mellor-Clark J, Margison F, Barkham M, Audin K, Connell J, McGrath C (2000). CORE: clinical outcomes in routine evaluation. *Journal of Mental Health* 9, 247–255.
- Fairburn CG, Jones R, Peveler RC, Hope RA, O'Connor M (1993). Psychotherapy and bulimia nervosa. Longer-term effects of interpersonal psychotherapy, behavior therapy,

- and cognitive behavior therapy. *Archives of General Psychiatry* 50, 419–428.
- Fletcher J, Lovell K, Bower P, Campbell M** (2005). Process and outcome of a non-guided self-help manual for anxiety and depression in primary care: a pilot study. *Behavioural and Cognitive Psychotherapy* 33, 319–331.
- Gilbody SM, House AO, Sheldon TA** (2002). Psychiatrists in the UK do not use outcomes measures. National survey. *British Journal of Psychiatry* 180, 101–103.
- Gillespie K, Duffy M, Hackmann A, Clark DM** (2002). Community based cognitive therapy in the treatment of post-traumatic stress disorder following the omagh bomb. *Behaviour Research and Therapy* 40, 345–357.
- Hollon SD** (2006). Randomised clinical trials. In *Evidence-based Practise in Mental Health: Debate and Dialogue on the Fundamental Questions* (ed. J. C. Norcross, L. E. Beutler and R. F. Levant), pp. 96–105. American Psychological Association: Washington, DC.
- Kendrick T, Simons L, Mynors-Wallis L, Gray A, Lathlean J, Pickering R, Harris S, Rivero-Arias O, Gerard K, Thompson C** (2006). Cost-effectiveness of referral for generic care or problem-solving treatment from community mental health nurses, compared with usual general practitioner care for common mental disorders: randomised controlled trial. *British Journal of Psychiatry* 189, 50–59.
- NICE** (2004a). Anxiety: Management of anxiety (panic disorder, with and without agoraphobia, and generalised anxiety disorder) in adults in primary, secondary and community care (clinical guidance 22). National Institute for Clinical Excellence: London, UK ([www.nice.org](http://www.nice.org)).
- NICE** (2004b). Depression: Management of depression in primary and secondary care (clinical guide 23). National Institute for Clinical Excellence: London, UK ([www.nice.org](http://www.nice.org)).
- NICE** (2004c). Eating disorders: core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders (clinical guideline 9). National Institute for Clinical Excellence: London, UK ([www.nice.org](http://www.nice.org)).
- NICE** (2005a). Obsessive-compulsive disorder: core interventions in the treatment of obsessive-compulsive disorder and body dysmorphic disorder (clinical guideline 31). National Institute for Clinical Excellence: London, UK ([www.nice.org](http://www.nice.org)).
- NICE** (2005b). 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, UK ([www.nice.org](http://www.nice.org)).
- Siev J, Chambless DL** (in press). Specificity of treatment effects: cognitive therapy and relaxation for generalized anxiety disorder and panic disorders. *Journal of Consulting & Clinical Psychology*.
- Stangier U, Heidenreich T, Schramm E, Berger M, Clark DM** (2007). Cognitive therapy versus interpersonal psychotherapy in the treatment of social phobia: long-term results from a randomized controlled trial. World Congress of Behavioral and Cognitive Therapies, Barcelona, July 2007.
- Stiles WB, Barkham M, Mellor-Clark J, Connell J** (2007). Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies in UK primary care routine practice: replication in a larger sample. *Psychological Medicine*. Published online: 10 September 2007. doi:10.1017/S0033291707001511.
- 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* 26, 555–556.
- 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.
- Wessely S** (2007). A defense of randomized controlled trials. *Biosocieties* 2, 115–127.