Primary Care Mental Health 2006;4:00–00

International research

What happens when people wait for therapy? Assessing the clinical significance of the changes observed over the waiting period for clients referred to a primary care psychology service

Charles Young
Cambridgeshire and Peterborough Mental Health Partnership NHS Trust, Psychological Treatment Services, Addenbrooke's Hospital, Cambridge, UK

ABSTRACT

A statistical analysis of CORE-OM scores obtained at the intake assessment and again immediately before the start of therapy for 230 people seen in a primary care psychology service – a service designed to offer brief cognitive behavioural therapy (CBT) for the treatment of mild-to-moderate depressive and anxiety disorders – reveals a significant improvement for the total score and the domain scores of symptoms, wellbeing, life functioning and risk. The differences reported in this study are larger than the test–retest differences reported for CORE-OM, and are not associated with the length of the wait. However, while other studies have also reported psychometric scores over the duration of the wait, this study considers the clinical significance of the changes observed. Individual change statistics indicate that 16.5% of the sample achieved reliable and clinically significant improvement, 12.2% showed reliable improvement only, 67.8% did not show any reliable change, while 3.5% reliably deteriorated before the start of therapy. As for those scoring on or above the clinical cut-off points at the assessment, the proportions were 25.2%, 14.6%, 57.6% and 2.6% respectively.

Keywords: CORE-OM, primary care psychology service, waiting list

Introduction

While there is a great deal of quantitative evidence indicating that the mean scores obtained before and after the waiting period for psychological therapy show an overall reduction in psychological symptoms, very little is known about the clinical relevance of this apparent improvement.[1,2] The debate for many of these researchers is not about the clinical significance of these changes, but whether the improvement is real or the result of a testing artefact.[1,3,4] The studies that have considered changes over the waiting period have typically employed large samples so that even small differences that may or may not be clinically relevant are often statistically significant.

Posternak and Miller conducted a meta-analysis of psychotherapy outcome studies using waiting-list controls to determine the extent of the spontaneous remission of depressive symptoms over the duration of the waiting period.[1,2] Their results indicate that the mean decrease in Beck Depression Inventory

scores and Hamilton Rating Scale scores were 15.7% and 11.9% respectively. An analysis of the seven studies reporting individual data show that 19.7% of subjects improved to a degree comparable to what would be regarded as a positive response in antidepressant trials. However, this was based on a small total sample of 76 depressed patients, and the situation for the diagnostically heterogeneous populations presenting in typical NHS primary care settings is less clear. Also, while there appear to be changes on the symptom domain, far less is known about changes on some of the other relevant domains, such as client wellbeing or life functioning.

What we do know about the apparent improvement over this waiting period is that it has been observed in samples waiting for therapy for depression, agoraphobia with or without panic disorder, panic disorder with agoraphobia, obsessive–compulsive disorder (OCD) and other diagnoses.[1,2] Also, this improvement does not appear to be affected by the duration between the testing sessions, and when multiple measurements are taken, seems be confined to a period between the first and second time the measures are administered, after which time the scores appear to stabilise.[1,3,4] The changes also appear to be limited to instruments that measure negative self-characteristics of the respondent, and are not recorded for positive mood measures.[1,3,4]

The aim of the present study is to consider the clinical significance of the changes in test scores observed over the duration of the waiting period for each of the domains of symptoms, wellbeing, life functioning and risk. To do this, changes in scores before and after the waiting period reported by a sample of clients offered brief therapy in a primary care psychological treatment service are analysed. The service offers brief cognitive therapy based at general practitioner (GP) surgeries to people with common psychological disorders of a mild to moderate severity.

The study is a naturalistic design and some clients may have sought other psychological or psychiatric treatments while waiting. Posternak and Miller's meta-analysis, however, reports that the two studies that specifically excluded subjects who sought treatment elsewhere showed a decrease in depression scores of more than double the rate found in the remaining studies that did not exclude these subjects, suggesting that these 'waiting list' effects are not simply the result of therapy obtained elsewhere.[2]

The advantage of a naturalistic design is that the emphasis is on external validity rather than internal validity, making the results applicable to real-world clinical settings.[5]

Method

The sample comprises 230 consecutive clients (68 male and 162 female clients) who completed the CORE-OM questionnaire at the time of the assessment and before the start of their therapy in a primary care psychology service. Each person referred to the department was assessed by one of three experienced psychologists who recorded the diagnosis (using DSM-IV criteria but without any formal assessment instruments) and other clinical and biographical details.

The mean age at the date of the assessment was 36.4 years (standard deviation (SD) = 11.2 years; range: 18–64 years). Of the 230 participants, 68 were diagnosed with depression (29.6%), 48 with panic disorder (20.9%), 9 with health anxiety (3.9%), 38 with generalised anxiety (16.5%), 14 with specific phobias (6.1%), 52 with mixed anxiety and depression (22.6%), and 1 with an unrecorded diagnosis. The average wait between the assessment and first therapy session at the time the data was collected was 29.4 days (SD = 27.8; range: 3–225 days).

The research sample is drawn from the 385 people offered therapy in the primary care service over the duration of the data-collection period; the difference of 155 being those people who did not complete both CORE-OM forms. A comparison between the CORE-OM scores at assessment between the groups that did and did not compete the second questionnaire reveals that there is no statistical difference between these scores (U = 13363.5, N_1 = 121, N_2 = 226, P = 0.728, two-tailed). Therefore, there is nothing to suggest that the sample of 230 does not adequately represent all of the people offered therapy in the service. A similar source of attrition bias worth considering is that caused by people who are offered therapy but fail to attend their first therapy session and are, therefore, discharged from the service. Over the duration of the data collection period, however, only one person failed to attend their first appointment (although others did terminate prematurely after attending their first therapy session).

The measure: CORE-OM

The CORE-OM measure was designed with the specific purpose of developing the evidence base for the various psychological therapies on offer in the UK.[6,7] The rationale for the development of CORE-OM is to offer a single measure that can be used by therapists representing the different theoretical models across different settings. Psychologists have, in the past, employed a plethora of measures

making comparisons between different studies and audits difficult, if not impossible. CORE-OM is the product of an attempt to address this problem and provide useful benchmarks for all psychotherapy services.

Indeed, the benchmarking data for primary and secondary care settings is available. Barkham *et al* provide the service intake data obtained from 32 primary care and 17 secondary care NHS settings involving 5733 and 1918 patients respectively (the same primary care data are also reported by Evans *et al*).[8,9]

The outcome measure comprises 34 items and provides a general measure of distress as well as addressing four domains relevant to psychotherapy: subjective wellbeing (4 items), problems/symptoms (12 items), life functioning (12 items), and risk/harm to self or others (6 items; 4 'risk to self' and 2 'risk to others').

Each item is scored on a five-point scale from 0 (not at all) to 4 (most or all of the time). Positively framed items, eight of the 34 items, are scored in reverse. A mean score is calculated for the total score and each of the domain scores, which means that the measure can be 'pro-rated' if there are a limited number of incomplete responses and that the total and domain scores range from 0 to 4 (although recent practice is to multiply these scores by ten so that meaningful differences between scores are whole numbers).[10] Normative data for the four domains and the total score are available for clinical and non-clinical populations in the UK for both sexes.[11]

The measure shows good reliability and convergent validity with other measures commonly used in psychiatric or psychological settings.[7,8,12] Also, the measure has small sex effects and, appropriately, shows large differences between clinical and non-clinical populations. Importantly, the measure is sensitive to change. To be user-friendly, CORE-OM is short, fitting onto two sides of an A4 page, and is freely available. In addition, it can be easily hand-scored or scanned by computer.

Measuring reliable and clinically significant improvement

The main outcome is the difference between pre- and post-test CORE-OM scores. Changes over time between the first measurement and the second measurement are analysed for the total score and each of the four domains. Jacobson and Truax report that there is a growing dissatisfaction with the way treatment efficacy has been traditionally evaluated: the main concern being that although the usual statistical comparisons between means tell us how likely it is that the supposed difference is simply the result of chance, they do not tell us much about the clinical significance of the difference.[13] The latter refers to the degree to which treatment effects are associated with meaningful difference to the lives of the clients.[14] Specifically, clinical relevance, as defined by Jacobson and Truax, is the extent to which an intervention moves an individual from the range of a dysfunctional population into the range of a functional population.[13]

Reliable change is a related index that is concerned with whether the size of the change recorded for any individual is sufficiently large to be regarded as statistically reliable and not the result of measurement error. To be more precise, it is the size of the difference that would occur in only 5% of cases if the change were the result of measurement error.[12]

Therefore, change is referred to as *reliable and clinically significant improvement* (RCSI) if the individual's score moves from one that is characteristic of a dysfunctional population to one characteristic of a functional population and the magnitude of the difference between the scores is large enough to be regarded as reliable. If, however, the difference between the two scores is large enough to be regarded as reliable but the score does not move from a dysfunctional population to a functional population, then this would be regarded as *reliable change only*. For the CORE-OM questionnaire, change is reliable and clinically significant when the mean score moves from above to below the clinical cut-off (1.19 for males; 1.29 for females), and is greater than 0.48.[11]

Including clients who are below the clinical threshold at the assessment in the analysis obviously depresses the proportion of people showing reliable and clinically significant improvement.[10] This is because there will be a proportion of people who cannot show a change from the clinical to non-clinical scores, since their scores are already located in the non-clinical range to start with. In order to permit comparisons with studies that do and studies that do not include clients who report sub-threshold scores, the proportions obtaining RCSI will be calculated with and without those scoring below the cut-off points at the assessment.

Other statistical analyses

Although not the primary focus of this research, the statistical significance of all the changes to the mean scores are reported, as are the within-group effect sizes. Using Cohen's rule-of-thumb definitions, 0.20 is a small effect, 0.50 a medium effect and 0.80 a large effect.[15]

Results

The descriptive statistics reported in Table 1 show that the total and domain CORE-OM scores were lower at the start of therapy than they were at the time of the assessment.

Table 2 reports the test statistics and significance for each of the differences in scores between the assessment and start of therapy for each of the CORE-OM dimensions. The tests are one-tailed because any change is expected to be in the direction of improvement. The results indicate that there are statistically significant differences for each of the CORE-OM dimensions.

It is also worth reporting that there are either very small or no linear relationships between CORE-OM scores and the length of wait, and none of these correlations is significant, which is consistent with other research findings.

What is more relevant, given the objectives of this study, is the degree to which these changes are clinically significant. The proportions of each of the different change categories are reported in Table 3.

Table 3 reveals that 28.7% of the sample showed at least reliable improvement, while only 3.5% of the total sample show reliable deterioration. Thus, the analysis of individual scores reveals that most people remain stable or improve (96.5%), while only a tiny proportion deteriorates.

The proportion of clients above or equal to the clinical cut-off threshold of 1.19 for men and 1.29 for women is revealing: only 65.7% of the sample meet the clinical cut-offs at the assessment (79 below and 151 equal or above). If the proportion of people who do not meet the clinical cut-off are subtracted from the sample, then the percentage of people achieving reliable and clinically significant improvement is 25.2%. Therefore, one-quarter of the clinical population awaiting therapy in a NHS

Table 1 Mean, standard deviation (SD) and effect sizes

	CORE-OM						
	Assessment		First session		Difference		Effect sizes
	Mean	SD	Mean	SD	Mean	SD	
All items	1.48	0.62	1.17	0.63	0.30	0.46	0.50
Wellbeing	1.90	0.87	1.53	0.89	0.37	0.82	0.42
Problems	1.95	0.77	1.50	0.79	0.44	0.61	0.58
Life functioning	1.48	0.73	1.24	0.74	0.23	0.57	0.33
Risk	0.26	0.41	0.15	0.34	0.11	0.34	0.29

Table 2 Significance of differences before and after the wait

	The Wilcoxon signed ranks test difference scores					
	Median	Range minimun	Range maximum	*z*	*N*-ties	Significance (one-tailed)
All items	0.25	–0.89	1.95	8.849	224	$P<0.001$
Wellbeing	0.25	–1.50	3.25	6.194	199	$P<0.001$
Problems	0.41	–1.00	2.59	9.367	221	$P<0.001$
Life functioning	0.17	–1.42	2.00	5.609	215	$P<0.001$
Risk	0.00	–1.17	225	5.349	105	$P<0.001$

Table 3 Reliable and clinically significant change

	n (%)
Reliable and clinically significant improvement (RCSI)	38 (16.5)
Reliable improvement only	28 (12.2)
No reliable change	156 (67.8)
Reliable deterioration	8 (3.5)
Total	230 (100)

Table 4 Adjusted reliable and clinically significant change

	n (%)
Reliable and clinically significant improvement (RCSI)	38 (25.2)
Reliable improvement only	22 (14.6)
No reliable change	87 (57.6)
Reliable deterioration	4 (2.6)
Total	151 (100)

primary care service may obtain reliable and clinically significant improvement before therapy starts. Table 4 provides the proportion for each of the change categories for the sub-sample that equal or exceed the clinical cut-off.

The mean and standard deviation of this *adjusted* sample are 1.87 and 0.40 respectively.

Finally, it should not be forgotten that although a regression to the mean effect cannot explain the overall change in mean scores, it might inflate the categories of those showing improvement and those showing deterioration. These results, therefore, should be treated with some caution.

Discussion

The results reported suggest that there is a general improvement as measured by CORE-OM over the duration of the wait for cognitive therapy for people referred to the primary care service with mild-to-moderate anxiety or depressive disorders. The changes for each of the domain scores of symptoms, well-being, life functioning and risk and the overall score are all statistically significant. The differences reported in this study are larger than the test–retest findings reported by Evans *et al*.[12] In their non-clinical sample of 43 students who returned complete data, the difference in mean scores was 0.10 for all of the items, 0.06 for wellbeing, 0.14 for problems/symptoms, 0.11 for functioning and 0.02 for risk. Any comparisons, however, should be treated with caution: the sample employed by Evans *et al* was both small and drawn from a non-clinical population that was assessed on two occasions only, one week apart.

The mean scores at assessment are lower than those reported for other primary care settings by Barkham *et al* and Evans *et al*, suggesting that the clients in the present sample are, on average, less severe than the clients presenting in other NHS primary care services.[8,9] In the data reported by Barkham *et al*, the mean scores at assessment across 32 primary care settings are 1.81 for all of the items, 2.38 for wellbeing, 2.30 for symptoms, 1.80 for functioning, and 0.47 for risk.[8] Moreover, only 65.7% of the present sample meet the clinical cut-offs at the assessment (79 below and 151 equal or above), lower than the 78.6% reported by Barkham *et al*.[8] However, the data of Barkham *et al* are drawn from a number of services with a wide range of admission or acceptance criteria, while the present data are drawn from a service with that only accept referrals of common, mild-to-moderate depression and anxiety disorders. It is noteworthy that the mean scores for all items on the CORE-OM are identical for their primary and secondary care data, when one would expect the secondary care means to be higher then the primary care means. Still, one weakness of the present study is that one cannot be certain to what extent the results can be generalised to other primary care settings.

The primary objective of this study, though, is to consider the clinical relevance of any change observed. The individual change statistics suggest that only a very small minority of people show reliable deterioration, while a sizeable minority show either reliable or reliable and clinically significant improvement. This suggests that there is no apparent harm in waiting for therapy, apart from having to endure a reasonably stable state of distress for longer than would be the case if people were offered therapy immediately (although the relationship between waiting time and outcome of psychological therapy has received very little attention). It would be useful to be able to identify those likely to deteriorate over the course of their wait at the time of their assessment appointments, an issue that certainly warrants further study.

When clients who do not meet the clinical cut-off points are deducted from the analysis, then the proportion of people showing reliable and clinically significant improvement is slightly more than one-quarter, while almost 40% of this sample show either RCSI or reliable improvement only. Thus a large proportion of people referred to a primary care psychology service are either partially or fully recovered by the time therapy starts.

The 25.2% showing RCSI in the adjusted sample is greater than the 19.7% of the Posternak and Miller sample who show an improvement (defined as a change of psychometric scores that would regarded as a positive response in antidepressant trials) over a similar waiting period.[2] However, direct comparisons are problematic: The clinical cut-off points employed by the two studies are not necessarily equivalent, and neither are the inclusion criteria (for this particular comparison the present sample included people with common psychological disorders scoring above the CORE-OM clinical threshold, while Posternak and Miller set their threshold at 20 or above on the Beck Depression Inventory or the Hamilton Rating Scale, or a formal diagnosis of major depressive disorder). Other obvious difference between the two samples is that the one is a diagnostically homogenous sample while the other is diagnostically heterogeneous.

Of the domain scores, problems/symptoms show the largest effects, followed by wellbeing, life functioning and risk respectively. Therefore, although the effects seen over the duration of the waiting period are not confined to the symptom domain, it is the domain showing the largest effect. According to Howard *et al*, the changes achieved in psychological therapy follow a sequence of *remoralisation*, the recovery of wellbeing; *remediation*, the reduction of symptoms; and *rehabilitation*, the restoration of life functioning.[16] Therefore, at the start of therapy, one might expect to see the size of the effects ranked in decreasing order to follow this sequence rather then the actual sequence observed.

The data, therefore, do not appear to fit the phase model of psychotherapy change. However, since this phase of change involves the instillation of hope, it is possible that much *remoralisation* has occurred between the time clients discuss their problems with their general practitioners and arrive for their assessment appointment. In this case, the improvements in wellbeing are not associated with the largest effects, precisely because much of that improvement has already occurred before the first measurement. Thus, although these data do not appear to support the phase model of psychotherapy change, one may need to track change from the moment the client seeks help through to the completion of therapy and beyond to adequately address this and similar issues.

Conclusion

A substantial proportion of people meeting the clinical thresholds on CORE-OM who are required to wait for cognitive therapy in a primary care psychology service for common anxiety and depressive disorders achieve partial or full remission at the start of their therapy. The majority, however, show neither reliably improved nor deteriorated scores, while only a very small proportion seem to worsen while waiting.

REFERENCES

1 Arrindell WA. Changes in waiting-list patients over time: data on some commonly used measures. *Beware! Behaviour Research and Therapy* 2001;39:1227–47.

2 Posternak MA and Miller I. Untreated short-term course of major depression: a meta-analysis of outcomes from studies using wait-list control groups. *Journal of Affective Disorders* 2001;66:139–46.

3 Jorm AF, Duncan-Jones P and Scott R. An analysis of the re-test artefact in longitudinal studies of psychiatric symptoms and personality. *Psychological Medicine* 1989;19:487–93.

4 Patrick Sharp J and Gilbert DG. Effects of repeated administration of the Beck depression inventory and other measures of negative mood states. *Personality and Individual Differences* 1998;24:457–63.

5 Nathan PE, Stuart SP and Dolan SL. Research on psychotherapy efficacy and effectiveness: between scylla and charybdis? In: Kazdin AE (ed). *Methodological Issues and Strategies in Clinical Research*. Washington: American Psychological Association, 2003, pp. 505–46.

6 Barkham M, Evans C, Margison F *et al*. The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health* 1998;7:21–35.

7 Evans C, Mellor-Clark J, Margison F *et al*. CORE: Clinical Outcomes in Routine Evaluation. *Journal of Mental Health* 2000;9:247–55.

8 Barkham M, Gilbert N, Connell J *et al*. 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* 2005;186:239–46.

9 Evans C, Connell J, Barkham M *et al*. Practice-based evidence: Benchmarking NHS primary care counselling services at national and local levels. *Clinical Psychology and Psychotherapy* 2003;10:374–88.

10 Barkham M, Connell J, Stiles W *et al*. Dose–effect relations and responsive regulation of treatment duration: the good enough level. *Journal of Consulting and Clinical Psychology* 2006;74:160–7.

11 CORE System Group. *CORE System User Manual*. Leeds: CORE System Group, 1988.

12 Evans C, Connell J, Barkham M *et al*. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry* 2002;180:51–60.
13 Jacobson NS and Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* 1991;59:12–19.
14 Kazdin AE. Clinical significance: measuring whether interventions make a difference. In: Kazdin AE (ed). *Methodological Issues and Strategies in Clinical Research*. Washington: American Psychological Association, 2003, pp. 691–710.
15 Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. Hillsdale NJ: Lawrence Erlbaum and Associates, 1988.
16 Howard KI, Lueger RJ, Maling MS *et al*. A phase model of psychotherapy: Causal mediation of outcome. *Journal of Consulting and Clinical Psychology* 1993;61:678–85.

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Charles Young, Cambridgeshire and Peterborough Mental Health Partnership NHS Trust, Psychological Treatment Services, Box 190 – S Block, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK. Tel: +44 (0)1223 217 939; email: charles.young@cambsmh.nhs.uk or youngcs2001@yahoo.com

Received ?????
Accepted ?????