Elsevier Editorial System(tm) for Journal of Affective Disorders
Manuscript Draft

Manuscript Number: JAD-D-10-00881R2

Title: Implementation of Psychological Therapies for Anxiety and Depression in Routine Practice: Two Year Prospective Cohort Study

Article Type: Research Paper

Keywords: Stepped care, implementation research, anxiety and depression, cohort study, reliable and clinically significant change

Corresponding Author: Professor David Richards,

Corresponding Author's Institution: University of Exeter

First Author: David A Richards, PhD, BSc, RN

Order of Authors: David A Richards, PhD, BSc, RN; Gunilla Borglin, PhD, MSc

Abstract: Introduction: Worldwide, health systems are improving access to empirically supported psychological therapies for anxiety and depression. Evaluations of this effort are limited by the cross sectional nature of studies, short implementation periods, poor data completeness rates and lack of clinically significant and reliable change metrics.

Objective: Assess the impact of implementing stepped care empirically supported psychological therapies by measuring the prospective outcomes of patients referred over a two year period to one Improving Access to Psychological Therapies service in the UK.

Method: We collected demographic, therapeutic and outcome data on depression (PHQ-9) and anxiety (GAD-7) from 7,859 consecutive patients for 24 months between1st July 2006 and 31st August 2008, following up these patients for a further one year.

Results: 4,183 patients (53%) received two or more treatment sessions. Uncontrolled effect size for depression was 1.07 (95% CI: 0.88 to 1.29) and for anxiety was 1.04 (0.88 to 1.23). 55.4% of treated patients met reliable improvement or reliable and clinically significant change criteria for depression, 54.7% for anxiety. Patients received a mean of 5.5 sessions over 3.5 hours, mainly low-intensity CBT and phone based case management. Attrition was high with 47% of referrals either not attending for an assessment or receiving an assessment only.

Conclusions: Recovery rates for patients receiving stepped care empirically supported treatments for anxiety and depression in routine practice are 40 to 46%. Only half of all patients referred go on to receive treatment. Further work is needed to improve routine engagement of patients with anxiety and depression.

Introduction/background

Despite more than fifty years of research into, and development of, effective behavioural and cognitive-behavioural treatments (CBT) for depression and anxiety, the availability of such treatments worldwide remains poor (Andrews & Tolkein II Team, 2006; McManus et al, 2009). Although the evidence in support of CBT as an empirically supported therapy (Chambless & Hollon, 1998) is strong, trained therapists are in short supply and the organisation of treatments in many countries remains at best *ad hoc*. Recently, established arguments for empirically supported psychological therapy have been strengthened by the increased prominence of the evidence based medicine movement in health care generally (Sackett et al, 1996). The concept that health care should be based on scientific evidence is now mainstream.

In many countries, practice is guided by clinical health care guidance, for example by the APA Task Force on Promotion and Dissemination of Psychological Procedures (1995) and the UK National Institute of Health and Clinical Excellence (NICE). NICE in particular has identified the important role that empirically supported psychological treatments, mainly CBT, should have in the treatment of depression (NICE, 2009) and anxiety disorders (NICE, 2005a; 2005b; 2007). Once guidance is issued, they highlight the inability of health care providers to deliver the treatments recommended in the guidelines. For example, in the UK, no more than 10% of people with anxiety or depression receive psychological treatments for their problems and only 5% of the total disorder prevalence had access to an evidence-based psychological treatment (McManus et al, 2009). Worldwide the economic burden of this untreated anxiety and depression to economies runs to hundreds of billions of dollars, estimated to be £19billion in the UK alone (Layard, 2006).

Although increased investment is one solution to the lack of treatment availability, other organisational strategies have been proposed. Foremost amongst these is stepped care (Haaga, 2000), a system of delivering and monitoring treatments so that the most effective yet least resource-intensive treatment is delivered to patients first (Davison, 2000). Stepped care is included in Australian and NICE guidelines (Andrews & Tolkein II Team, 2006; NICE, 2007; 2009) as the method by which treatments for depression and anxiety, including CBT, should be delivered.

In order to address the severe under-provision of treatments, the UK government has instigated a highly ambitious programme of Improving Access to Psychological Therapies (IAPT) in England by funding the implementation of NICE guidelines for people suffering from depression and anxiety disorders. The IAPT programme aims to address under-provision of these treatments by training 3,600 new psychological therapists between 2008-2011 to enable 900,000 more people to access treatment, with half of those engaging in treatment moving to recovery and 25,000 fewer on sick pay and benefits by 2010/11.

However, the successful implementation of results of randomised clinical trials into routine clinical practice is not a foregone conclusion (Bero et al, 1998; Glasgow, Lichtenstein & Marcus, 2003; NHS Centre for Reviews and Dissemination, 1999). Treatments developed in trials may not translate into situations of contextual heterogeneity (Medical Research Council, 2008). Add to this factor, the nuances of training and delivery of apparently standardised treatments by many hundreds or thousands of individual therapists, and one typically finds far greater variation in outcomes than the original trials results (Lambert, Hansen & Harmon, 2010). Indeed, although Stewart and Chambless (2009) report that for anxiety disorders, data from routine practice may closely approximate outcomes seen in trials, they call for more observational cohort studies of routine clinical populations, echoing the UK Medical Research Council's (MRC) position that the least well performed element of the research-practice cycle is implementation (MRC, 2000; 2008). Long-term surveillance is recommended in the form of uncontrolled, longitudinal observational cohort studies to measure the extent to which the effects of treatment evidence gathered in RCTs is effectively translated into routine practice.

Therefore, aside from the evidence marshalled by NICE for the effectiveness of psychological treatments for depression and anxiety, the IAPT has been underpinned by the results of two 'demonstration sites' which acted as pilot test beds during 2006/7. In one of these sites, psychological therapies were delivered using a stepped care organisational protocol (Richards & Suckling, 2008) whereby the majority of patients received a low-intensity form of CBT such as guided self-help (Gellatly et al, 2007; Hirai & Clum, 2006). This site treated around five times more patients than an alternative where many more patients were allocated to 'high-intensity' i.e. standard face to face, CBT (Clarke et al, 2009; Richards & Suckling, 2009), although the focus in the second site included more patients with those anxiety disorders for which there is no good evidence for low-intensity CBT, particularly PTSD. Nonetheless, these results showed that the routine implementation of stepped care

psychological therapies, primarily CBT, could deliver recovery rates of 55-56%, broadly in line with that predicted from RCTs of the constituent psychological treatments.

Although this data (Clarke et al, 2009; Richards & Suckling, 2009) were a significant advance on other evaluations of routine psychological therapies practice, which have managed to collect data on no more than 33-38% of clinical outcomes for all patients (Stiles et al, 2006; 2008), it was limited by the cross sectional nature of the study. A significant proportion of our patients were still 'in treatment' so that in the stepped care site pre-post outcome data were only available on 46% of the 2,795 patients who were assessed during the first year of operation. At that point it was, therefore, unknown as to how many of the patients still 'in the system' would complete treatment and meet criteria for recovery after they ceased contact. It might be the case that a large proportion of those patients completing contact within the first year were more likely to recover, thereby leading to inflated effect sizes. Further, although we were able to report effect sizes and recovery rates, we did not analyse the data using reliable and clinically significant change criteria (Jacobson & Truax, 1991; McMillan, Gilbody & Richards, 2010). These criteria may be a better representation of patient recovery than other less sophisticated methods which rely on patients falling below a cut point on a clinical outcome measure, no matter what their pre-treatment starting point. Finally, one might also argue that the intense scrutiny placed upon the site during the first year of operation might question the extent to which our previous data could be said to be 'routine'.

To remedy these limitations we conducted a prospective study of all patients entering the pilot site until the point that they exited the system. Our objective was to determine the clinical impact of providing evidence-based psychological treatment in a complete cohort of patients with anxiety and depression treated at the site. This remedies the major limitation in our previous cross-sectional analysis (Richards and Suckling, 2009) in which a large number of patients who had not completed their contacts with the service had unknown outcomes. Our prospective method adopts a procedure analogous to an intention to treat analysis in clinical trials, but for observational data, where rigorous efforts are made to ensure outcomes for all patients are collected. We set a census date three years after the site commenced operations, 12 months after referral data on all patients in our cohort had been logged at the site, to maximise the chances that all patients would have completed contact. We analysed outcome data conventionally as well as against the reliable and clinically significant change criteria described by Jacobson & Truax (1991), in which the threshold for clinically significant

change on any continuous measure of morbidity is set at the point where the probability of coming from a clinical and non-clinical distribution is equal, the recommended method of quantifying improvement when clinical and non-clinical distributions overlap (McMillan, Gilbody & Richards, 2010).. Any patient with a score above this cut point is regarded as having a clinically significant level of morbidity. To achieve reliable as well as clinically significant change, a patient must improve to below this cut point and improve to a greater extent than could be accounted for by the level of test-retest unreliability of the specific measure.

Method

Prospective, longitudinal observational study.

Sample

The sample consisted of 7,859 consecutive patients referred during 24 months between 1st July 2006 and 31st August 2008 and completing treatment by 31st August 2009.

Setting

The study was conducted in a post-industrial city of 290,000 inhabitants in the north of England. The population was overwhelmingly white British (97%) and although there are areas of affluence, 41% of the population were in the UK nationally defined most deprived quintile. Average population mortality from cancer, smoking, heart disease and stroke were all significantly above the national UK average.

Interventions

Clinical care was organised according to stepped care principles. Patients were assessed and offered a choice of treatments suitable for their condition. A case manager undertook a face-to-face assessment, following which patients were offered telephone delivered contacts for subsequent appointments, although face-to-face appointments were available. The principle adopted was that a low burden/low-intensity treatment should be discussed and offered to patients first. Patients were then monitored at each subsequent contact to determine whether they required stepping up to a higher intensity of treatment, a decision which was taken after assessing patient progress using clinical outcome measures and discussion with the patient and the case manager's supervisor. On very rare occasions high-intensity treatment could be offered as a first step where patients specifically requested it and after discussion with the

clinician's supervisor. Low intensity treatment included a CBT-based guided 'Recovery Programme' (Lovell & Richards, 2008) for depression and a guided self-help manual for anxiety (Williams, 2003), both including education about depression and/or anxiety and a choice of cognitive and behavioural strategies such as behavioural activation, cognitive restructuring, exposure, sleep and panic management and problem solving. Two computerized CBT programmes 'Beating the Blues' and 'Fear Fighter' were also offered to patients. Pharmacological treatment remained the responsibility of the patient's general practitioner at all times but patients also received advice on managing their medication if prescribed and if they requested it. Workers could also liaise with social care agencies for activities such as housing support and debt relief. High-intensity treatment was CBT delivered according to Beck's (1979) manual.

At any one time the service employed 25 low- or high-intensity therapists. During the course of the study 51 degree educated and/or experienced mental health workers delivered case management. They were specifically trained to deliver high-volume, low-intensity treatments via a custom designed 25-day training programme at the University of York – a forerunner of the national IAPT low-intensity clinical training curriculum (www.iapt.nhs.uk/). The course covered the engagement and assessment of patients with common mental health problems together with clinical skills to deliver evidence-based, low-intensity treatments and required participants to pass clinical skills competency assessments before seeing patients. Eleven mental health professionals with a post-graduate qualification in CBT delivered high-intensity CBT.

Measures

All data was collected by mental health workers as part of their routine clinical practice and entered onto a secure electronic management system, 'PC-MIS' (www.pc-mis.co.uk) designed for managing patient referrals, allocation and progression in the stepped care system and for facilitating supervision of low-intensity case managers by more senior workers. Clinical measures were recorded at every contact between workers and patients.

Demographic data, collected at intake, included contact details, date of birth, gender, ethnicity, referral route, and primary and secondary problems identified by referrers. Workers also recorded treatment type given, delivery method (face-to-face or telephone), session duration and patient disposition at the end of each contact, including date of treatment ending.

Depression: nine-item Patient Health Questionnaire (PHQ-9) (Kroenke, Spitzer & Williams, 2001), range 0-27. The PHQ-9 has been validated in a UK depressed population (Cameron et al., 2008; Gilbody et al., 2007), has good psychometric properties – Cronbach's (1951) alpha 0.89 (Kroenke, Spitzer & Williams, 2001), with a score of 10 being optimum for identifying depressive symptoms likely to be of clinical severity (Gilbody, Richards & Barkham, 2007; Kroenke, Spitzer & Williams, 2001).

Anxiety: seven-item General Anxiety Disorder Questionnaire (GAD-7) (Spitzer, Kroenke, Williams & Lowe, 2006), range 0-21. The GAD-7 has not been validated in a UK population but has good psychometric properties from studies in the US – Cronbach's (1951) alpha 0.92 with a score of 8 being optimum for identifying symptoms of general anxiety disorder, panic disorder, post-traumatic stress disorder or social anxiety disorder (Kroenke, Spitzer, Williams, Monahan & Lowe, 2007).

Data download and cleaning

Data was downloaded in an anonymised, non-identifiable manner from PC-MIS in the form of a series of .csv files which were converted into MSExcel and then analysed with SPSSv14. Downloads were obtained in the form of case-by-case datasets and contact-by-contact datasets. As this was a routine dataset, with clinical contact data entered by 63 mental health workers it required cleaning. Case-by-case data items were examined by GB and cross checked by unique patient identifiers against contact-by-contact data and then cleaned of inconsistencies, errors and missing data. DR validated all data cleaning. For example, where no end of episode date was recorded and no clinical contact had been entered in the last six months, the case was coded as 'closed' and the last appointment date taken as the end of the treatment episode. Further, PC-MIS did not differentiate between episodes of care where patients had been discharged and then returned for another treatment episode. We identified these cases (n=81) and included data from first episodes only in our analyses.

Statistical analysis

We divided patients into those referred but not assessed, those receiving one assessment contact only and those 'treated', i.e. with two or more contacts including the assessment and at least one other treatment contact. Workers recorded the final disposition of the treated group, using pre-selected values on the electronic patient management system, into those that

jointly agreed with workers to complete treatment, those that dropped out from treatment without agreement with workers and those that the workers subsequently found unsuitable during treatment (figure 1). We calculated means, standard deviations, frequencies and percentages as appropriate for data on patient characteristics and treatment received. We calculated pre- and post-treatment change based on several common methods used to define treatment outcome: 1) rates of likely depression and anxiety pre- and post-treatment by PHQ-9 and GAD-7 cut-offs; 2) pre/post effect sizes (initial assessment score minus post-treatment score divided by the post-treatment standard deviation – in our case the largest and hence most conservative of the pre, pooled or post-treatment standard deviations); 3) relative risks (event rate for likely anxiety or depression post-treatment divided by the event rate pre-treatment); 4) 'IAPT Recovery Rates' where recovery = (pre-treatment PHQ-9 > 9 or pre-treatment GAD-7 > 7) and (post-treatment PHQ-9 < 10 and post-treatment GAD-7 < 8) capturing the proportion of patients who were unlikely to be cases of depression and/or anxiety post-treatment of those who were likely cases of depression and/or anxiety pre-treatment; 5) we also calculated reliable and clinically significant change (Jacobson & Truax, 1991). Clinically significant change is defined as the score at which the probability of coming from a clinical and non-clinical distribution is equal. Scores below this point are classified as the non-clinical range. Clinically significant change requires that a person is above the cut-off pre-treatment (i.e. is in the clinical range) but below it at post-treatment. Reliable change is where the change in scores must be greater than that which could be due to the inherent unreliability of the measure. We calculated these criteria using estimates from clinical and non-clinical population distributions (means and standard deviations) for the PHQ-9 from the clinical pre-treatment scores of the current study and the non-clinical distribution and estimate of internal reliability from the original PHQ-9 validation study (Kroenke, Spitzer, & Williams, 2001). On the basis of these, reliable improvement required an improvement of 6 points or more and clinically significant change required a person to score 9 or above pre-treatment and 8 or below post-treatment. The same approach was used to calculate these criteria for the GAD-7. Clinical means and standard deviations were taken from the current dataset, and non-clinical distribution values and estimates of internal reliability were taken from the original validation study of the GAD7 (Spitzer, Kroenke, Williams, & Lowe, 2006). For the GAD-7 the corresponding figures were: a pre- to post-treatment improvement of 5 points or more (reliable improvement) a move from 10 or above pre-treatment to 9 or below post-treatment (clinically significant change).

In addition, a sensitivity analysis was conducted to test whether recent onset cases were less severe and more likely to improve as suggested by Andrews (2001) and Kendler, Walters and Kessler (1997). Pre and post scores were calculated separately for patients whose problem durations were 0-2, 3-4, 5-6, 7-12, 13-24 and >24 months.

Ethical approval

This study did not fall under the UK National Patient Safety and National Research Ethics Service (NRES) definition of research, (National Patient Safety Agency, 2010) as we did not manipulate clinicians' treatment decisions nor used experimental interventions in this project. Therefore, permission to access the routine, anonymised, non-patient identifiable data required for this study was granted by the NHS service provider under its data sharing agreement, approved by the Trust' Caldicott Guardian according to national NHS policy. As part of routine clinical practice in the NHS, all patients are informed that their data may be used in this manner but that all data will be anonymised and they will not be identified by those analysing the data or in any publications arising from it. Data was stripped of all sensitive patient identifiers (name, address, post-code, NHS number) before being downloaded.

Results

Patient flows: 7859 patients were referred to the service during the census period. Of these, 5717 (73%) had a recorded assessment in the 12 months following referral. As a percentage of these 5717 patients assessed, 1534 (27%) received an assessment and/or advice appointment only, with the remaining 4183 (73%) receiving more than one contact. 2949 (71%) of patients receiving more than one contact completed an agreed programme of treatment, 969 (23%) dropped out of such treatment, 262 (6%) were considered no longer suitable and three patients died from causes unrelated to their psychological difficulties (figure 1).

Insert Figure 1 about here

Demographic data: the mean age of patients referred was 38 years, almost two thirds of whom were women (table 1). Three quarters of patients were identified by the referrer as depressed with most of the remainder identified as having mixed depression and anxiety or

generalised anxiety disorder. Very few patients were identified by referrers as having any of the other anxiety disorders. Of the 5717 patients who had a recorded assessment only 52% (n=2975) had a recorded date of onset (mean problem duration in months = 32, S.D. 60.95) and of the 4183 patients with more than one contact 55% (n=2297) had this data recorded (mean problem duration = 31.77, S.D. 58.55).

Insert Table 1 about here

Clinical outcomes: tables 2 and 3 report outcome data for all patients receiving two or more contacts with the service. Effect sizes were large for both anxiety and depression, the largest effects being seen in those patients completing treatment. At assessment 91% of patients were above the cut-of points on either the PHQ-9 or GAD-7 which indicate they meet the criteria for a diagnosis, compared to 54% at the last contact. Using reliable and clinically significant change criteria, 40% of anxious and 41% of depressed patients were reliably and clinically significantly improved with a further 15% of both groups reliably improved, leaving 45% of patients scoring above GAD-7 or PHQ-9 criteria as unimproved. Problem duration had no effect on pre-treatment scores but a small effect on improvement rates, accounting for a mere 0.7% of the variance in improvement for both depression ($F=2.746$, $p=0.012$) and anxiety ($F=2.602$, $p=0.016$). Post-hoc analyses (Tukey's-b) indicated that this effect was caused by those patients with chronic problem durations of greater than two years, whose post-treatment scores remained higher than the other groups.

Insert Table 2 about here

Insert Table 3 about here

Interventions received: the vast majority of patients who received treatment following an assessment did so using written guided self-help for depression (n=3311; 79.2%) and/or anxiety (n=2021; 48.3%) or used computerised CBT (n=491; 11.7%). Medication management advice was given to 1459 (34.9%) of patients with 1279 (30.6%) receiving liaison advice with other services. Eighty one (1.9%) patients were treated with counselling. On average patients were treated in relatively few sessions (five) with a short combined contact time (three hours), reflecting the essentially low-intensity nature of the stepped care system in place (table 4). A total of 275 patients (7% of all patients receiving treatment)

received high-intensity CBT, 251 of whom did so after being stepped up from unsuccessful low-intensity treatment. These 275 patients comprised 6% (185/3311) of patients who had previously received low-intensity depression treatment, 9% (178/2021) low-intensity anxiety treatment, 9% (84/919) who had received information only, 8% (109/1459) of patients who had received medication advice and 8% (37/491) who had received cCBT (percentages not cumulative as patients could receive more than one low-intensity intervention). Outcomes for these patients were then similar to those receiving low -intensity treatments only, with those that completed high-intensity treatment doing so in an average of ten hours in a median number of sessions of 13 (table 5).

Insert Table 4 about here

Insert Table 5 about here

Discussion

We have shown that large scale implementation of evidence-based psychological therapies into routine service settings are associated with pre/post effect sizes >1.0 for anxiety and depression with reliable or reliable and clinically significant change in 55% of cases. These results were achieved by a stepped care system where the majority of cases were treated in less than six sessions of contact with mental health workers. This study is unique in that our procedures allow us to report an almost complete account of all patients' outcomes who received treatment, the nearest comparable analysis to an intention to treat analysis in a clinical trial. This compares to only 62% in our previous cross-sectional analyses (Richards & Suckling, 2009) and 33 and 38% in previous reports of routine implementation (Stiles et al, 2006; 2008).

We can report, therefore, the outcomes of the IAPT demonstration site with far greater confidence than previously. Our uncontrolled effect sizes were similar to those observed in the more limited cross sectional analyses for depression (1.07 compared to 1.09) and anxiety (1.04 compared to 1.07) although this complete longitudinal sample's post-treatment morbidity rates were around 5% more than those seen in the cross sectional sample in both depression (41% versus 35%) and anxiety (45% versus 40%). Furthermore, there was no evidence that patients with recent onset of symptoms improved more than others, indeed it was only patients with very chronic disorders of greater than two years duration that showed improvement rates worse than the other duration groups, thus providing no support to the

contention that these rates of recovery were unduly associated with greater levels of spontaneous remission seen in new onset cases of depression.

Using the 'IAPT formula' to calculate recovery (patients who are above case threshold on *either* GAD-7 *or* PHQ-9 at assessment and who are then below threshold on *both* measures post-treatment) 46% of patients were 'recovered'. Using more conventional reliable and clinically significant change criteria, between 40-41% of patients experienced reliable and clinically significant improvement on either depression or anxiety, with a further 15% experiencing an improvement that is unlikely to be due to the unreliability of the measure. Outcomes for those patients who ended treatment in an agreed and planned manner (2,949 – 71% of those assessed and receiving two or more contacts) were between 6-9% greater than those who dropped out or who were considered no longer suitable for treatment, depending on the specific analysis used.

The majority of patients were treated using only low-intensity psychological therapy, predominantly low-intensity CBT, in an average of 5.5 sessions (6 sessions for completers) taking up to 3.5 hours on average. Telephone contacts occupied more than half of the time delivering treatment.

Although remaining a small percentage of the total treated population, the proportion of patients receiving high-intensity CBT in this full longitudinal sample was almost three times (five times in completers) that of our previous cross sectional analysis, (table 4). Surprisingly, almost no patients received CBT without going through an initial low-intensity step – almost all high-intensity patients were stepped up from an ineffective course of low-intensity treatment. Patients who required a step up to high-intensity CBT treatment needed a median of more than three times more high-intensity appointments to achieve comparable treatment effect sizes to those receiving low-intensity treatment only (table 5).

Limitations

Our findings should be interpreted in light of the many limitations of conducting uncontrolled implementation studies in routine clinical settings. First and foremost, such uncontrolled observational designs cannot ascertain what proportion of the recovery rates could be accounted for by factors such as spontaneous remission over time. We tried to investigate the potential for higher spontaneous remission rates in those with shorter durations of illness and

we found no evidence that patients with very short durations of illness contribute disproportionately to the improvement rates reported. However, problem duration data was missing for 45% of patients and so conclusions here are tentative. Additionally, it is possible that effect sizes for depression may be inflated by the 'remoralisation' effect observed in patients with anxiety disorders, whose mood may initially lift as a consequence of feeling hopeful, even through their specific anxiety symptoms may not have yet ameliorated.

A further significant limitation is the lack of standardised diagnostic interviews. Many patients were not even given a 'probable' diagnosis by clinical staff (low-intensity staff in particular are not trained to formally diagnose patients). Although our self-report measures, the PHQ-9 and GAD-7, have good psychometric properties and the cut points we used are those reported in validation studies as being the scores with the best balance of diagnostic specificity and sensitivity, they do not formally 'diagnose'. Although it is likely that most of our patients would attract a formal diagnosis if interviewed by a diagnostician, we cannot say for sure what the exact diagnostic status of our population is.

Another potential bias arises from the collection of measures by staff as part of their everyday clinical work, rather than by independent raters. Some protection against observer bias is provided by the patient-completed self-report method of data collection but since many questionnaires were completed in the presence of workers, the potential for the demands of this situation to influence our results remains.

Implications

Notwithstanding the caveats above, which apply to most uncontrolled clinical implementation studies, the results of this observational prospective cohort study have considerable implications for the dissemination of evidence-based psychological treatment to routine practice. The contribution of our study is enhanced by the fact that clinical outcome measures are complete for more than 90% of all patients and final dispositions are known for the total population referred, assessed and treated. We know that patients were considerably distressed by their symptoms. Despite the lack of formal diagnoses in this service, 91% of patients met the PHQ-9 or GAD-7 sensitivity and specificity criteria associated with having a diagnosis. Consequently, we can make a number of reasonable assumptions about the extent to which low-intensity CBT provided as a part of a stepped care system can be effective. Once patients engage in and receive treatment, we could expect between 40-45% of patients treated in such

a system to be below the criteria for having a diagnosis on all measures after a short period – six sessions over 3.5 hours – of low-intensity CBT delivered predominantly using the telephone. More than half of patients treated could demonstrate either clinical and reliable, or reliable change. When patients are stepped up to high-intensity treatment, similar proportions of patients might recover where low-intensity treatment has failed. However, we are able to say less about the performance of high-intensity CBT..The number of high-intensity treatments delivered to patients in our sample is low, albeit higher than in our previous study. Our dataset does not enable us to determine whether this is due to patient choice, poor management of non-response to low-intensity treatment, lack of available high-intensity resource – a key factor noted in other observational studies of stepped care (Richards et al, 2010) – or other unknown clinical or operational factors.

Of those that complete treatment, 40% are likely to show no improvement, some even after a trial of high-intensity CBT. High-intensity CBT, interpersonal therapy, brief psychodynamic, behavioural couples therapy and counselling approaches are all recommended to some extent by NICE in the UK and by other guidelines internationally. The extent to which at least some of these 40% would be helped by the provision of more of these high-intensity empirically supported treatments cannot be determined by this study. Further research is needed to investigate different configurations of low- and high-intensity treatments to identify the optimum proportions required to maximise recovery rates.

Although we have demonstrated the feasibility of managing high volumes of patients with depression and anxiety through a stepped care system, attrition rates are still higher than we would desire. Unlike many studies, we chose to present our data from as near to the beginning of the patient pathway as possible, not merely from the point of assessment and allocation to treatment within the service. This means that we have no data whatsoever on 2,142 patients who did not make it to an assessment following referral from a GP. Curiously, in our study, at every point in the patient flow 27% of patients either don't turn up, don't come back or drop out. Despite waiting a full year after patients had been referred to, and logged by, the service, only 4183/7859 (53%) of referred patients received two or more sessions of assessment and then treatment. Whilst some patients may have been satisfied with a single advice session, many more were lost to the service before and after assessment. This is despite the service adopting many of the principle features of collaborative care (Katon et al, 1999; 2010) such as telephone case management, designed to improve patient engagement and improve both

treatment compliance and outcomes (Gilbody et al, 2006). In routine practice it has previously been observed that we are still relatively poor at engaging and retaining our patients (Barrett et al, 2008; Gilbert et al, 2005; Wierzbicki & Pekarik, 1993). Disappointingly, even when adopting collaborative care procedures, routine services might expect to loose at least a quarter of their patients at each stage as they progress through the referral, assessment and treatment patient pathway. Systems of quality improvement clearly need further refinement to stem the haemorrhage of patients from routine care.

Conclusion

Recovery rates for patients entering a stepped care service utilising empirically supported treatments for anxiety and depression in routine practice are between 40 to 55%, depending on specific metric used. However, only half of all patients referred go on to receive treatment. Further work is needed to understand and improve engagement and utilisation for patients with anxiety and depression in routine services.

References

Andrews, G. (2001) Placebo response in depression; bane of research, bone to therapy. *British Journal of Psychiatry*, 178, 192-194.

Andrews, G. & Tolkein II Team (2006). *A needs-based, costed stepped-care model for Mental Health Services*. Sydney: CRUFAD, University of New South Wales.

Barrett, M.S. Chua, W-J. Crits-Christoph, P. Gibbons, M. Casiano, D. & Thompson, D (2008) Early withdrawal from mental health treatment: implications for psychotherapy practice. *Psychotherapy,* 45, 247–267

Beck, A.T. (1979). *Cognitive Therapy and the Emotional Disorders*, Meridian, New York.

Bero, LA., Grilli, R., Grimshaw, J., Harvey, E., Oxman, A.D., Thomson, M.A. (1998). Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *British Medical Journal, 317*, 465-8.

Cameron, I.M., Crawford, J.R., Lawton, K., Reid, I.C. (2008). Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *British Journal of General Practice, 58,* 32-36.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66,* 7–18.

Clark, D.M, Layard, R., Smithies, R., Richards, D.A., Suckling, R., Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy, 47,* 910-920.

Cronbach LJ. Coefficient alpha and the internal structures of tests. (1951). *Psychometrika, 3,* 297–334.

Davison, G. (2000). Stepped care: doing more with less? *Journal of Consulting and Clinical Psychology, 68,* 580-585

Gellatly, J., Bower, P., Hennessey, S., Richards, D., Gilbody, S., and Lovell, K. (2007). What makes self-help interventions effective in the management of depressive symptoms? Meta analysis and meta regression. *Psychological Medicine, 37* 1217–1228.

Gilbody, S., Bower, P., Fletcher, J., Richards, D. A., Sutton, A. J. (2006). Collaborative care for depression. A cumulative meta-analysis and review of long-term outcome. *Archives of Internal Medicine, 166,* 2314-2331.

Gilbody, S., Richards, D. A., & Barkham, M. (2007). Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. *The British Journal of General Practice, 57,* 650-652.

Gilbert, N., Barkham, M., Richards, A. and 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, 4, 241-251

Glasgow, R.E., Lichtenstein, E., Marcus, A.C. (2003). Why don't we see more translation of health promotion research into practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health*, 93(8), 1261-7.

Haaga, D.A.F (2000). Introduction to the special section on stepped care models in psychotherapy. *Journal of Consulting and Clinical Psychology*, 68, 4, 547-548

Hirai, M., & Clum, G.A. (2006). A meta-analytic study of self-help interventions for anxiety problems. *Behavior Therapy, 37,* 2, 99-111.

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

Katon, W., Von Korff, M., Lin, E., Simon, G., Walker, E., Unützer, J., Bush, T., Russo, J., & Ludman, E. (1999). Stepped collaborative care for primary care patients with persistent symptoms of depression: a randomized trial. *Archives of General Psychiatry 56,* 1109-1115.

Katon, W., Unutzer, J., Wells, K. and Jones, L. (2010). Collaborative depression care: history, evolution and ways to enhance dissemination and sustainability. General Hospital Psychiatry, doi:10.1016/j.genhosppsych.2010.04.001

Kendler, K.S., Walters, E.E., Kessler, R.C. (1997) The prediction of length of major depressive episodes: results from an epidemiological survey of female twins. *Psychological Medicine*, 27, 107-117.

Kroenke, K., Spitzer, R.L., Williams, J.B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16,* 606-613.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., Monahan, P.O. and Lowe, B. (2007). Anxiety Disorders in Primary Care: Prevalence, Impairment, Comorbidity, and Detection. *Annals of Internal Medicine*, 146, 317-325

Lambert, M.J., Hansen, N.B., Harmon, S.C. (2010). Outcome Questionnaire System (The OQ System): Development and Practical Applications in Healthcare Settings. In: Barkham, M., Hardy, G.E., Mellor-Clark, J. (eds). *Developing and Delivering Practice-Based Evidence*. Chichester: John Wiley & Sons, Ltd

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

Lovell, K., & Richards, D.A. (2008). *A recovery programme for depression*. London, Rethink.

McManus, S., Howard Meltzer, H., Brugha, T., Bebbington, P., & Jenkins, R. (2009). *Adult Psychiatric Morbidity in England, 2007: results of a household survey*. London: The Health & Social Care Information Centre

McMillan, D., Gilbody, S., & Richards, D.A. (2010). Defining successful treatment outcome in depression using the PHQ-9: A comparison of methods. *Journal of Affective Disorders*. 127, 122-129.

Medical Research Council (2000). A framework for the development and evaluation of RCTs for complex interventions to improve health. London: Medical Research Council.

Medical Research Council. (2008). *Developing and evaluating complex interventions: new guidance*. London: Medical Research Council.

National Patient Safety Agency, (2010). *Defining Research*. London: National Patient Safety Agency.

National Institute for Health and Clinical Excellence. (2005a). *Core interventions in the treatment of obsessive-compulsive disorder and body dysmorphic disorder*. London: NICE

National Institute for Health and Clinical Excellence. (2005b). *Post-traumatic stress disorder (PTSD): The management of PTSD in adults and children in primary and secondary care*. London: NICE

National Institute for Health and Clinical Excellence. (2007). *Anxiety (amended): management of Anxiety (Panic Disorder, with or without Agoraphobia, and Generalised Anxiety Disorder) in Adults in Primary, Secondary and Community Care*. London: NICE.

National Institute for Health and Clinical Excellence (2009). *Depression: Management of depression in primary and secondary care.* London: NICE.

NHS Centre for Reviews and Dissemination. (1999). Getting evidence into practice. *Effective Healthcare,* 5(1).

Richards, D.A., & Suckling, R. (2008). Improving access to psychological therapies: The Doncaster demonstration site organisational model. *Clinical Psychology Forum, 181,* 9-18.

Richards, D.A., & Suckling, R. (2009). Improving access to psychological therapies: Phase IV prospective cohort study. *British Journal of Clinical Psychology, 48,* 377–396

Richards, D.A., Weaver, A., Utley, M., Bower, P., Cape, J., Gallivan, S., *et al.* (2010). *Developing evidence based and acceptable stepped care systems in mental health care: an operational research project. Final report.* NIHR Service Delivery and Organisation programme.

Sacket, D.L., Rosenberg, W.M.C., Gray J.A.M., Haynes, R.B., Richardson W.S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal, 312,* 71-72.

Spitzer, R.L., Kroenke, K., Williams, J.B.W., & Lowe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine, 166,* 1092-1097.

Stewart, R. E., & Chambless, D. L. (2009). Cognitive-behavioural therapy for adult anxiety disorders in clinical practice: A meta-analysis of effectiveness studies. *Journal of Consulting and Clinical Psychology, 77,* 4, 595-606.

Stiles, W.B., 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, W.B., Barkham, M., Mellor-Clark, J., & Connell, J. (2008). Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies in UK primary care routine practice: replication in a larger sample. *Psychological Medicine, 38,* 677-688.

Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically validated psychological treatments. *The Clinical Psychologist, 48,* 3–23.

Williams, C. (2003). *Overcoming Anxiety: a Five Areas Approach*. London: Hodder Arnold.

Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice*, 24, 190-195.

Table(s)

Table 1. Patient characteristics

	One contact only (n = 1534)	**Two or more contacts** (n = 4183)	**Not assessed** n=2142	**All referred to services** (n = 7859)
Age (mean/SD)	38.3 (13.7)	39.1 (13.1)	35.2 (13.4)	37.9 (13.4)
Female/Male (%)	63.5/36.5	66.6/33.4	64.6/35.4	65.4/34.6
White British (%)	94.0	94.1	90.7	93.1
GP referral (%)	92.0	93.2	90.8	92.3
Primary problem identified by referrer (%/n)				
- Depression	76.0 (1166)	77.2 (3231)	69.7 (1494)	75.0 (5891)
- Mixed anxiety & depression	7.5 (115)	7.6 (319)	9.9 (211)	8.2 (645)
- General anxiety disorder	4.2 (64)	5.6 (235)	3.7 (80)	4.8 (379)
- Other anxiety disorders	1.3 (20)	1.9 (74)	0.6 (13)	1.4 (107)
- Other disorders[1]	0.6 (10)	0.3 (13)	0.6 (13)	0.4 (36)
- No disorder reported	10.4 (159)	7.4 (311)	15.5 (331)	10.2 (801)
Secondary problem identified by referrer (%/n)				
- Depression	1.5 (23)	1.8 (74)	0.7 (15)	1.4 (112)
- Mixed anxiety & depression	0.2 (3)	0.1 (6)	0.3 (6)	0.2 (15)
- Generalised anxiety disorder	63.4 (972)	64.5 (2698)	54.5 (1167)	61.5 (4837)
- Other anxiety disorders	0.4 (6)	0.8 (34)	0.5 (11)	0.6 (51)
- Other disorders[1]	0.7 (11)	1.0 (42)	1.1 (24)	1.1 (77)
- No disorder reported	33.8 (519)	31.8 (1329)	42.9 (919)	35.2 (2767)

[1] mental and behavioural disorder due to alcohol, bipolar affective disorder, somatoform disorder, eating disorder, disappearance and death of family member, mental disorders not otherwise specified.

Table 2. Improvement and effect size for patients with more than one contact

	Completed treatment (n = 2949)	**Dropped out of treatment** (n = 969)	**Unsuitable for further treatment** (n = 262)	**Total** (n = 4183)
Improvement				
Pre treatment				
PHQ-9 (mean/SD)	15.75 (6.26)	17.1 (5.73)	18.50 (5.57)	16.21 (6.15)
Median (IQR)	17 (12-21)	18 (13-22)	20 (15-23)	17 (12-21)
Post treatment				
PHQ-9 (mean/SD)	7.64 (6.87)	11.73 (6.93)	14.97 (7.12)	9.04 (7.27)
Median (IQR)	5 (2-12)	11(6-17)	16 (9-20)	8 (3-14)
- Effect size (CI)	1.24 (1.01 to 1.48)	0.85 (0.49 to 1.28)	0.56. (-0.12 to 1.42)	1.07 (0.88 to 1.29)
Pre treatment				
GAD-7 (mean/SD)	13.78 (5.24)	14.57 (4.71)	15.41 (4.63)	14.06 (5.11)
Median (IQR)	15 (10-18)	15 (12-18)	16 (13-19)	15 (11-18)
Post treatment				
GAD-7 (mean/SD)	6.91(6.07)	10.52 (6.14)	12.73 (5.79)	8.10 (6.37)
Median (IQR)	5 (2-11)	10 (5-16)	14 (8-17)	7 (3-13)
- Effect size (CI)	1.21 (1.03 to 1.43)	0.75 (0.45 to 1.13)	0.51 (-0.05 to 1.22)	1.04 (0.88 to 1.23)

Table 3. Relative risk, reliable and clinical change and recovery rates for patients with more than one contact

	Completed treatment	Dropped out of treatment	Unsuitable for further treatment	Total
PHQ-9 (%/n)				
- Rate of depression pre treatment	81.7 (2410/2945)	88.4 (857/967)	91.2 (239/262)	83.9 (3508/4177)
- Rate of depression post treatment	31.9 (941/2906)	58.2 (564/951)	71.4 (187/250)	40.5 (1695/4110)
- Relative Risk (95% CI)	0.40 (0.37 to 0.41)	0.67 (0.63 to 0.70)	0.82 (0.75 to 0.88)	0.48 (0.47 to 0.51)
- Reliable improvement & clinical significant change	47.3 (1378/2906)	27.1 (237/952)	16.4 (41/250)	40.8 (1676/4110)
- Reliable improvement only	13.6 (394/2906)	17.9 (169/952)	14.8 (37/250)	14.6 (600/4110)
- No improvement	39.0 (1134/2906)	57.0 (539/952)	68.8 (172/250)	44.6 (1834/4110)
GAD-7 (%/n)				
- Rate of anxiety pre treatment	84.8 (2502/2945)	90.2 (874/967)	92.4 (242/262)	86.5 (3620/4177)
- Rate of anxiety post treatment	36.4 (1073/2900)	62.6 (607/944)	74.4 (195/251)	44.9 (1878/4098)
- Relative Risk (95% CI)	0.44 (0.41 to 0.46)	0.71 (0.68 to 0.75)	0.84 (0.78 to 0.91)	0.53 (0.51 to 0.55)
- Reliable improvement & clinical significant change	47.1 (1367/2900)	25.1 (237/945)	15.9 (40/251)	40.1 (1643/4098)
- Reliable improvement only	13.8 (399/2900)	17.9 (169/945)	12.4 (31/251)	14.6 (599/4098)
- No improvement	39.1 (1134/2900)	57.0 (539/945)	71.7 (180/251)	45.3 (1856/4098)
IAPT [1] recovery rate (%/n)				
- Case at assessment	89.9 (2647/2949)	94.9 (918/969)	96.2 (252/262)	91.4 (3817/4183)
- Recovery rate	55.4 (1446/2609)	27.0 (244/904)	15.4 (37/240)	46.0 (1724/3756)

[1] Pre treatment PHQ-9 > 9 or GAD-7 % > 7 and post treatment PHQ-9 < 10 & GAD-7 % < 8

Table 4. Description of treatment received by patients with more than one contact

	Total patients (n = 4183)	**Completed treatment** (n = 2949)	**Dropped out of treatment** (n = 969)	**Unsuitable for further treatment** (n = 262)
Mean time: hours & minutes (SD)	3.15 (3.09)	3.32 (3.27)	2.28 (2.03)	2.49 (2.13)
Mean numbers of session of contacts (SD)	5.49 (4.31)	6.01 (4.58)	4.15 (3.15)	4.53 (3.61)
Number of face-to-face contacts (mean/SD)	2.33 (2.96)	2.49 (2.23)	1.90 (2.14)	2.03 (1.33)
Number of telephone contacts (mean/SD)	3.17 (3.01)	3.53 (3.30)	2.25 (2.21)	2.50 (2.78)
Patients with at least one[1] session of:		(%/n)	(%/n)	(%/n)
Depression self-help book (n=3311; 79.2%)[2]		70.0 (2319)	24.6 (780)	6.3 (209)
Anxiety self-help book (n=2021; 48.3%)		71.0 (1431)	22.8 (461)	5.8 (118)
Medication management (n=1459; 34.9%)		70.1 (1017)	26.7 (373)	6.7 (97)
Liaison (n=1279; 30.6%)		69.0 (883)	20.3 (259)	10.6 (135)
Information only (n=919; 22.0%)		75.4 (685)	20.0 (181)	5.8 (53)
Computerized CBT (n=491;11.7%)		76.7 (373)	21.0 (103)	3.1 (15)
High-intensity CBT (n=275; 6.6%)		79.6(219)	16.4 (45)	4.0 (11)
Counselling (n=81; 1.9%)		64.2 (52)	24.7 (20)	11.1 (9)

[1] Patient could receive more than one treatment

[2] Percentage of all patients receiving any treatment

Table 5. The high and low intensity treatment groups (n= 2949)

	Low intensity treatment[1] only (n=2626)	**High intensity treatment[2]** (n=219)
Gender female/male (%)	65.9/34.1	69.4/30.6
Age (mean/SD)	39.9 (13.15)	38.9 (12.80)
White British (%)	93.9	94.1
N treatments (mean/SD)	5.37 (3.44)	13.87 (7.99)
Median (IQR)	4 (3 – 7)	13 (8 – 19)
Mean time: hours and minutes (SD)	2.53 (1.54)	10.00 (6.36)
Depression (PHQ-9)		
Pre treatment (mean/SD)	15.66 (6.22)	16.85 (6.36)
Median (IQR)	16.0 (11 – 21)	18.0 (13 – 22)
Post treatment (mean/SD)	7.49 (6.86)	9.25 (6.96)
Median (IQR)	5.0 (2 - 12)	8.0 (3 - 15)
- Effect size (CI)	1.25 (1.01 to 1.51)	1.14 (0.3 to 2.06)
Anxiety (GAD-7)		
Pre treatment (mean/SD)	13.68 (5.23)	15.12 (5.04)
Median (IQR)	15.0 (10 - 18)	16.0 (12 – 19)
Post treatment (mean/SD)	6.77 (6.06)	8.49 (6.07)
Median (IQR)	5.0 (2 - 11)	7.0 (3 - 14)
- Effect size (CI)	1.22 (1.02 to 1.46)	1.19 (0.53 to 2.00)

[1] Low intensity treatment only = depression self-help book, anxiety self-help book, medication management, signposting, information only, computerized CBT.
[2] High intensity treatment includes patients gone straight to HIT (n = 19) as well as stepped up patients = CBT

Figure(s)

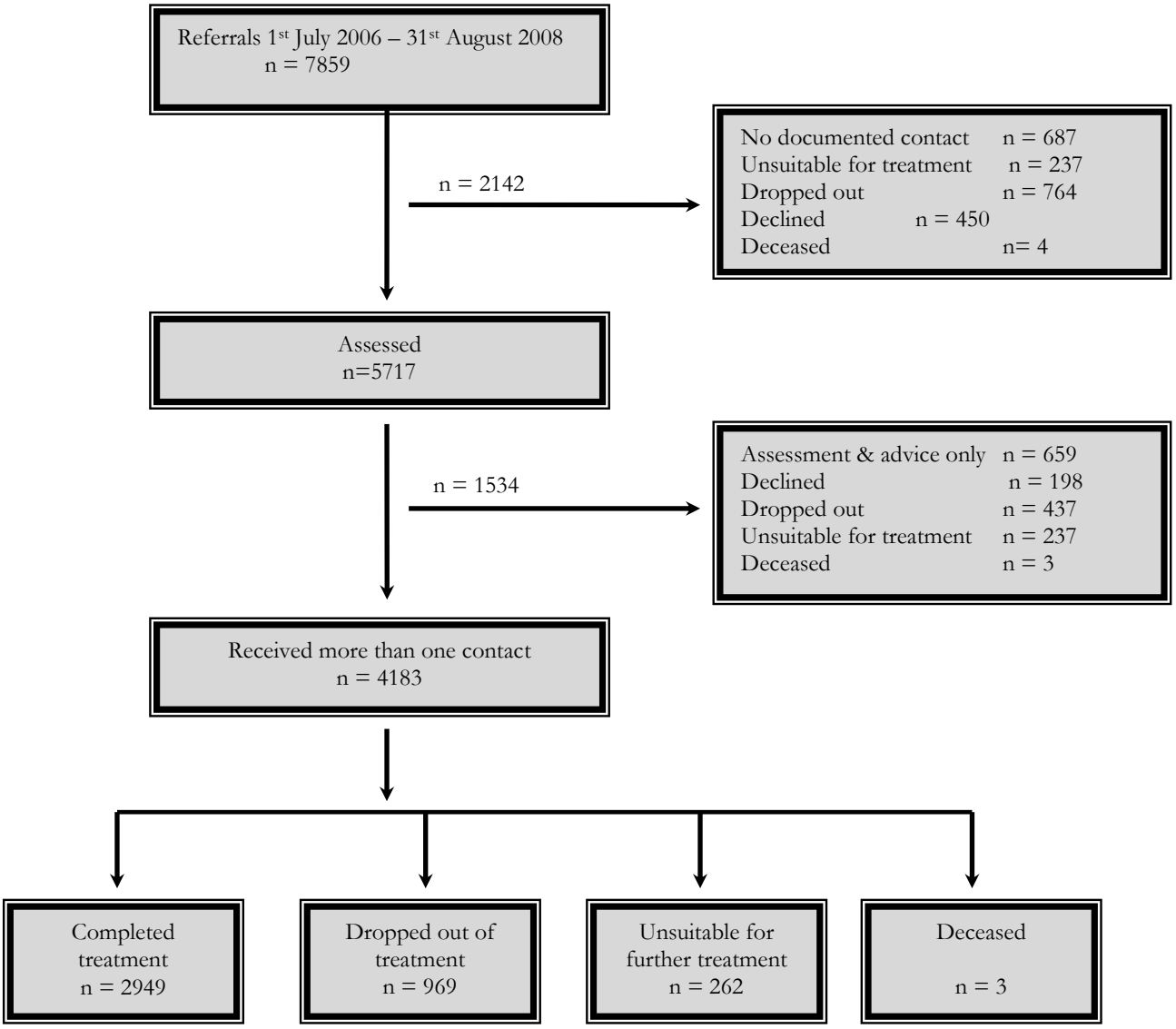

Figure 1. Flow diagram of patient pathways & status

Acknowledgement

We would like to acknowledge Rupert Suckling and the staff and managers of the clinical site. We would also like to acknowledge Simon Gilbody, Dean McMillan and David M. Clark for helpful comments and advice on this article

***Conflict of Interest**

David Richards receives financial support from the Department of Health IAPT Programme. Gunilla Borglin declares that she has no conflicts of interest.

***Role of the Funding Source**

David Richards receives financial support from the NIHR EME programme, the PenCLAHRC programme, UK and the Department of Health IAPT Programme. Gunilla Borglin was supported by a strategic post-doctoral funding grant from Blekinge Institute of Technology, Sweden. The funders had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

***Contributors**

David Richards conceived and designed the study. Gunilla Borglin undertook the data cleaning and analyses. Both authors undertook the statistical analysis. Both authors contributed to and have approved the final manuscript.