Durability of Clinical Gains:

Do the treatment gains seen in IAPT services last?

Michelle J Lee[1], Annette G M O'Toole[2], Ineke Wolsey[3], David M Clark [4]

[1]Charlie Waller Institute/Oxford Academic Health Science Network

School of Psychology and Clinical Language Sciences

University of Reading, Berkshire, UK.

[2]Healthy Minds (IAPT Service), Buckinghamshire, UK.

[3]Oxford Academic Health Science Network

Anxiety and Depression (IAPT) Network Manager, UK.

[4]Professor and Chair of Experimental Psychology, Oxford University. Oxford, UK.

Anxiety and Depression (IAPT) Network Clinical Network Lead

Abstract

Background: The Improving Access to Psychological Therapies (IAPT) service regularly report recovery and improvement figures showing the success of the programme. However, it is unclear how long patients remain well following treatment within IAPT. There is little research looking at relapse rates in anxiety and depression within IAPT, and methodologies have varied in terms of outcome definitions, treatment focus and scope. This has resulted in an imprecise and mixed picture in terms of the relapse rates of IAPT patients. **Aims:** The main aim of this study was to complete a 6-month and 12-month follow-up of patients receiving routine treatment at Step 2 *and/or* Step 3 across several IAPT services in order to ascertain durability of clinical gains more generally. **Method:** Patients discharged in June 2016 from six IAPT services within the south of England were invited to complete a 6-month and 12-month follow-up. Baseline, end of treatment and follow-up data (using measures of depression and anxiety used routinely within IAPT services) were used to calculate reliable improvement/recovery rates. **Results:** We found rates of reliable improvement and recovery to be high at the end of treatment; for most patients, these gains were maintained at 6 and 12 months. **Conclusions:** Based on our findings, those who do well in treatment, appear to maintain their gains. However, these findings are limited by a small and selective sample, which may not be representative of the broader range of patients seen within IAPT services.

Keywords: anxiety, depression, recovery, improvement, IAPT services

Introduction

The Improving Access to Psychological Therapies (IAPT) programme (Layard et al, 2006) was developed to offer greater access to treatment for people experiencing anxiety disorders and/or depression. The IAPT programme only offers evidence-based interventions approved by the National Institute for Health and Care Excellence (NICE) for depression or anxiety disorders (NICE, 2000), including (but not limited to) Cognitive Behavioural Therapy (CBT).

This programme follows a stepped care approach to therapy – a system of delivering treatment so that the most effective, yet least resource intensive treatment is delivered to patients in the first instance, only stepping up to more intensive treatments as clinically necessary. In practice, this means that those patients who fall in the mild to moderate range on standardised measures of depression and anxiety are offered low intensity interventions such as group work or telephone guided self-help with Psychological Well-being Practitioners or computerised CBT and are stepped up to high intensity interventions such as face to face therapy if they do not recover after their initial intervention. When necessary, patients are offered high intensity treatment as a first option – for example, if they present with severe depression or have more complex anxiety disorders such as post-traumatic stress disorder. Additionally, the programme differs from existing services in several other key ways, such as allowing for self-referral (as well as referral from primary care and other sources) and measuring symptoms at every contact (either in person or over the phone), using a minimum data set (MDS).

Common mental illnesses such as anxiety and depression have a huge financial impact in terms of welfare costs as lost tax revenue; providing effective treatment is very important from an economic point of view and providing CBT has been shown to be helpful in getting people back to work (Proudfoot et al, 1997). In light of this, the programme also hoped to improve patients' well-being to help them remain in

work or facilitate a return to work. Employment coaches worked with the services to help with this, the assumption being that the investment in psychological therapies would pay for itself by reducing "other depression and anxiety-related public costs" (Clark et al, 2009; p. 911).

Prior to national rollout of the programme, the Department of Health funded a pilot looking at two demonstration sites in Doncaster and in Newham, which offered CBT to patients in primary care with depression and/or anxiety, resulting in around 5500 referrals between August 2006 and September 2007, of which around 3500 received treatment (Clark et al, 2009).

Both sites achieved recovery rates between 55-56% at end of treatment and 5% of patients had improved their employment status (Clark et al, 2009), however, the two demonstration sites *did not* routinely follow up their patients to check whether clinical gains had been maintained in the longer term. A one-off postal follow-up survey was given to patients from both demonstration sites to specifically examine this, using PHQ-9 and GAD-7 scores (Clark et al, 2009). Doncaster followed up around half of all eligible patients, finding that 50% of these people were recovered at 42 week (average) follow up (compared with 56% at end of treatment). Newham collected data on just over one third of their eligible clients, finding that 42% of their sample were recovered at 42 week (average) follow up (compared with 57% at end of treatment). The team recommended that future IAPT sites should offer routine follow-ups 3-6 months post-discharge and offer booster sessions at that stage if, patients show signs of deterioration when followed up (Clark et al, 2009). Despite the importance of this finding, IAPT services are not currently commissioned to collect follow up data; instead, KPIs focus on access rates and recovery rather than *maintenance* of clinical gains. Consequently, follow-up data is not routinely collected or reported within annual IAPT reports.

The Government target is that 50% of eligible referrals to local IAPT services should move to recovery and the most recent report indicates that this target is being met (Community & Mental Health Team, 2017). The latest Key Performance Indicators

report using data from referrals in Q4 of 2017/18 (i.e. January – March 2018) reported 378,574 new referrals into IAPT services across the country. Of those referred, 263,636 patients entered treatment and of those 139,053 finished a course of treatment through IAPT with 51.7% moving to recovery at discharge.

Durability of clinical gains has only been addressed in a small number of research studies, which are limited by focusing on either a *single* IAPT service and/or *a single* treatment step. Because of the *nature* of treatment offered at the two demonstration sites, the Clark et al (2009) study focused predominantly on durability of clinical gains at Step 2 within the Doncaster demonstration site and Step 3 within the Newham site.

A more recent study by Ali et al (2017) examined the durability of clinical gains by looking at relapse rates, but again, this was within a *single* IAPT service, and they only looked at patients receiving low-intensity treatment (patients receiving Step 2 treatment who were subsequently stepped up to high-intensity treatment were excluded from the study). They followed up 439 patients with remission of symptoms at discharge for a year (out of a pool of 2100 potentially eligible patients who were identified from treatment discharge records) by examining monthly anxiety and depression (GAD-7 and PHQ-9) scores. Overall, they found that 53% of these patients had relapsed within one year, and 80% of these relapses occurred within the first 6 months after treatment. They also found that having residual depression symptoms at the end of treatment increased the risk of relapse – and that these individuals tended to relapse sooner (although residual anxiety symptoms were not a predictor of relapse; Ali et al, 2017; Paykel, 2008). This indicator of poor clinical gains contrasts with data from the follow up of the demonstration sites which suggested the majority of clients stayed well in the first 10 months after discharge. The cause of the discrepancy in findings is unclear but may reflect the specific services examined here.

Further research, looking at *reliable* improvement and *reliable* recovery rates of follow up of patients treated at Step 2 *and* Step 3 across s*everal* IAPT sites is clearly

needed to obtain a more comprehensive picture around the durability of clinical gains within IAPT services more generally. In order to address this need, the current research sought to follow up a cohort of patients seen within several IAPT services and across different treatment steps (i.e. at Step 2 and/or Step 3) for 12 months after discharge. Patients discharged during June 2016 across six IAPT services within the south of England were followed up at 6-month and 12-month periods in order to ascertain whether reliable improvement and recovery seen at the end of treatment was maintained in the longer term.

Method

Design

This was a prospective, longitudinal cohort study of patients who had received treatment at Step 2 and/or Step 3 within one of six IAPT services in the south of England and had completed treatment and been discharged during June 2016. All six services fall under the umbrella of the Oxford Academic Health Science Network (OAHSN) who led on this research, and included Healthy Minds (Buckinghamshire), Talking Space (Oxfordshire), Talking Therapies (Berkshire), Talk for Change (Milton Keynes), Bedfordshire Wellbeing Service, and Luton Wellbeing Service. The study was considered a routine service evaluation and was given clearance by Oxford Health Audit committee. Patients were followed up on two occasions – six months and twelve months post-discharge (i.e. during December 2016 and June 2017).

Participants and procedure

Clinical staff from all six IAPT services involved in this study were asked to seek consent from patients discharged in June 2016 who had at least two treatment sessions (including their initial assessment) to be contacted for a follow up.

Patients who had indicated they were willing to be contacted for a follow-up at six- and twelve-months post-discharge were assigned an ID number by their local service. Relevant data about these individuals (gender, age, ethnicity, primary

presenting problem, baseline and end of treatment MDS scores and ADSMs, employment status and treatment step(s) received) were recorded, along with those patients who had been discharged but had not given consent to be contacted for follow-up (anonymity was preserved using the assigned ID numbers).

In December 2016 and July 2017, patients who had given permission to be contacted again were each sent a hard copy of the follow-up measures (see below); they then had the option of completing these and returning them to their service in the stamped addressed envelope provided, or using a link provided in the information letter sent out to complete the survey online.

Online surveys and returned hard copies were checked on a daily basis during the working week. If anyone scored more than zero (not at all) on question 9 of the PHQ-9, ('Been bothered by thoughts that you would be better off dead or of hurting yourself in some way'), this was flagged up by researchers and services made arrangements to call them to undertake a risk assessment. Additionally, contact details of local IAPT services and other sources of support and help were also detailed in the follow-up measures (both the hard copy and online version), in case any of the participants needed further help.

Patients requiring a translator were identified by services and this was provided where necessary (via a follow-up over the telephone). Wherever possible (i.e. when there were enough admin/clinical staff within the local IAPT service), those who had given consent to be contacted by phone were contacted to remind them about the study and when time permitted, participants were also given the option of doing the follow-up over the phone.

Measures and sources of data

Measures consisted of the Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer & Williams, 2001) and the Generalised Anxiety Disorder Questionnaire (GAD-7; Spitzer, Kroenke, Williams & Lowe, 2006), both of which are used routinely within IAPT services to measure symptoms of depression and anxiety. The PHQ-9 is a nine-item

screening tool for depression; patients rate the frequency with which each item has bothered them over the previous two weeks from 0-3, with total possible scores ranging from 0-27 and a cut-off of 10 is used to detect clinically significant depression. The GAD-7 is a seven-item screening tool for depression; patients rate the extent to which each item has bothered them over the previous two weeks from 0-3, with total possible scores ranging from 0-21 and a cut-off of 8 is used to detect clinically significant anxiety.

Additional Anxiety Disorder Specific Measures (ADSMs) were used where relevant (i.e. if the patient had been treated for that particular disorder) as follows:
The Agoraphobia Mobility Inventory (Chambless, Caputo, Jasin, Gracely & Williams, 1985) is a 27-item self-report questionnaire used to determine the extent to which different situations are avoided because of discomfort or anxiety. Scores can range from 27-135, with a cut-off of 60 used to detect clinically significant agoraphobia.
The Social Phobia Inventory (SPIN; Connor et al, 2000) is a 17-item self-report questionnaire used to determine the extent to which an individual has experienced different symptoms over the previous week. Total scores can range from 0-68, with a cut-off of 19 used to detect clinically significant agoraphobia.
The Obsessive-Compulsive Inventory (OCI; Foa, Kozak, Salkovskis, Coles & Amir 1998) is a 42-item self-report questionnaire used to determine the extent to which an individual has experienced different symptoms over the previous month. Total scores can range from 0-168, with a cut-off of 40 used to detect clinically significant Obsessive Compulsive Disorder (OCD).

The Impact of Events Scale Revised (IES-R; Weiss & Marmar, 1996) is a 22-item self-report questionnaire used to determine the extent to which an individual has experienced different symptoms after a stressful life event over the previous month. Total scores can range from 0-88, with a cut-off of 33 used to detect clinically significant Post Traumatic Stress Disorder (PTSD).

The Health Anxiety Inventory (HAI; Salkovskis, Rimes, Warwick and Clark, 2002) is an 18-item self-report questionnaire used to determine the extent to which an

individual has experienced thoughts about illness over the past week. Total scores can range from 0-54, with a cut-off of 18 used to detect clinically-significant somatoform disorder.

Results

Response rates:

3027 patients were discharged across the six services during June 2016. Of these, consent was obtained from 323 patients (11%). Practical issues meant that not all patients may have been informed about the study (hence the low rate of response).

Of the 323 patients who agreed to be contacted after discharge, 85 patients completed the first follow-up in December 2016 (26% response rate) and 57 patients completed the second follow-up in June 2017 (18% response rate) – 21 of whom had not completed the first follow-up. A total of 106 patients completed *at least one* of the follow-ups (33% of those consenting to follow up) and 36 patients completed both the first *and* second follow-up (11% of those consenting to follow-up).

It should be noted that response rates varied widely from service to service in terms of giving consent to be followed up (which ranged from 6%-49% across the six services), and participating in the first and second follow up once consent had been given (which ranged from 8% to 67% and 0-67% respectively of those consenting to be contacted). Response rates were higher in services with a member of staff who had time available to call and remind participants about the study and offer the option of doing their follow-up over the phone.

Mi*ssing data at 6-month follow-up:*

When data was missing for the 6-month follow up but *had* been gathered for the 12-month follow-up, this latter data point was included for the 6-month reporting of reliable improvement and reliable recovery.

Sample characteristics of those participating in follow-up

Of those participating in at least one follow-up (n = 108), 52% were female and 83% were white British. The average age of the sample was 43 years (SD=14.68, range from 18-79 years).

The distribution of primary problems was as follows: 34% presenting with Depressive Disorder, 17% presenting with Generalised Anxiety Disorder, 11% presenting with PTSD, 10% presenting with Mixed Anxiety/Depression, 9% presenting with Recurrent Depressive Disorder, 4% presenting with OCD, 3% presenting with Panic Disorder, 2% presenting with Adjustment Disorder and the remaining presenting with Health Anxiety (2%), Specific Phobia (2%) and Social Phobia (1%).

Most of the sample (63%) had received treatment at either Step 3 or a combination of Step 2 and Step 3, with 37% of the sample receiving treatment only at Step 2. Those receiving treatment only at Step 2 had an average of 6.36 treatment sessions, and those at Step 2 and 3 or just Step 3 had an average of 15.04 sessions.

Representativeness of the sample

Key characteristics of those who did not consent to be contacted again after discharge, those who did consent to be contacted but did not participate in the follow-up, and those who consented and participated in the follow-up were calculated separately and are shown in Table 1.

Table 1 Characteristics of those not giving consent, giving consent without participation and giving consent with participation.

	No Consent	Consent but no participation	Consent and participation	p
Age	45.70[a]	42.72[a]	50.00[a]	>.05
PHQ-9 baseline	12.86[a]	16.91[b]	14.27[ab]	<.001
PHQ-9 change[1]	5.37[a]	9.51[b]	9.36[b]	<.001
GAD-7 baseline	10.64[a]	13.51[b]	11.45[ab]	<.005
GAD-7 change[1]	4.30[a]	6.58[b]	6.81[b]	<.005
Number of sessions	6.48[a]	17.41[b]	14.95[b]	<.001

[1] These scores represent the change from baseline to end of treatment

Please note that scores on the same row sharing the same superscript letter do not differ significantly from each other.

The three groups did not differ from each other in terms of age. However, there were significant differences across groups in terms of PHQ-9 scores (both baseline scores and change scores from start to end of treatment), GAD-7 scores (baseline scores and change scores from start to end of treatment), and number of treatment sessions received. Of note, baseline and change scores for anxiety and depression symptoms were significantly higher in both groups who had given consent as compared to those not giving consent, and those who gave consent received significantly more treatment sessions than those who did not give consent.

Reliable improvement

A patient was deemed to have shown reliable improvement if there was a *decrease* in one or both assessment measures (the PHQ-9 and/or the relevant ADSM) that surpassed the measurement error on that assessment measure, with no *increase* beyond the measurement error on either measure. Reliable improvement can be shown regardless of 'caseness' at the start of treatment.

Durability of clinical gains

Reliable Improvement 6-month follow-up data – all participants

Reliable improvement at 6-month follow-up (data for the 12-month follow-up was used if 6-month follow-up data was missing) could only be calculated for 88 patients. This is because three patients did not complete their PHQ-9 or GAD-7 measure at follow-up, and 15 patients had agreed to take part in the follow-up but had not consented to the up flow of their baseline/end of treatment information for the purposes of the study (so reliable improvement could not be calculated). Table 2 shows the reliable improvement of patients at 6-month follow-up (compared with end of treatment status).

Table 2 Reliable improvement at 6-month follow-up (all patients)

		6-month follow-up			
		Reliable *Deterioration*	No Reliable Improvement	Reliable *Improvement*	Total
End of Treatment	Reliable *Deterioration*	0	0	0	0
	No Reliable Improvement	1	8	8	17
	Reliable *Improvement*	1	9	61	71
	Total	2	17	69	88

By end of treatment, 81% of all patients showed reliable improvement from baseline, 19% showed no reliable improvement, and no-one showed reliable deterioration. *Of those who showed reliable improvement at end of treatment (n=71),* 86% continued to be reliably improved at 6-month follow-up, 13% were no longer reliably improved, and 1% had reliably deteriorated. *Of those who showed no reliable improvement at end of treatment (n=17),* 47% continued to not be reliably improved, 47% had moved to being reliably improved, and 6% had reliably deteriorated.

Reliable improvement at 6-month follow-up was also calculated for patients separately according to which treatment step they had received (Step 2 only versus Step 2 and 3 or Step 3 only). Table 3 shows reliable improvement at six months for

patients receiving Step 2 treatment only (compared with their end of treatment status).

Table 3 Reliable improvement at 6-month follow-up (Step 2 only patients)

		6-month follow-up			
		Reliable *Deterioration*	No Reliable Improvement	Reliable *Improvement*	Total
End of Treatment	Reliable *Deterioration*	0	0	0	0
	No Reliable Improvement	0	3	5	8
	Reliable *Improvement*	0	1	23	24
	Total	0	4	28	32

Table 3 shows that by end of treatment, 75% of all patients receiving Step 2 treatment showed reliable improvement from baseline (24/32), 25% showed no reliable improvement (8/32), and no-one showed reliable deterioration. *Of those who showed reliable improvement at end of treatment (n=24),* 96% continued to be reliably improved at 6-month follow-up, and 4% were no longer reliably improved (no-one had reliably deteriorated). *Of those who showed no reliable improvement at end of treatment (n=8),* 63% continued to not be reliably improved, and 37% had moved to being reliably improved (no-one had reliably deteriorated).

Table 4 shows reliable improvement at six months for patients receiving either Step 3 treatment only or a combination of both Step 2 and Step 3 treatment (compared with end of treatment status).

Table 4 Reliable improvement at 6-month follow-up (Step 3 only or Step 2&3 patients)

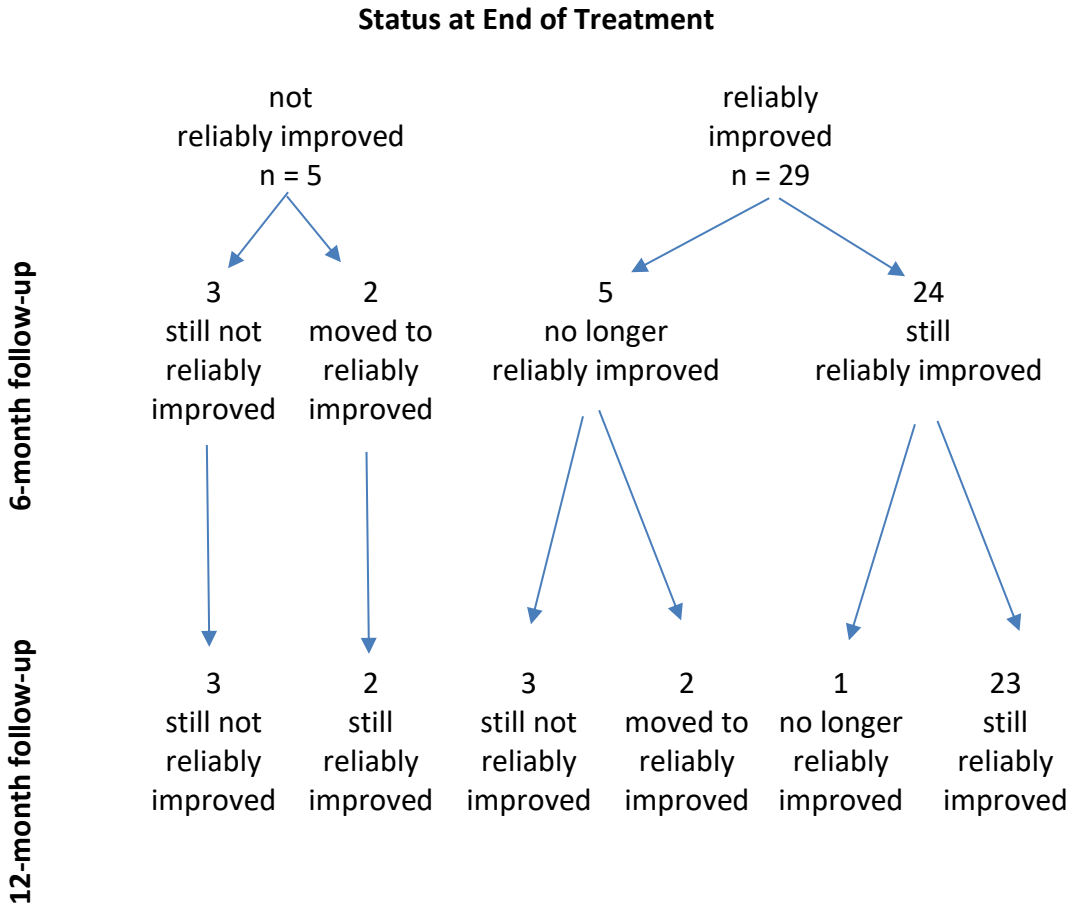
		6-month follow-up			
		Reliable *Deterioration*	No Reliable Improvement	Reliable *Improvement*	Total
End of Treatment	Reliable *Deterioration*	0	0	0	0
	No Reliable Improvement	1	5	3	9
	Reliable *Improvement*	1	8	37	46
	Total	2	13	40	55

By end of treatment, 84% of all patients receiving Step 3 treatment only or a mixture of Step 2 and Step 3 treatment showed reliable improvement from baseline (46/55), and 16% showed no reliable improvement (9/55). No-one showed reliable deterioration. *Of those who showed reliable improvement at end of treatment (n=46),* 80% continued to be reliably improved at 6-month follow-up (37/46), 17% were no longer reliably improved (8/46) and 2% had reliably deteriorated (1/46). *Of those who showed no reliable improvement at end of treatment (n=9),* 56% continued to not be reliably improved (5/9), 33% had moved to being reliably improved (3/9), and 11% had reliably deteriorated (1/9).

Reliable improvement for participants completing both follow-ups

Reliable improvement for the patients who completed both follow-ups was also calculated and is presented in Figure 1. This could only be done for 34 out of the 36 patients who did both, because two had not consented to the up flow of their baseline/end of treatment data, so reliable improvement could not be calculated.

Figure 1 Reliable improvement at 6 and 12-month follow-up for all patients who completed both follow-ups. Status is also indicated at 12-month follow-up.



It can be seen from Figure 1 that of those reliably improved at end of treatment, 83% continued to be reliably improved at 6-month follow-up and 79% continued to be reliably improved at 12-month follow-up.

Reliable recovery

A patient is deemed to have shown a reliable recovery if they show a reliable *improvement* in their scores (see definition above) *as well as* a change from '*caseness*' at baseline to *non-caseness at end of treatment or follow-up*. Caseness at

baseline means that either the baseline PHQ-9 score *or* the relevant ADSM score (or both) are *above* the relevant caseness threshold (i.e. at least one measure needs to be above the threshold). Non-caseness at end of treatment or follow-up means that the PHQ-9 score and relevant ADSM score at end of treatment or follow-up are *both* below the relevant caseness threshold for that particular measure (i.e. if scores on the PHQ-9 are below threshold but the score on the relevant ADSM is above threshold, this would not constitute non-caseness). Patients that start their course of treatment below caseness are not included in reliable recovery counts.

Reliable recovery rates at 6-month follow-up – all participants

Reliable recovery rates were calculated for all participants at 6-month follow-up and these data are shown in Table 5.

Table 5 Reliable Recovery at 6-month follow-up (All participants)

		6-month follow-up		
		No reliable recovery	Reliable recovery	Total
End of treatment	No reliable recovery	27	9	36
	Reliable recovery	9	38	47
	Total	36	47	83

By end of treatment, 57% of all patients demonstrated reliable recovery from baseline (47/83), and 43% showed no reliable recovery from baseline (36/83). *Of those who showed reliable recovery at end of treatment (n=47),* 81% continued to be reliably recovered at 6-month follow-up (38/47), and 19% were no longer reliably recovered (9/47). *Of those who showed no reliable recovery at end of treatment (n=36),* 75% continued to not be reliably recovered (27/36), and 25% had moved to being reliably recovered (9/36).

Reliable recovery rates were also calculated separately for participants according to which step of treatment they had received (those receiving Step 2 treatment only,

and those receiving either Step 3 treatment or a combination of Step 2 and Step 3 treatment. Table 6 shows recovery rates for those receiving Step 2 treatment only.

Table 6 Reliable Recovery at 6-month follow-up (Step 2 only participants)

		6-month follow-up		
		No reliable recovery	Reliable recovery	Total
End of treatment	No reliable recovery	8	6	14
	Reliable recovery	2	15	17
	Total	10	21	31

By end of treatment, 55% of all patients receiving Step 2 treatment only demonstrated reliable recovery from baseline (17/31), and 45% showed no reliable recovery from baseline (14/17). *Of those who showed reliable recovery at end of treatment (n=17),* 88% continued to be reliably recovered at 6-month follow-up (15/17), and 12% were no longer reliably recovered (2/17). *Of those who showed no reliable recovery at end of treatment (n=14),* 57% continued to not be reliably recovered (8/14), and 43% had moved to being reliably recovered (6/14).

Table 7 shows recovery rates in those patients either receiving Step 3 only or a combination of Step 2 and Step 3 treatment.

Table 7 Reliable Recovery at 6-month follow-up (Step 3 only and 2&3 participants)

		6-month follow-up		
		No reliable recovery	Reliable recovery	Total
End of treatment	No reliable recovery	19	3	22
	Reliable recovery	7	23	30
	Total	26	26	52

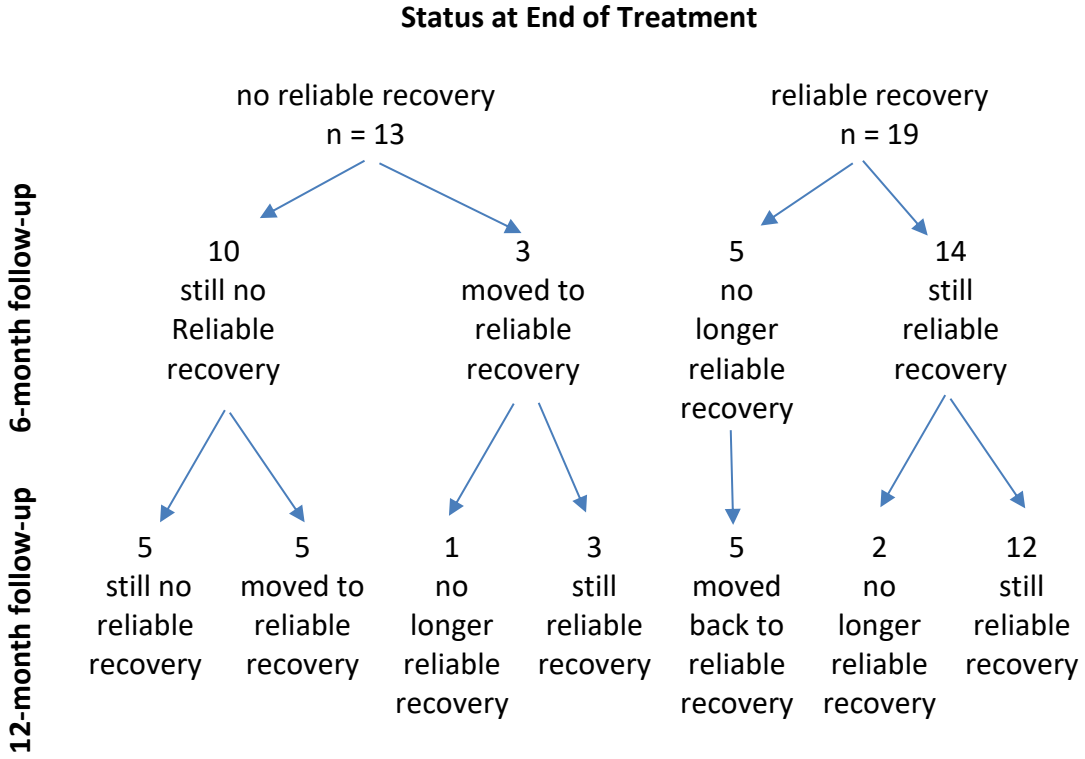
By end of treatment, 58% of all patients receiving either Step 3 or a combination of Step 2 and Step 3 treatment demonstrated reliable recovery from baseline (30/52),

and 42% showed no reliable recovery from baseline (22/52). *Of those who showed reliable recovery at end of treatment (n=30),* 77% continued to be reliably recovered at 6-month follow-up (23/30), and 23% were no longer reliably recovered (7/30). *Of those who showed no reliable recovery at end of treatment (n=22),* 86% continued to not be reliably recovered (19/22), and 14% had moved to being reliably recovered (3/22).

Reliable recovery for participants completing both follow-ups

Reliable recovery for the patients who completed both follow-ups was calculated at both time points. This could only be done for 33 out of the 36 patients who completed both, because two had not consented to the up flow of their baseline/end of treatment data, and one did not meet 'caseness' at the start of treatment, so reliable recovery could not be calculated. This information is shown in Figure 2.

Figure 2 Reliable recovery at 6 and 12-month follow-up (patients who completed both follow-ups)



It can be seen from Figure 2 that of those reliably recovered at end of treatment, 74% continued to be reliably improved at 6-month follow-up and 63% continued to be reliably improved at 12-month follow-up.

Discussion

The aim of the current research was to examine the durability of clinical gains obtained in patients who had been treated at one of six IAPT services across the south of England. This is an important contribution to the field, given that to the best of our knowledge, no one has conducted similar research following up patients from across *both* treatment steps and across *multiple* IAPT services.

The overwhelming message from this research is that gains seen in treatment are largely maintained at follow-up. For example, of those patients with *reliable improvement* at end of treatment (i.e. 81% of our sample), 86% continued to be reliably improved at 6-month follow-up, and of those patients with *reliable recovery* at end of treatment (i.e. 57% of our sample), 81% continued to be reliably recovered at 6-month follow-up.

We do, however, acknowledge that our findings are based on a very small data set, and so these results must be interpreted with a degree of caution. Additionally, our follow up sample formed only a very small percentage of those discharged during June 2016 and may not have been particularly representative of this larger group. Whilst around a quarter of those who had given consent to follow-up participated in the research (26%), this only represents around 3% of the patients who were *actually discharged* during June 2016 (i.e. 3% of all patients who could have been included in the current study). Additionally, comparison of those consenting to and completing the follow-up versus those not giving consent highlighted significant differences on several indices, including initial severity of clinical presentation and

treatment impact. Those participating in the study had significantly higher baseline anxiety and depression scores, received a substantial number of sessions from their service (15 sessions on average) and demonstrated greater reduction in anxiety and depression scores by the end of treatment. Whilst this suggests that patients with significant difficulties who receive more than the average number of treatment sessions *and* do well clinically tend to maintain these clinical gains in the longer term, it does call into question how representative these individuals are of all those patients discharged from services during June 2016. A higher response rate of those consenting to be followed up, as well as those completing the follow up itself is needed in order to get a clearer picture of how well clinical gains are maintained more generally in other types of patient.

These findings are consistent with Clark et al (2009)'s findings in the two demonstration sites that clinical gains are largely maintained (in their case at 10 month follow up); however, clinical gains were less likely to be maintained by patients followed up within the Ali et al (2017) study.

In terms of data collection, practical issues may have also impacted on the response rate and number of patients giving consent in the first instance; undertaking a large-scale study involving several different IAPT services is challenging. The dates chosen to complete the follow-ups were not ideal; conducting the research during busy holiday periods (in this case December and June) is likely to have impacted the rate of participation. Obtaining consent to contact patients for the follow-up was challenging. Services were asked to obtain consent to follow up patients (via therapists for each patient discharged in June 2016) at their final treatment session. However, this may not have happened in all cases for a number of reasons (lack of communication within services may have meant that some therapists were simply not aware of the need to ask for consent, and therapists may have simply forgotten to ask), meaning that rates of consent were very low (consent was only received from 11% of patients across the six services as a whole). In this study, services were given guidance around obtaining consent, but in most services, lack of staff meant that there was no dedicated time to ensure that this was happening systematically;

by contrast, in one service with a dedicated researcher, consent was obtained for around *half* of all the patients being discharged. For future research, it may be preferable to introduce consent-taking at the *start* of treatment – possibly at the initial assessment, in order to ensure a more routine and systematic approach.

There was also a great deal of variance in terms of response rate from service to service; this may possibly reflect the need for a dedicated researcher within each IAPT service to contact patients who have consented to but not yet participated in the follow-up. It is not always feasible for services to allocate resources (in terms of staff) to research which is outside of their remit. Having a member of the research team working on an honorary contract within each of the services may be something to consider for future research, as this will reduce the workload for already over-stretched services.

Thought should also be given to the nature of *how* the data is collected. Patients were offered the choice of completing a 'hard copy' of the questionnaire and sending it back via an enclosed stamped addressed envelope, or filling in an online survey (both options were used equally), and in some special cases, to answer questions over the telephone. In future, data could ideally be collected in more innovative ways, for example through the development of an app, which could also be designed to facilitate treatment gains. This would allow for more frequent monitoring (monthly) and would alert services sooner of any signs of recurrence of symptoms. Top-up sessions could then be offered to prevent future relapses.

In conclusion, we have reported on an important study, which for the first time has attempted to assess the durability of clinical gains seen in treatment at IAPT services by following up a range of patients from a variety of services in the south of England. Several methodological issues have been highlighted, which illustrate the difficult nature of carrying out such research, and suggestions have been made for how research might be modified in future studies. However, notwithstanding these difficulties, the research has highlighted the importance of such work and has

concluded that within our small sample, the treatment gains seen in these services are largely maintained after discharge.

Acknowledgements

The project team would like to thank John Pimm (Healthy Minds, Oxford Health NHS Foundation Trust) Jo Ryder and Samantha Sadler (Talking Space, Oxford Health NHS Foundation Trust); Joanne Rollings, Nicola Farrin and Nicola Graham (Talking Therapies, Berkshire Healthcare NHS Foundation Trust); Jon Olsen and Sophie Fitzgerald (Talk for Change, Milton Keynes, Central and North West London NHS Foundation Trust); Douglas Hiscock (Bedford Wellbeing Service & Luton Wellbeing Service, East London NHS Foundation Trust) for all their support with the project. We would also like to thank the members of the patient forum group within the OAHSN who helped to guide the content and wording of the follow up surveys, and all the patients who so kindly gave up their time to complete the follow up surveys.

Ethical statement

Ethical approval was not needed for this study. It was considered a routine service evaluation and was given clearance by Oxford Health Audit committee.

Conflict of interest

Michelle Lee, Annette O'Toole, Ineke Wolsey and David Clark have no conflicts of interest with respect to this publication.

References

Ali, S., Rhodes, L., Moreea, O., McMillan, D., Gilbody, S., Leach, C., Lucock, M., Lutz, W and Delgadillo, J. (2017). How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. *Behaviour Research and Therapy, 94,* 1-8.

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

Community & Mental Health Team, NHS Digital. (April 2017). Improving Access to Psychological Therapies (IAPT). Executive Summary (online) Available at: *http://www.digital.nhs.uk/catalogue/PUB30035*. Date Accessed: 29 August 2017.

Community & Mental Health Team, NHS Digital. (October 2016, Revised June 2017). Psychological Therapies: Annual Report on the use of IAPT services (online). Available at: *https://digital.nhs.uk/catalogue/PUB22110.* Date Accessed: 8 September 2017.

Connor, K.M. Davidson, J., R., Churchill, L., E., Sherwood, A., Foa, E., Weisler, R., H. (2000) Psychometric Properties of the Social Phobia Inventory, *The British Journal of Psychiatry, 176*, 379-386.

Foa, E. B., Kovak, M. J.Salkovskis, P. M.Coles, M. E., & Amir, N. (1998). The validation of a new obsessive-compulsive disorder scale: The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment, 10,* 206-214.

Jones, P. W., Harding, G., Berry, P., Wiklund, I., Chen, W. H., and Leidy, N. (2009). Development and first validation of the COPD Assessment Test. *The European Respiratory Journal*, 34, 648-654.

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

Layard, R., Clark, D., Bell, S., Knapp, M., Meacher, B., Priebe, S., Turnberg, L., Thornicroft, G., & Wright, B. (2006). The depression report; A new deal for depression and anxiety disorders. *The Centre for Economic Performance's Mental Health Policy Group*, LSE.

Meyer TJ, Miller ML, Metzger RL, Borkovec (1990) TD: Development and Validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, *28,* 487-495.

Paykel, E. S. (2008). Partial remission, residual symptoms, and relapse in depression. *Dialogues in Clinical Neuroscience, 10,* 431-437.

Proudfoot, J., Guest, D., Carson, J., Dunn, G., and Grey, J. (1997) Effect of cognitive behavioural training on job-finding among long-term unemployed people. *Lancet, 350,* 96-100.

Salkovskis, P.M., Rimes, K.A., Warwick, H.M.C. & Clark, D.M (2002) The *Health Anxiety Inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. Psychological Medicine, 32,* 843-853.

Shear, M. K., Brown, T. A., Barlow, D. H., Money, R., Sholomskas, D.E., Woods, S. W., Gorman, J. M., and Papp, L. A. (1997). "Multicenter collaborative panic disorder severity scale.". *The American Journal of Psychiatry, 154,* 1571–5.

Spitzer, R. L., Kroenke, R., Williams, S. B., & Lowe, B. (2006). A brief measure of assessing generalised anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166,* 1092-1097.

Weiss, D.S. (2007). The Impact of Event Scale: Revised. In J.P. Wilson & C.S. Tang (Eds.), Cross-cultural assessment of psychological trauma and PTSD (pp. 219-238). New York: Springer