

Improving Access to Psychological Therapies

A review of the progress made by sites in the first roll-out year

Gyles Glover

Michelle Webb

Felicity Evison

North East Public Health Observatory

July 2010

IAPT Programme Foreword

In spring 2009, the Improving Access to Psychological Therapies (IAPT) programme commissioned a detailed analysis of the first wave (2008/09) IAPT implementation sites in order to evaluate whether the commitments relating to accessibility, the provision of NICE-approved therapies and detailed outcome monitoring were progressing appropriately. Specifically, the Review was asked to consider the:

- Equity of provision in relation to geographic, gender, age, ethnicity, range of disorders, language and disability coverage of the new services
- Profile of therapy types provided, including the pattern and length of interventions and the frequency of multi-step interventions; and the relationship of these to presenting problems, staff grades, medication usage, outcomes (clinical symptomatic, work and social).

North East Public Health Observatory's report enables the programme to examine how effectively these broader policy aims have been achieved.

It is clear that while there has been a high degree of success in rolling services out across the country, more needs to be done by local commissioners and services to ensure that those services are provided equitably and to the similar standards of quality.

Equality Impact Assessments (EqIA) focus on groups stemming from the existing legislation in the United Kingdom that covers discrimination. The groups and target areas include; age, sexuality, faith or belief, race, ethnicity, disability and gender. The groups are not homogeneous and people within these groups have different and individual needs.

In addition, it is also clear that services need to address the low referral and treatment volumes for some of the anxiety disorders to ensure that the scope of the service covers the range of need, which IAPT is committed to addressing.

Consequently, the issues highlighted in this report will form the basis of a full IAPT Equalities Review in 2010. Both this report and the Equalities Review will serve to underpin the publication of updated IAPT Commissioning Guidance in Autumn 2010.

Looking ahead, the report demonstrates that further analysis of the database will be of great value in informing future developments in the quality of care provided in IAPT services and informing the commissioning of such services. Therefore, additional analysis of this data has now been commissioned by the IAPT Programme and will be published later in 2010.

Contents

Abstract	1
Introduction	5
Background	5
Design and methods	7
Results	9
<i>Overview</i>	9
<i>Overall flow of patients</i>	10
<i>Characteristics of study group</i>	12
<i>Ethnic profile</i>	12
<i>Disability</i>	13
<i>Sources of Referral</i>	14
<i>Work and benefit status</i>	14
<i>Diagnosis</i>	15
<i>Initial symptom ratings</i>	16
<i>Variations in symptom and W&SAS ratings</i>	18
<i>Symptom and W&SAS scores as continuously distributed variables</i>	19
<i>Treatment</i>	20
<i>Treatment endings</i>	22
<i>Numbers of sessions</i>	22
<i>Diagnosis and choice of treatment</i>	23
<i>Which patients received higher intensity treatments?</i>	25
<i>How successful was treatment?</i>	27
<i>Associations between outcomes and service level factors</i>	32
<i>Outcome in relation to the phobia questions</i>	33
Discussion	34
<i>Data collection</i>	34
<i>Usability of the rating scales</i>	35
<i>Assignment of diagnoses</i>	36
<i>Patients receiving assessment and treatment</i>	36
<i>Patterns of treatment received</i>	37
<i>Outcomes</i>	37
<i>Equality issues</i>	39

Conclusion	40
References	41
Tables	42
Annex 1. Dataset finally requested from providers	121

Abstract

Background

Improving Access to Psychological Therapies (IAPT) is a programme designed to make psychological therapies for common mental health problems widely available. Following pilot work in Doncaster and Newham, the first wave of roll-out sites started operation in 2008. We report a study of their work in the year from October 2008 to September 2009. Allowing for varying start times, this was roughly their first full year of operation.

Sources and methods

Detailed monitoring was an integral part of the programme from the outset, and sites were required to collect patient-level data about their clinical work in accordance with the IAPT minimum dataset [2]. We arranged to collect an anonymised, patient level extract of this from all sites operational in the relevant year in the late autumn of 2009. Extracts were collated into a single database and analysed in the North East Public Health Observatory. Our aims were to report on the characteristics of the patients and their care from the perspectives of equity, the patterns of treatment provide and its outcome.

Data received

We received datasets from 32 sites, in one case in two separate sections. We had contact with two further sites which failed to produce data. Data collection proved reasonably successful in providers where one of the two on-line NHS-net based data gathering services was being used. In some other sites it proved more problematic as they had not used the standard coding frames specified for the minimum dataset.

Out of 137,285 total records of referrals in the relevant period, data submitted to us indicated that 79,310 had an initial assessment. These patients formed our study group and our descriptions of patients referred to the service relate to them. For 47.4% of these, it would have been premature to report on treatment and outcome patterns as they were still in contact with service. 52.6% had terminated their episodes, 18.8% after only one contact, making symptom change unmeasurable. 33.8% had terminated after more than one contact. This group formed the basis for our outcome analyses.

Characteristics of referred patients

95% of the study group were of working age, 64% were women and 87% were White British. The services seem initially to have been more accessible to younger people, particularly women. People from Asian and Black ethnic groups were under-represented, as were men from minority white ethnic groups. Disability was very poorly reported, with only six sites completing this aspect of the dataset to a usable extent. 84% of patients were referred by their GP, 8.6% self-referred and 2.8% were referred by another agency, commonly other health services. Referral patterns varied widely between sites.

Work and benefit status were reported for most patients, but there was wide misunderstanding of the nature of Statutory Sick Pay, suggesting it would be unwise to report this separately from other social security benefits. For analysis patients were thus simply categorised as being on any benefits or none. 78% of patients at initial assessment were economically active and 54% were in full or part-time employment. 30% reported receiving some type of benefit.

Diagnoses were relatively poorly reported with only 54% of patients having a usable diagnosis. The commonest reported primary conditions were depressive episode (29%), mixed anxiety and depressive disorder (29%), generalised anxiety disorder (18%) and recurrent depressive disorder (7%). OCD, PTSD and family loss each accounted for 2%. Examination of the rating scales suggested that the diagnosis of mixed anxiety and depressive disorder (described in ICD10 as a sub-syndromal condition in which patients have significant symptoms of both conditions but do not reach either case threshold) was probably being commonly used when diagnoses of a depressive episode and/or anxiety disorder would probably have been more appropriate. The importance of this is that NICE guidelines on evidence based treatment approaches relate to ICD10 categories and there is no guideline for the mixed disorder.

Three established symptom measures were used: the PHQ-9 for depressive symptoms, the GAD-7 for anxiety symptoms and the Work and Social Adjustment Scale (W&SAS). In addition to these a new set of three questions asking about the presence of social phobia, agoraphobia and specific phobias was used. Completeness of these ratings was good, with 93% of study group members having complete ratings on PHQ-9, GAD-7 and the phobia questions. Two sites only implemented the W&SAS to a very limited extent, otherwise it too was completed by over 90% of patients at initial assessment. We explored the question of the sensitivity, specificity and positive and negative predictive power of the new phobia questions to the extent this was possible with the other data available. They appeared to perform very poorly, however against this it must be added that we were only able to test them against a diagnostic yardstick we knew to be of low quality. Similarly the two symptom rating scales were poor measures as the questions were devised specifically to identify patients with phobic symptoms not apparent on these.

72% of patients scored at or above the PHQ-9 threshold for depression, 77% at or above the GAD-7 threshold for anxiety disorder and 53% positively on one or more of the phobia questions. 84% were at case level on either of the first two scales – the measure used in pilot work, and on which programme targets are based, 88% were positive on a wider measure of caseness on any of these three. Using the categorisations of Mundt et al [3], 40.7% were significantly functionally impaired on the W&SAS; 35.6% had moderately severe problems or worse. In the Doncaster and Newham pilot sites, slightly more (90%) of the patients taken on for treatment scored at case level on either the PHQ-9 or the GAD-7. Symptom ratings were similar for the two sexes, slightly lower for people over 65. Depressive, but not anxiety symptoms were more severe in Mixed, Asian and Black ethnic groups.

Treatment

Data about the treatments received by patients was contained in three separate counts of numbers of attendances: by purpose of attendance, by type of intervention given and by the employment grade of the therapist seen. The results need to be taken with some caution as totals of these counts differed in unaccountable ways.

Sites differed widely in the proportion of sessions delivered by low intensity workers (Agenda For Change grades 1 to 5) and high intensity therapists (Agenda For Change grades 6 and above) staff. The median pattern was 45% low intensity / 55% high intensity, but the inter decile range was from 25% to 93% high intensity. There were also wide variations in the breakdown of the low and high intensity groupings themselves. The most common form of low intensity treatment was guided self

help, followed (in order of number of patient-sessions delivered) by pure self help, psycho educational groups, behavioural activation, computerised CBT and structured exercise. The commonest form of high intensity treatment was CBT, followed by counselling. Small amounts of Interpersonal therapy and Couple therapy were also provided.

Of the 41,724 patient who had terminated their contact episodes, 95% received some treatment. Of those treated, 61% received some low intensity treatment, 46% some high intensity treatment, and 19% both, however sites differed widely on these figures. The best marker available for patients who had 'stepped up' was those receiving both low and high intensity treatment (19% of those treated), but this overall figure conceals rates in four sites above 30% and in four below 2%; the median step up rate was 13%.

29% of patients receiving low intensity treatment had more than one type; sites appeared to have characteristic combinations. The employment grade of staff seen by patients matched the intervention types to a considerable extent. 38% of patients finished treatment episodes by being judged to have completed treatment, 22% dropped out, 9% declined treatment, 12% were judged unsuitable (mostly after treatment had started) and in 20% of cases this datum was missing. Again there were wide site variations in these patterns.

Reported numbers of treatment sessions per patient were surprisingly low in comparison to NICE recommendations, though this finding should be taken with caution as session counts seemed unreliable. While much treatment was in line with NICE guidelines, much was not.

Outcomes of treatment

At its simplest, a recovery rate can be defined as the proportion of patients with case level symptoms at the start who have lost them by termination. However it is necessary to consider also those with missing second symptom ratings and those who developed case level ratings on one or more scale during treatment not initially there. Using our most conservative approach, which considered all patients with initial scale ratings and two or more attendances, assuming those with missing second ratings to be unchanged, 37% of those with initial case level ratings on either PHQ-9 or GAD-7 did not have case level ratings on either at termination. Allowing for the development of new cases, the prevalence of case level symptoms fell by 29% during treatment. Outcomes were substantially better for patients completing their treatment (56% of cases recovering, net change in prevalence -44%). There was a small though statistically significant fall in benefit claimancy (-1.6%) but not in unemployment.

However the effectiveness of treatment varied substantially between sites. The 95% confidence intervals for recovery rates and net prevalence change figures, for the best and worst thirds of sites showed almost no overlap.

Discussion

Several aspects of data collection, and technical aspects of the rating scales were considered. Most pressing, the diagnostic code frame needs to be extended to include panic disorder, and more attention is needed on how reliable diagnoses are to be obtained. The phobia questions, as derived are not yet usable. However routine service roll-out evaluations (like the present study) are not an appropriate context for evaluative studies of new instruments, so if it is concluded that questions for this roll are required, research studies will be needed.

Important aspects which will need attention are the variations in site drop-out rates and clinical outcome figures. It will also be important to explore in depth the reasons for the significant differences in treatment rates for members of minority ethnic groups.

Introduction

This report describes a study of the clinical work undertaken by the first wave of teams set up under the Department of Health's Improving Access to Psychological Therapies (IAPT) programme. Following initial pilot work, the first roll-out phase teams were funded in the financial year 2008/9. Our study covers their work in the year from October 2008 to September 2009. Allowing for an initial start-up period, this is roughly the first full year in which a substantial number of the new teams were operational.

Background

The need for a much larger psychological therapies service was signalled by the National Director for Mental Health in his review of the National Service Framework 'Five years on'. Pilot work was undertaken in two services, Doncaster and Newham [4]. Following this, a national implementation plan for the programme was published early in 2008 [1]. This set out a description of the type of service envisaged in each PCT area. Box 1 sets out the details. In most cases these services would need to undertake a substantial amount of therapist training, building up to an appropriate complement of both high and low intensity therapists in a ratio of roughly 3:2. Roll-out to at least 20 new areas in 2008/9 was agreed for the first year, with full national roll-out to follow in the two succeeding years of the 2007 Comprehensive Spending Review period.

Box 1. Description of local services

- Stand alone team of therapists
- Referrals from GPs, as well as self-referrals
- Delivering NICE-compliant therapies at the level required
- In convenient settings in primary care or elsewhere in the community
- Supported by
 - employment advisors (with access to other relevant social supports, such as housing services),
 - GP advisor (to provide medical advice and liaise with other GPs) and
 - administrative staff.
- Size dependent on PCT population and level of need – for average need, 40 trained therapists for 250k population
- Preference for working as a single team, led by senior therapists
- Most therapy delivered close to people's homes – in GP surgeries, Jobcentres, or voluntary organisation premises.
- Support and some low-intensity therapy (guided selfhelp) delivered in part over the telephone.
- Central base for supervision sessions, some therapy sessions, record-keeping and administration.
- Expertise in employment, housing and benefits available to enable integrated service enabling people to return to normal functioning.
- Team members qualified in the therapy they are delivering.

Source [1]

Ongoing monitoring and evaluation of the programme was an integral component of the programme. A minimum dataset, recording the care provided to each service user alongside their

clinical progress, was developed. Implementation of this was mandatory for funded sites. And the programme as a whole was given three headline performance indicators:

- **PCT coverage** – at least 20 PCTs to implement IAPT services in 2008/09, this coverage to increase over 2009/10 and 2010/11
- **Building a skilled workforce** – training programmes to deliver 3,600 therapists by 2010/11 with an appropriate skill mix and supervision arrangements and
- **Extending access to NICE-compliant services** – 900,000 more people accessing treatment, with half of those who complete the programme moving to recovery and 25,000 fewer on sick pay and benefits, by 2010/11.

However it was apparent that it would take a year or more to organise routine national submission of the dataset. The present study was set up to examine progress in relation to the third, patient based goal, specifically to provide early feedback on how the programme was progressing.

In this context, the detailed aims for this study were to examine:

- The equity of provision in relation to geographical , gender, age, ethnic, language and disability coverage of the new services;
- The pattern of durations of interventions and the frequency of multi-step interventions;
- The profile of therapy types provided, and the relationship of these to presenting problems, staff grades, medication usage, outcomes (symptomatic, work and social);
- The emerging pattern of outcomes, including the variability of this within and between services, as services mature and as the large cohort of newly trained therapists gathers experience.

Design and methods

All of the new service providers were in regular contact with the unit in the Department of Health responsible for programme implementation. Details of the patient level dataset they were required to collect had been circulated. This was intended to support performance monitoring generally, and to anticipate the introduction of a full minimum dataset for periodic returns to the Information Centre. Initially this was circulated as the draft document, *Improving Access to Psychological Therapies Outcomes Toolkit*. A revised version was subsequently published as the programme's *Technical Guidance for IAPT Key Performance Indicators* [2]. Thus, in theory at least, there was an established standard set of data items, including patient demographic data, markers such as ethnicity, gender, age and disability, for issues relevant to equity of access, details of episode milestones, initial and repeat symptom and social functioning assessments and numbers of sessions with details of their broad content, specific treatment interventions and the characteristics of staff members seen.

During the summer of 2009, the arrangements for our study were developed in discussion between the programme director and his staff, the clinical advisors to the programme, the suppliers of information systems supporting collection of the minimum dataset in IAPT sites and the present authors. From these deliberations it was agreed that an anonymised, patient-level subset of the data, sufficient to inform about the issues forming the subject of the study, would be extracted by, or on behalf of all currently operational sites, in respect of the year October 2008 to September 2009. Details of the sub-set of data to be included are set out at annex 1. Each data item was explicitly referenced to the minimum dataset which provides definitions. No data item capable of identifying a patient was included, thus patients' names, NHS numbers, addresses, postcodes, and dates of birth were omitted and derived patient ages were requested. Details of the dataset were cleared with the secretariat of the National Information Governance Board for Health and Social Care.

Individual service providers were responsible for arranging production of a set of records comprising the agreed fields, and covering each case assessed or treated in the relevant period from their information system. These datasets were transmitted to the North East Public Health Observatory, usually as encrypted spreadsheets, through the NHSnet. On receipt, data files were checked to ensure no identifiable data had been included in error. A lengthy process of data cleaning and consolidation followed, at the end of which data from all sites were consolidated into a single master file. In this format, field codings were checked, made uniform and then extended to facilitate analysis. Analyses were undertaken with the data in a Microsoft Access database. Detailed tabulations were mainly done directly in SQL. Simple statistical testing was undertaken using custom written NEPHO statistical functions in Visual Basic. More complex statistical analyses were undertaken using STATA v10. Presentation and graphs were developed in Microsoft Excel. Confidence intervals for population based rates and for standardised service use ratios were calculated using Byar estimations of confidence intervals for counts. Confidence intervals for proportions were calculated using Wilson's method.

Multiple and logistic regression models were developed using forward stepwise selection methods, in most cases using a likelihood ratio method with a threshold of 10% significance for inclusion. Multiple regression models were calculated using Wald tests instead of likelihood ratio tests to provide robust standard errors because of heteroskedasticity. Hosmer-Lemeshow tests for

goodness of fit were undertaken on the logistic regressions. Dummy variables were created for primary diagnosis, gender, age group, ethnic group, referral method and site, using the most common category as the reference.

Results

Overview

Altogether, 32 IAPT sites data supplied data documenting a total of 138,541 episodes of clinical contact or near-contact. We were in communication with two further sites, but did not succeed in getting usable data from them. One site supplied data in two sections, reflecting a change of information gathering arrangements midway through the year. A further site showed clear evidence of a similar change in their data collection arrangements. Table SS1 (the first table in the site-specific data series) shows the type of information system used in each site. The majority of sites used one of two information system suppliers both of whom provided a service operating through the NHSnet, with data held on remote servers managed by the system suppliers. This had the advantage of making very few technical demands locally beyond the availability of ordinary desktop computers with NHSnet connections.

A large proportion of the records supplied to us lacked any assessment or treatment data, suggesting that individuals referred had been entered onto an administrative system, but had either not made contact with services, or not proceeded to have an initial assessment. Sites were asked to provide referral, assessment and treatment start dates, however not all did. We developed an operational 'start date', defined as the earliest of these. At least one date was available for all but 163 (0.01%) episodes. Start dates indicated that many sites had included data about episodes starting prior to the defined study window of 1st October 2008 to 30th September 2009. Seventeen sites reported a total of 1086 episodes (0.8%) with earlier start dates. Eighteen sites provided data on episodes with start dates up to the end of September 2009 (12 months data). Four included start dates only up to earlier September dates, nine to the last week of August 2009 (11 months data), and one to mid July (9.5 months data). Altogether 80,020 (58%) had at least one of the initial symptom ratings (PHQ-9, GAD-7 or phobia ratings).

We took the 79,310 patients, with a start date on or after 1st October 2008 and with at least some evidence of an initial assessment, as the principal study group. Seven episodes with clearly erroneous start dates in November and December 2009 (postdating data submission) were excluded. Table 1 summarises this.

Data were submitted to us in the form of a summary record for each patient. Table 2 shows the level of completeness of the main data items in the records describing study group patients. This study is intended to convey the consistency of the roll out across sites as well its overall achievements. This table introduces a format we have used extensively in subsequent tables for this purpose, which is to show, alongside the overall figures, percentile points from the distribution across sites. Seven percentile points are shown: the maximum and minimum figures, the median figure (in our case the mid point between the 16th and 17th sites in order), the 25th and 75th percentiles giving the boundaries of the middle half of the sites, known as the inter-quartile range (IQR), and the 10th and 90th percentiles. The last two, which effectively give the boundary of the top and bottom three sites are intended to show whether the actual maximum and minimum figures are outliers or reflect a continuous trend.

In most respects data completeness was remarkably good. Most of the fields reporting universal characteristics (expected to be completed for all patients) were present in over 90% of cases. Three areas were less complete. Fields reporting disabilities were completed only by a small number of sites, though reasonably completely in these. Ethnic categories and primary diagnoses were missing in 25.4% and 36% of patients respectively, these omissions representing generally poor reporting. Treatment and episode ending details were less completely reported and showed more inter-site variation. These are only expected for patients where treatment or and ending has occurred and, in part this reflected differences in the proportions of patients whose episodes had ended. Data on contact activity by treatment type and session purpose (assessment / treatment/ review etc) were notably more complete than data on activity by therapist grade. In most cases where completeness was less than excellent, a small number of sites accounted for a large proportion of the missing data.

We examined whether there were differences in completeness between the submissions from the two major information systems by performing Kruskal Wallis tests to compare the ranks for proportion of records with complete data in each field between the groups of sites served by each major type of information system. There were highly significant differences in completeness on nineteen out of thirty four major variables, sixteen favouring one system, three another. This would seem to suggest that the system used is important and that both major systems could be improved.

Two sites (one worse than the other) showed evidence of locally developed information systems not using standard minimum dataset coding frames. In these cases fields such as the primary diagnosis and source of referral appeared to have been entered by hand. We recoded these to the extent they aligned unambiguously with standard codes. However this was a laborious job and we would recommend that for future returns, sites be advised that improperly coded data will be reported as missing.

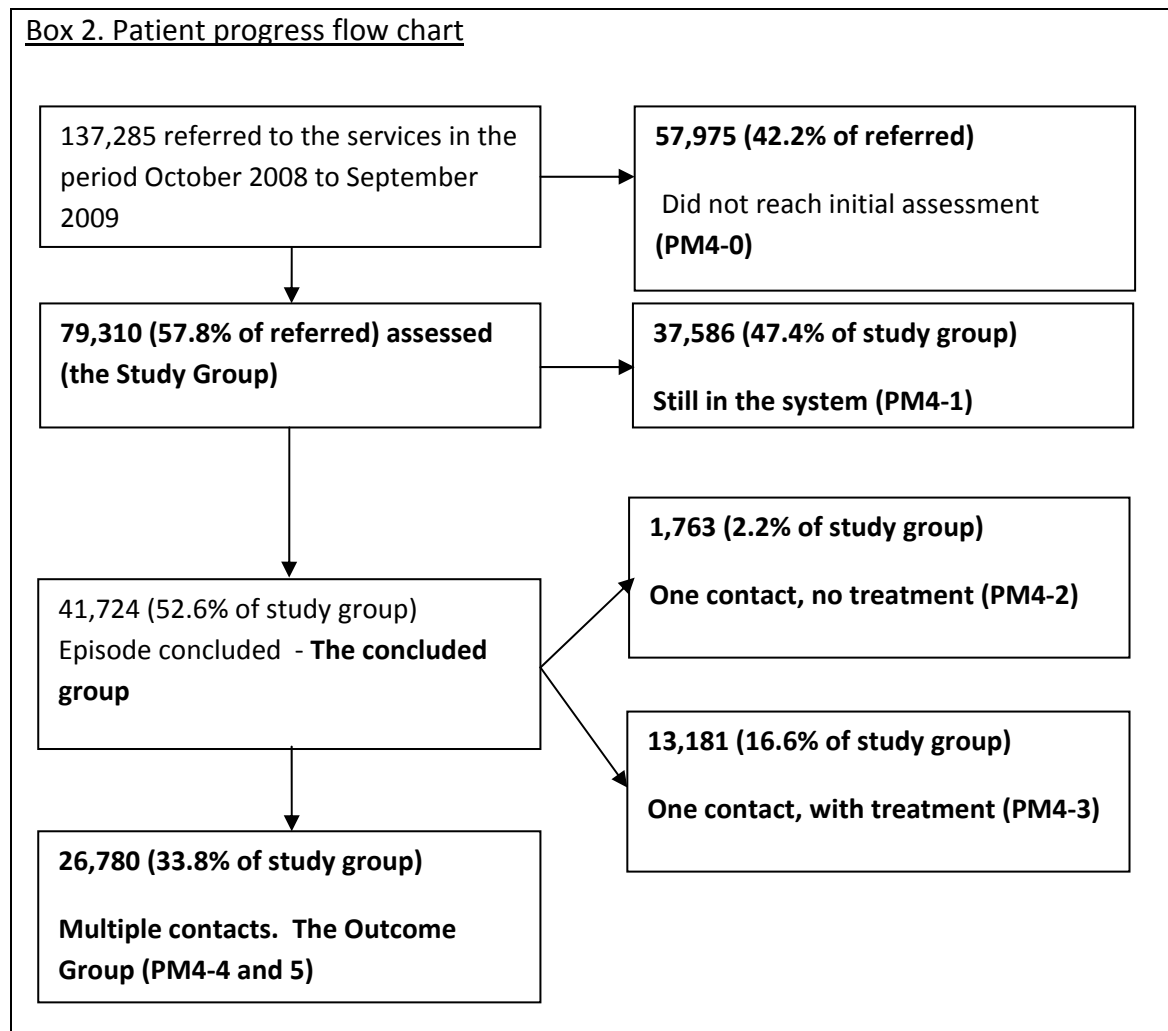
We considered further exploration of which patients were more or less likely to progress from preliminary contact to initial assessment. Table 3 gives an overview and table SS1 provides the detail for individual sites. We did not pursue this as it appeared likely that findings would primarily reflect the differing practices of sites in respect of the stage of assessment at which system-record were initiated. Site 32, for example had no record of individuals who had not gone on to have a full assessment, and sites 18, 22, 29 and 33 all had less than 5% in this category, while in eight sites less than half of those registered went on to full assessment.

Overall flow of patients

Clark and his colleagues [4], describing the pilot studies in Newham and Doncaster, were able to work sufficiently closely with the sites to provide a detailed narrative of the working patterns and care pathways in each. We were not able to work at this level of detail with the much larger number of sites in the wider roll-out and working retrospectively. However to some extent it was possible to identify the stages in care as documented by the data received. Box 2 summarises this simple perspective. Of the 137,285 total individuals with contact recorded in the appropriate period, , 79,310 had an initial assessment.

We can only satisfactorily analyse treatment and outcomes where episodes have finished. As far as our records showed, 47.4% of this initially assessed group were, still in the system. 52.6% had some evidence of having finished their episode of contact, in the form either of an ending date or a reason for termination.

Of those who had finished their episodes of care, 35.8% had had only one attendance, and 11.8% of these had no reported treatment. The remaining 64.2% (of patients with finished episodes) had been seen more than once, making analysis of their symptomatic progress potentially possible. 99.5% of these had received some treatment.



There was considerable variation between the sites in relation to the proportion of patients who had reached each stage. Details are set out in table S2. Charts 1 to 3 serve to illustrate different aspects of these data. Chart 1 shows the proportions of patients whose episode was concluded after more than one contact. This group is important since they are necessarily the group on whom outcome measurement is based. Numerically they range from none to 64.6% of the total cases in the 32 sites, with a median of 33.7% and an inter-quartile range (IQR) of 28.5% to 41.0%. Ignoring two obvious outliers, the range in the proportion of patients concluded and treated (irrespective of the number of contacts) is from 19.3% to 71.5% (median 51.3%, IQR 36.1% to 61.9%) (Chart 2), while the range in those simply concluded is 30.2% to 72.1% (median 52.6%, IQR 40.3% to 63.3%) (Chart 3).

Characteristics of study group

The next section describes the characteristics of all the study group patients. The intention is to provide an overall picture of the individuals being served.

Table 4 shows the age and gender breakdown of the study group. The apparently large number of patients (8%) with missing age groups arose as a result of one site submitting ages in an incompatible decennial grouping. Excluding this site, only 3% of patient's ages were missing.

1% of patients were aged under 18 and 4% 65 or older. Overall the gender ratio was just under two women to each man. This figure was higher at both extremes of age, and lowest, at 1.7, in the 35-44 age group. A separate analysis using the decennial grouping provided by the individual site concerned showed no difference in the age profile for men, but a significantly greater representation of older women (chi square = 22.47, df = 7, $p < 0.005$).

Several sites agreed to take a special interest in exploring relevant treatment models for particular groups. Older people, and children and young people were designated in this way. The one site designated for children and young people did not supply data. The proportion of patients aged 65 and over in the four sites designated for older people (Sites 7, 19, 29 and 30) was not significantly different from that in the rest (chi square values: women 2.03, men 0.43, df = 1 in each case).

In all but one of the sites, we were able to relate the service to the population of a single PCT (or in one case two PCTs combined), thus permitting calculation of population-based rates of service use. We calculated rates separately for women and men. Allowing for the one missing site, and a small amount of loss due to cases with their gender not reported, we can say that 73,984 cases arose from an overall population of 11.21 million, just over one fifth of the population of England. This gives a crude rate of 6.60 per 1000 (95% Confidence interval 6.55 to 6.65) for all ages. Detailed age specific rates for sites are shown in table SS.4 and SS.5, with age specific rates for all sites combined in table 5 and chart 4. Overall differences in rates for sites are probably not illuminating here as they most likely reflect mainly the accessibility of the service to the various parts of the PCTs served and the overall volume of staff available to provide a service, but with these caveats, differences in age and gender profile are interesting. Higher rates were seen for women, and for younger adults. Overall the rate for women was just under double the rate for men, but in younger age groups the differences was greater. Chart 5 shows the prevalence of any neurotic disorder by age and gender from the 2007 adult psychiatric morbidity survey for comparison [5]. The gender differences and the fall off with older age were both much sharper in the IAPT study group rate, suggesting that for similar levels of need, younger women were 'early adopters'. Gender ratios for specific age groups seldom differed significantly between sites. However these patterns were not consistent between sites: sites 6, 11 and 18 had a significantly more dominantly female caseload and sites 15 and 28 notably less so.

Ethnic profile

The ethnic profile of the study group patients is shown in table 6. This table divides patients into three broad age categories (18 to 34, 35 to 64 and 65 and over), and condenses ethnic categories

into the conventional six broad groupings¹ as numbers in many ethnic categories in the oldest age group are small. Site-specific table SS7 gives further details of case numbers and proportions for individual sites. The ethnic profile differed considerably between sites; this was to be expected as the populations they serve also differ. The question of how well the ethnic profile of patients reflects that of the areas from which they were drawn is examined more thoroughly in table 7 for broad ethnic groups, and in table 8 for each of the 16 ethnic groups of the 2001 census classification. Site specific table SS8 provides the same analysis for individual sites. This analysis uses the technique of indirect standardisation: the number of patients of each ethnic group is compared with the number expected, if the overall age- and sex-specific service use rates for each site had applied evenly across the all ethnic elements of the population they served. This approach allows for the fact that the sites provided different overall levels service to their respective population, and that they served populations with widely differing ethnic compositions; it simply tests the question whether, given this situation, the ethnic mix of patients in the study group suggested some degree of ethnic selection. In terms of the broad ethnic classification, people assigning themselves to one of the Asian, Black or 'Other' groups were significantly under-represented for both genders, minority white groups were also under-represented for men. It seems likely that these differences – in the ethnic profile simply of those receiving an initial assessment - reflect mainly patterns of referral to services.

Four sites were not included in this analysis. Site 25 was omitted because of uncertainty about the relationship of the service to current PCT boundaries. Sites 29, 32 and 33 were omitted because patients' ages were either not reported, or reported incompatibly.

Sites were asked to report on whether patients were 'able to communicate in spoken English'. Only seven sites produced reasonably complete data on this (usable answers in 50% or more of records). Amongst these, usable answers were provided overall for 80% of patients. 30 (out of 10,543 - 0.28%) were reported as being unable to. Of these, three gave their ethnic category as White British, two had ethnic categories of 'not stated' (they declined to say) and three had no entry. This left a rate of 2.1% (22/1047) of patients from minority ethnic groups unable to communicate in spoken English. Full data on reports of patients spoken English are provided in the site-specific table SS8.

Disability

Sites were asked to record whether each patient had any of a range of disabilities. Visual, speech and hearing disabilities were widely reported by only six sites; two for virtually all patients, the others on 86%, 76% 73% and 56%. The same sites, plus one other reported mobility disability extensively, with a further 10 sites recording the presence or absence of this for between 10 and 50% of patients.

In the six sites that reported, of those patients for whom data were provided, 2.5% had a visual disability, 1.0% a speech disability and 2.9% a hearing disability. Overall reporting completeness in this group was 75%. In the 7 sites reporting substantially on mobility, with 78% of records usably

¹ Minority white groups - White Irish, Any other White; Mixed groups - Mixed White and Black Caribbean, Mixed White and Black African, Mixed White and Asian and Any Other Mixed Background; Asian – Indian, Pakistani, Bangladeshi and Any Other Asian; Black groups - Black Caribbean, Black African and Any Other Black; Other – Chinese and Any other ethnic group.

coded, 5.8% of patients reported a disability. Full details by site are provided in site specific table SS9.

Sources of referral

Sources of referral were remarkably completely reported. Overall figures are shown in table 9. Of the 32 sites, half provided useably coded referral information for in excess of 99% of episodes; only five reported this for less than 95%, the weakest figure being 82.6%. Overall 83.6% of patients were referred by their general practitioner and 8.6% referred themselves. 2.8% were referred by another clinical specialty and 1.3% by local community health services. No other identified source accounted for as much as 1%.

However patterns of referral varied substantially between sites. Table 9 also gives some impression of this by showing percentile points on the distributions of the proportion referred by each type of source between sites. Thus in 5% of sites (at or above the 95th percentile), one third or more of patients were self referred, the maximum being 99.8%. By contrast, for half the sites, the proportion self-referred was 2.1% or less (the 'median' figure). 'Other clinical specialties' and community health staff also stood out as groups which while insignificant in the group as a whole, were major referral channels in a small number of sites. Full site specific details are given in table SS10.

The Newham site in the pilot study found that self-referral was particularly relevant for members of minority ethnic groups [4]. In our own earlier study for the Delivering Race Equality Dashboard we found that while rates of referral of minority ethnic group members to IAPT services were significantly lower than those for White British, rates of starting treatment were not [6]. We explored this here, comparing the proportion of these patients referring themselves with that for White British patients. Overall there was a highly significant difference. 9.5% of White British patients had referred themselves whilst the corresponding proportion for minority ethnic group patients was 17.1% (Chi Square = 252.5, df = 1, p<0.0001). However this was not a consistent finding. Amongst the eleven sites where the overall proportion of patients self-referring exceeded 5%, there was a significant difference in only four (sites 11, 12, 23 and 31).

Work and benefit status

96.9% of study group members had valid ratings for their initial employment status. Rating of Statutory Sick Pay (SSP) commonly appeared anomalous. This benefit is payable to people who are employed (full or part time) and who are unable to work because they are ill. It is administered through their employer. However 23% of those reported as receiving SSP were also reported as not in employment. This anomaly was found in a substantial proportion of SSP cases in all sites, suggesting that the distinction between SSP and other social security benefits was not clear to respondents. For this reason we developed wider markers of 'any income support benefits' at first and last assessment which was marked True if the patient reported receiving either SSP or any benefits, False if they positively reported not receiving either, and unknown in other cases. 94% of cases had a known rating for this characteristic.

The overall pattern for employment and benefit status of study group members is shown in table 10. Overall 77.5% were economically active and 53.7% in employment. 23.8% were unemployed and 19.3% economically inactive. Just under 30% were receiving benefits of some kind. The percentile points for the distribution between sites on these figures shows that ranges were, in most cases, not great. Full site level details are provided in table SS11.

Diagnosis

Primary diagnosis, though an important aspect of the IAPT work because it constitutes the framework for NICE guidance about evidence-based treatments, was one of the less completely reported aspects of the dataset we received. Overall, a usable diagnosis was reported for 54.4% of episodes, the proportion varying between sites with a median of 64% (IQR 26.2% to 80.8%.) This is lower than the proportion with 'valid' diagnoses given in table 2, as a substantial proportion

The overall pattern for the sites is shown in table 11. The most common diagnoses were depressive episode and mixed anxiety and depressive disorder, each accounting for just under 30% of cases with a usable diagnosis. Generalised anxiety disorder was the only other category accounting for more than 10% of cases. Site specific figures are given in table SS.12 and illustrated in chart 5a. There was considerable variation between sites in the diagnostic profile, as shown by the percentile columns in table 11. Phobic states, whilst accounting for only 4.2% of cases overall, were concentrated in a small number of sites, comprising a third of the caseload in one, a quarter in a second and over 10% in two more. Three sites produced substantial numbers of text diagnoses which either could not be accommodated in the coding frame produced for the minimum dataset and used by the other sites, or could be assigned to more than one category. Box 3 shows the details.

Box 3. Reported diagnoses assigned to the category 'other specified mental disorder' from three sites. Explanatory notes in parentheses have been added.

Diagnosis	Records
F40 (Phobic anxiety disorder not otherwise specified)	19
F41 (Other anxiety disorder not otherwise specified)	33
F43 (Reaction to severe stress, and adjustment disorders not otherwise specified)	5
Anxiety	1096
Behavioural problem	116
Drug Misuse	12
Personality disorder	4
Acute Stress Reaction	97
Panic	31

The large number of cases diagnosed as 'mixed anxiety depressive disorder' reflects the diagnostic problem arising from the frequency with which these two conditions co-occur and the rather complex position in the current classificatory systems [7]. Officially, the phrase describes a 'sub-syndromal' condition in which there are symptoms characteristic of both anxiety and depression, but not sufficient of either to establish the diagnoses in its own right. This is the sense in which it is used by the authors of the UK national psychiatric morbidity survey programme [8]. However, examination of the scores on the GAD-7 and PHQ-9 rating scales (described more fully in the next section) suggest that usage of the term here is wider. Of the 12,627 patients assigned the diagnosis, 73.1% scored at or above these rating scales' thresholds for both anxiety and depression, 9.0% for

anxiety alone and 5.3% for depression alone. Only 11.3 % scored below threshold levels on both. Charts 5b and 5c show the distribution of PHQ-9 and GAD-7 scores respectively for patients with each of the four relevant categories as a primary diagnosis. The profile of PHQ-9 scores for patients diagnosed with depressive disorder and mixed anxiety depressive disorder are indistinguishable. Below the threshold score of 8, the profile of GAD-7 scores for patients with generalised anxiety disorder and mixed anxiety depressive disorder are very close. However more patients with GAD-7 scores at or above this threshold are categorised as having the mixed diagnosis or a depressive episode than generalised anxiety disorder. Secondary diagnoses do not seem to eliminate this issue. Table 11a shows overall patterns of secondary diagnoses for patients with a primary diagnosis of a depressive or generalised anxiety condition. Only 3.5% of these people overall had any secondary diagnosis and the patterns of GAD-7 and PHQ-9 scores in the combined situations suggest no obvious pattern. Site specific table SS12a shows at site level the mean initial PHQ-9 and GAD 7 scores for patients grouped by combinations of diagnoses, whether these were assigned as primary or secondary. The depression group includes both depressive disorder and recurrent depressive disorder; the anxiety group includes just generalised anxiety disorder. Patients with neither of these diagnoses but with a diagnosis of mixed anxiety depressive disorder are shown as 'MADD alone'. Overall this group is indistinguishable in scale scores from the (very much smaller) group with both types of diagnosis independently; in some sites the independent diagnoses group have higher scores, in others, lower. Table 11b shows a more detailed presentation of symptom scores for patients assigned to these four diagnostic categories for the study group as a whole, including ratings for the three phobia questions separately. The clear impression is that the group assigned the diagnosis mixed anxiety depressive disorder have, scores similar to those with diagnoses of depression and, for the GAD-7 and PHQ-9, higher than those with generalised anxiety disorder.

Initial symptom ratings

Table 2 shows the overall completeness of initial symptom ratings in study group members, table 12, and the site specific table SS13 explore this in more depth. 93.0% of study group members had complete initial ratings for all three symptom scales. The most frequent omission was the phobia scores, the refinement added to the rating protocol following the pilot study and intended to identify individuals with clinically significant phobias who may have failed to reach caseness on the GAD-7 rating for anxiety symptoms. The percentile figures, and the site specific table (table SS13) show that the majority of sites provided overall completeness levels above 95% on all ratings, with three quarters of sites providing all four ratings on more than 90% of study group patients. Initial Work and Social Adjustment Scale (W&SAS) scores were produced for slightly more of the study group members than the phobia scores overall. Unlike the phobia scores, where several sites performed less well, underperformance on W&SAS completeness was largely confined to two sites.

The phobia scales are new; designed specifically for the IAPT programme. There was thus some interest in the extent to which they appear to be sensitive and specific to the diagnoses they are intended to reflect. Tables 12a to 12c show numbers of patients, grouped by primary diagnosis, with mean scores and proportions reaching the case threshold (four) on the three phobia questions, for social phobia, agoraphobia and specific phobias respectively. Table 12d shows sensitivity, specificity and positive and negative predictive value calculations² for all possible cut points on the phobia

² The sensitivity and specificity of a test are the proportions of cases and non-cases it correctly identifies; the positive and negative predictive values are the probability that someone scoring positive has the condition and

questions. There is no cut point for any of the tests where sensitivity and specificity figures are both adequate. To be effective as a screening instrument, sensitivity should be above 80% and specificity higher (to avoid very high overall misclassification rates where the prevalence of the condition is low). These results need to be interpreted cautiously. First it seems unlikely that the diagnostic yardstick against which performance was being measured was reliable. It was certainly not based on a standardised process. Second, the point of the phobia questions in this context is not screening. They were introduced as a way to identify caseness, by a short standardised question instrument, in individuals with phobic conditions who failed to reach threshold scores on either the GAD-7 or the PHQ-9. In terms of their use at initial assessment, the key question for each is the proportion of patients assigned the relevant primary diagnosis, who do not reach caseness on the two main scales but who are correctly classified by the phobia question. Table 12e shows the result. Half of patients with primary diagnoses of social- or agoraphobia, not reaching caseness on the two main scales were identified by the corresponding phobia question. The specific phobias question was a little more successful, identifying 60%. However if the questions are also to be used as tests of outcomes, the issue of whether they can accurately identify whether patients do not have symptoms is also crucial.

The apparently poor specificity (if the yardstick of the assigned diagnoses is to be believed) for all three questions, combined with the low prevalence of the conditions they are designed to identify in the current patient group, and hopefully in any post treatment group, gives them exceptionally weak positive predictive values in this context. This would be likely to give spuriously low recovery rates on measures using them. We have included some results using them in this report for the sake of completeness. However these should be looked on very cautiously and we would recommend concentration on measures based on the more established instruments.

Were study group members 'cases'?

The simplest way to interpret all these instruments is by rating whether individuals reached threshold scores conventionally taken to indicate caseness (8 for the GAD-7 and 10 for the PHQ-9). For the three IAPT Phobia questions (designed for this programme) a positive answer to any (a score of 4 from a possible range of 0 to 8 on any one of the questions) was considered to indicate caseness. Mundt et al [3] concluded that scores above 20 on the scale reflected moderately severe or worse psychopathology, scores between 10 and 20 indicated significant functional impairment, while scores below 10 appeared in sub-clinical populations. Tables 12 and SS13 show the results. Of those patients with usable ratings in each case, 77.3% of patients scored at case level for anxiety symptoms, 72.3% for depressive symptoms, and 53.3% for phobic symptoms. 40.7% were significantly functionally impaired on the W&SAS; 35.6% had moderately severe problems or worse.

Two composite symptom measures were introduced here which will be used later to measure 'movement towards recovery'. MTR2 uses all three of the symptom scores. A person was counted as a MTR2 case if they scored at or above the caseness threshold on any one of them at initial assessment. 88.1% of the study group patients came into this category. MTR1 used only the GAD-7 and PHQ-9 scores. This allows comparison with the Doncaster / Newham pilot study, and is a more reliable measure in view of the concerns about the specificity of the phobia questions set out above.

that someone scoring negative does not. Sensitivity and specificity are properties of the test, PPV and NPV are also dependent on the prevalence of the condition in the population being tested.

Overall, 83.6% of patients reached MTR1 case level at initial assessment, lower than the pilot study findings of 90% in Doncaster and 86% in Newham. The percentile columns in table 12 and the corresponding site specific figures show that in this case the overall figure was a good reflection of the pattern between sites; the Newham figure was just below the 90th percentile of our sites, while the Doncaster figure was matched by only one site.

The wider MTR2 measure of initial caseness included 88.1% of study group patients overall. Site figures were fairly tightly clustered around a median of 88.0% (IQR 86.4% to 89.6%). There was one notable low outlier, site 30 at 68.4%. The converse of this is that 11.9% of study group patients overall had no case level rating, with a site median figure of 12.0% (IQR 10.5% to 13.5%).

Most patients (73.2%) reached case level ratings on more than one of the scales. This is to be expected as Depression and Anxiety symptoms commonly co-exist, and anxiety states and phobic states differ primarily in the extent to which the symptoms are related to specific situations. Table 13 shows the overall pattern of co-occurrence of case-level ratings. This can be compared to the analysis of psychiatric co-morbidity in the general population undertaken by the authors of the recent national psychiatric morbidity [ref – chapter 12]. These are not strictly comparable, since the latter used different rating instruments (the CIS-R for the conditions studied here) and included a larger number of conditions in its scope. However, table 12.1 of their report indicates that of the 23% of the general population with any psychiatric condition, 69% had one, 19% two and only 12% three or more diagnosable disorders. If reaching the rating thresholds on each of our instruments can be considered comparable to this, the extent of co-morbidity is much greater, with only 17%, of those having any threshold condition (from amongst the more restricted range) having only one, while 36% had two and 47% all three. This frequency of co-morbid conditions was not, however reflected in the assignment of secondary diagnoses. 42,212 study group members had a usable primary diagnosis of a mental disorder. Ignoring 94 illegal assignments (combinations of depressive or anxiety disorders with mixed anxiety depressive disorder), only 2,300 (5%) had some usable secondary diagnosis of a mental disorder.

Variations in symptom and W&SAS ratings

The proportion with case level ratings on each of the three instruments varied significantly with both age and gender. In the case of gender (tables 15 and 16) the differences were not substantial, though statistically significant because of the large numbers. Women were less likely than men to have no case level rating and more likely to have three. By contrast, a slightly higher proportion of men had severe problems on the W&SAS, though in the broader category of moderate or severe problems there was no difference.

Substantially fewer patients aged 65 and over reached case-level on each of the instruments and the proportion in this age group with no case-level rating was twice that seen for other age groups. Older patients also had substantially fewer diagnoses: 43% of those aged 65 and over had 0 or 1 case level rating compared to 27% of those under 65 (Tables 17 and 18). This is important as it suggests a reason why overall (as opposed to condition specific) outcome measures may look more favourable for older people groups. Three sites (29, 32 and 33) which did not provide age grouping details are not included in the age group analysis. W&SAS scores followed the same pattern with a substantially smaller proportion over 65, and a notably smaller proportion of under 18s in both the moderately and severely disabled categories.

Symptom patterns across the broad ethnic categories are shown in table 19. Substantially higher proportions of people giving their ethnic group as Mixed, Asian or Black had threshold levels of depressive symptoms on the PHQ-9, while all minority groups scored substantially higher than White British on the Phobia questions. Anxiety symptom scores showed relatively little difference. The proportions scoring at caseness for either depression or anxiety (the MTR1 test) followed the depression pattern, being higher for Mixed, Asian and Black people, but the differences were only slight in the MTR2 measure, suggesting that members of minority groups who were MTR2 cases had higher level of multiple case-level ratings than White British. This is reflected in the difference in numbers of conditions experienced by patients shown in table 20. Higher proportions of all the minority groups reached case level on all three rating instruments. As with the difference between older and younger patients, this is a potential reason why overall outcome measures may look less favourable for members of minority groups. On the W&SAS, substantially higher proportions of all minority ethnic groups scored as severely disabled. To the extent that higher figures were seen in the combined moderately and severely disabled grouping, this reflected severe cases; proportions in the moderate category itself were smaller for all minorities than for the White British group.

Table 21 shows the proportions of patients scoring at case level broken down by source of referral. While the differences between referral groups were statistically significant, because of the large number of cases, in most cases they were not substantial. There were two exceptions to this. The proportion of people referred from the range of 'other' sources scoring positively on the phobia questions was 13% higher than that among people referred by their GPs or themselves. And the proportion of individuals having no case level rating was 17% greater amongst those who referred themselves than in other people. However this last finding appears to have arisen largely from one site (site 30). Omitting this site from the analysis, only 12.1% of self referred patients failed to reach case-level on at least one rating. Table 22 shows that overall, a slightly higher proportion of self referred patients had no case level ratings, while a substantially higher proportion of those referred from 'other' sources – mainly from other clinical units – had case-level ratings on all three. W&SAS ratings showed higher proportions of all patients self referred and referred by 'other sources' had severe levels of disability.

Symptom and W&SAS scores as continuously distributed variables

Instead of using them to categorise patients on the basis of whether they reach specified cut-points, the three symptom scales and the W&SAS can be treated as continuous scores. While not strictly continuous (since they have fixed minima and maxima and can only take integer values) this is an attractive option as it may allow a more finely tuned set of analyses. In considering this approach, it is first necessary to see how scores on each are distributed. This is shown in charts 6 to 9.

None of these follows the statistical normal distribution. The GAD-7 and Phobia scores are very strongly skewed and the GAD-7 and W&SAS show truncation at one end of the distribution. The phobia score, a total of patients ratings on three questions, shows a distinct sawtooth pattern. This arose not from concordance between two scales, but from an apparent even number preference in the answers to all three questions in almost all sites. This may have arisen from the way the scales were presented, with 9 possible scores (0 to 8 inclusive) and 5 textual commentary points, possibly guiding respondents to the points 0, 2, 4, 6 and 8. These characteristics suggested that the most appropriate ways to reflect the scoring patterns for sub groups would be with medians and percentile distributions rather than means and standard deviations. Tables 23 to 26 show the

distributional patterns of all study group members on the four scales respectively, broken into the gender, age, ethnic and referral groups used above. Differences between groups are tested using Kruskal-Wallis tests. The findings are similar to those using the cut-points approach. Detailed profiles of the study group patients of each site, on each scale, are provided in site specific table SS14.

Treatment

The picture of treatment provided to the study group patients is described first. This is intended to convey an overall impression of the amount and nature of work being undertaken by the sites. Subsequent sections, describing the treatments received by patients, are confined to the patterns of treatment received by patients whose episodes have been completed. Inevitably these do not show the full amount of treatment reported, as they omit work to date with the 47.4% of study group patients still in treatment.

The dataset we received provided three types of information about the contact activity of, and treatment received by each patient. The first reported the number of sessions each patient had by the staff grade of the therapist seen or contacted, coded as 'agenda for change' grades 1 to 8d inclusive. A second gave a count of the number of sessions by purpose of session (assessment, treatment, review, follow up and reasonable combinations of these). The third provided details of the type of intervention provided in treatment sessions. Ten types of specified treatment intervention were counted and a further count provided for 'other' types of intervention. Overall figures for these counts are shown in table 27 with site specific tables in SS15. They vary considerably. Some differences can be anticipated: for example a number of contacts will be for assessment or review only and will thus not appear in the counts by type of therapy. However even allowing for this (the table shows a tally of counts by purpose where the purpose includes treatment) the differences are substantial and differ between sites. Generally, counts by types of therapy are higher than counts of sessions where therapy forms all or part of the purpose. This could be explained for the low intensity therapies (see below) if some sessions were considered to include more than one type of treatment. To clarify this issue further the site specific table presents ratios of number of sessions counted by treatment type to number by purpose where the purpose includes treatment for three subgroups of patients. The overall ratio for patients having only low intensity treatments was 1.8, for those having only high intensity treatments it was 1.36 and for those having both it was 1.78. Session counts by purpose were substantially higher than those by therapist grade in all 70% of sites – overall by 33%. There is no obvious legitimate explanation for this in terms of the scope of the questions. Taken together these observations suggested that the session count by purpose of session was probably the most satisfactory of the three as a measure of the number of attendances by patients.

The profile of session counts by staff grade overall is shown in table 28, with parallel site specific detail in table SS16. Chart 10 shows the figures graphically. As a rough categorisation, staff employed at 'Agenda for Change' grades 1 to 5 were classified as having low intensity skills, as opposed to 'high intensity skills' staff at grade 6 and above. The most striking thing about these figures is the wide diversity in staff profile between sites. The median pattern was 44.9% low intensity skills, 55.1% high intensity skills. However the IQR for the proportion at low intensity skill levels was 36.3% to 56.7%; the extremes for the highly skilled fraction were from 13.6% to 100%.

Within these two broad groupings there were also wide ranges in the proportions at grades 4 and 5 and at grades 6 and 7.

A similarly wide variation was seen in the types of treatment provided. The ten types of approach reported separately were broadly categorised into high and low intensity groups comprising, respectively CBT, counselling, IPT and Couple therapy (high) and computerised CBT, Guided self help, pure self help, structured exercise, behavioural activation and psycho-educational groups (low). Table 29 shows overall figures, with site specific figures in table SS17. The most commonly provided form of low intensity care appears to be guided self help. The two most commonly provided forms of high intensity therapy were CBT and counselling, with roughly twice as much of the former. High intensity session numbers are shown as proportions of the total volume of high intensity sessions; this is reasonable as their nature precludes combining approaches in a single session. Some forms of treatment, including structured exercise amongst the low intensity forms and IPT amongst the high intensity forms were provided in substantial quantity in only a few sites.

Patterns of treatment received by patients

41,724 patients in the study group had completed their episodes of contact. For 39,819 (95.4%) this included at least some treatment. Table 30 (and site specific table SS18) provide an overview of the levels of treatment provided. Overall 60.8% received some low intensity, and 46% some high intensity intervention. There was some overlap: 18.5% of patients received both. These patterns varied considerably between sites, as chart 11 illustrates. Receiving both low and high intensity care can be seen as taking a therapeutic step up (in the terms of the stepped care model). The median site figure for this was 13.1%, IQR 6.4% to 17.8%.

Within the group receiving low intensity therapy, 71% of patients had a single type, 22.5% two types and 6% three types or more. The frequency of different pair wise combinations is shown in table 31. Several sites seemed provide guided or pure self help approaches to the great majority (over 80%) of their clients. The most common combination was pure self help with guided self help. Twelve sites had more than 100 such patients. Three other sites had substantial numbers of cases in the second most common combination, pure self help with psycho educational groups. The patterns of combination seemed to suggest distinctive site approaches (site specific table SS19).

Sites also varied considerably in the patterns of high intensity therapies patients received. Of the 18,308 patients with finished episodes receiving some high intensity therapy, 57.8% received CBT and 50.1% counselling, with 1.3% and 0.6 % receiving IPT and couple therapy respectively. As shown in Site Specific table SS20, and chart 12 there were wide variations between sites and 8.7% of these patients were reported as receiving both CBT and counselling.

There was some degree of match in the intensity level of the therapist(s) seen and the type of intervention. Thus, overall, 32.8% of those with finished episodes who received at least one of the coded treatment types had at least one appointment with a therapist at AFC grade 6 or above. However this was the case for 58.3% of those having only high intensity treatments but only 15.0% of those having only low intensity interventions. For those having both, the proportion was 35.8%. A degree of overlap here is to be expected, as some of those currently employed at grade 5 would have been trainee high intensity therapists. A fuller exploration of this issue would require more extensive access to the data sources than we had.

Treatment endings

Chart 13 and site specific table 21 show the pattern of endings for study group patients with finished episodes. Overall, only 38% were reported as having finished as a result of the treatment being judged to be complete. 21.6% finished with the patient dropping out, 8.6% with the patient declining treatment. In 11.6% of cases the patient was judged not suitable, though 86% of these had some treatment reported. For 20.4% of cases the method of ending was unknown or uninterpretable coded. Sites differed widely in these figures; ignoring five sites with more than 20% of data missing, median proportions were: for completing treatment 45% (IQR 38.4% to 54.4%), for dropping out 22.9% (IQR 18.8% to 29.0%), for declining 9.6% (IQR 6.3% to 12.9%) and for appearing unsuitable 10.2% (IQR 7.4% to 18%).

Because of its significance for outcomes (described below), we undertook a multivariate analysis of factors associated with a terminations because the patient was considered unsuitable, or declined or dropped out. Detailed results are in table 31a. As with later models, the actual amount of variance explained by these models was modest, but where individual factors emerge as significant predictors this may be regarded as broadly reliable, though recognising that other, unidentified may be more important.

Higher PHQ-9 scores were associated with increased odds of both types of premature ending, lower GAD-7 scores were associated with reduced odds of proving unsuitable. High W&SAS scores were associated with proving unsuitable, whilst low ones were marginally associated with dropping out. Women and under 18s were more likely to prove unsuitable, 18 to 34s were more likely to drop out, whilst those over 65 were much less likely to. People of Mixed or Black ethnic groups were more likely to be considered unsuitable. Most of the specified diagnoses, if they predicted outcome at all were associated with lower odds of either type of premature ending. The exceptions to this were mental and behavioural disorders associated with the use of alcohol, bipolar disorder, eating disorders and family loss, all of which were associated with greater odds of proving unsuitable. Fourteen sites added significantly to the model for unsuitability while 24 contributed to the model for drop-out. We performed Hosmer-Lemeshow tests for goodness of fit on both models. For unsuitability it was non-significant (H-L 4.45, 8df, $p = 0.8144$) suggesting the model was broadly satisfactory statistically, for drop-out it was significant suggesting the model was less satisfactory (H-L 41.68, 8df, $p < 0.001$).

Numbers of sessions

The numbers of treatment sessions attended by patients who received treatment is illustrated in the charts 14 to 16 and a number of site specific numerical treatments are presented in table SS22. Following the comparison between the three different counts of session numbers (above) this analysis initially used the count of sessions by purpose where the purpose included treatment. This showed a highly skewed distribution, so medians and interquartile ranges are reported. The analysis was confined to individuals who had treatment interventions reported in this variable. Conclusions can only be tentative as a large proportion of data were missing³ (36.0% of patients having low intensity care only, 28.3% of those having high intensity care only and 14.6% of those having both). The proportion of cases with missing data varied considerably between sites; two submitted no usable data, one no missing data, and the proportion missing in the remainder had a

³ Missing data here means that the patient had one or more sessions by intervention type reported, but no session by appointment purpose where the purpose included treatment.

median of 27.1% (IQR 20.6% to 36.2%). Where reported, the medians and IQRs for treatment session numbers were 2 (1 to 4) for patients having only low intensity treatments, 3 (1 to 6) for those with only high intensity treatments and 3 (1 to 5) for those having both. The graphs however show that there was a clear distinction between session numbers for those patients who continued treatment until it was considered complete, and those who declined, dropped out, or were considered unsuitable (usually after treatment had started).

These numbers of treatment sessions were surprisingly low; well below the numbers recommended in NICE guidance (see box 4, next section). This raises the question of whether there might have been substantial failure to record all of each patient’s treatment sessions. The site specific tables showed that a small number of sites reported larger numbers. For patients receiving only low intensity treatment, in one site the median number of reported sessions was 5, whilst in two it was 4 and in a further four, 3. For patients having high intensity treatments alone, one site reported a median of 9 sessions, a second 8.5 a further three, six or more. Session counts using the total of sessions by type of treatment showed similar overall results: patients having low intensity treatments only -median 2 sessions, IQR 1 to 4, high intensity treatment only -median 3 sessions, IQR 1 to 6. For those having both low and high low intensity treatments, numbers were slightly higher - median 5, IQR 3 to 8, but still apparently low for people with the added complication of making a therapeutic step-up. Whether these records are accurate cannot be established without further corroborative detail. A total of sixteen or more treatment sessions (the NICE guideline recommended minimum for patients with depression or generalised anxiety disorder receiving high intensity treatments) was reached by only 165 of the 7,825 (1.38%) patients having high intensity treatment alone for whom count details are available. These were reported by 23 different sites, with no single site reporting more than 19.

Diagnosis and choice of treatment

A central element of the IAPT programme is that the treatments provided should be evidence based. Extensive guidance on the effectiveness of specific treatment approaches for particular clinical conditions has been collated by the National Institute for Health and Clinical Effectiveness (NICE), and is published in their Guidelines. Four Guideline documents are of relevance here. Box 4 sets out specific guidance about psychological treatments for conditions recognised in the primary diagnoses coded in the IAPT minimum dataset. Panic disorder has also been included here – this appears to have been an omission from the draft of the minimum dataset current at the time of the study (the diagnosis was used in 20 centres for a total of 769 patients).

Box 4. Treatment approaches recommended for specified conditions in current NICE Guidelines.	
Condition	Relevant guidance (relating to psychological interventions)
Depression in adults (Guideline 90)	Low intensity: Guided self help (6 to 8 sessions over 9 to 12 weeks), computerised CBT (9 to 12 weeks), structured group physical activity programme (3 sessions per week, 10 to 14

	<p>weeks).</p> <p>High intensity: CBT (16 to 20 sessions over 3 to 4 months), IPT (16 to 20 sessions over 3 to 4 months), behavioural activation (16 to 20 sessions over 3 to 4 months), behavioural couples therapy (15 to 20 sessions over 5 to 6 months) or, if these not agreeable, counselling (6 to 10 sessions over 8 to 12 weeks) or short term psychodynamic psychotherapy (16 to 20 sessions over 4 to 6 months).</p> <p>For relapse prevention: Individual CBT, mindfulness-based CBT.</p>
Generalised anxiety disorder (Guideline 22)	<p>CBT – 16-20 hours in 1-2 hour sessions over no more than 4 months.</p> <p>Briefer CBT 8-10 hours designed to integrate with self-help materials.</p> <p>Self help - bibliotherapy</p>
Panic disorder (Guideline 22) (not included in the list of diagnoses in the dataset)	<p>CBT – 7 to 14 hours, in 1 – 2 hour sessions over no more than 4 months.</p> <p>Briefer CBT – 7 hours, integrated with self help materials</p> <p><i>For a few people, brief, more intensive CBT</i></p>
Obsessive compulsive disorder (Guideline 31)	<p>CBT including exposure and response prevention – for mild, low intensity – initially up to 10 therapist hours; for moderate more than 10 therapist hours</p>
Post traumatic stress disorder (Guideline 26)	<p>Trauma-focussed CBT (8 to 12 sessions, usually weekly)</p>

Table 32 shows the broad pattern of treatment received by patients classified according to the primary diagnosis assigned to them. Just under 40% of patients had either no diagnosis or an un-illuminating one (mental disorder not otherwise specified). However several of the identified diagnosis groups had distinct treatment profiles. Of the patients with obsessive compulsive disorder, all three types of phobia and post-traumatic stress disorder, just over half received CBT without counselling while 20% to 40% received low intensity treatments only. For patients with depressive episode, mixed anxiety depressive disorder or generalised anxiety disorder, around 20% received CBT, a similar proportion counselling and a little over 40% only low intensity treatments.

Patients with recurrent depression were, (appropriately), more likely to get CBT. Counselling was the most common approach in the treatment of reactions to family loss, but, scarcely less common in the treatment of generalised anxiety disorders (for which NICE does not recommend it) than for depressive episodes (for which it is recommended as a treatment where patients do not want CBT or pharmacological treatments).

Table 33 shows a more detailed picture of the types of high and low intensity treatments that were provided to patients in the largest diagnostic groups. The less common types of high intensity treatment and individual low intensity techniques are shown here. This table simply shows the proportion of patients in the diagnostic category receiving the treatment; patients commonly receive more than one type of treatment. Patients with OCD and PTSD were notably less likely to receive no high intensity treatments and more likely not to receive low intensity. IPT and, to a small extent behavioural activation, were provided more for patients with depressive and recurrent depressive disorders than other conditions. Psycho-educational groups were targeted predominantly at patients with depressive episodes, generalised anxiety disorder and mixed anxiety and depressive disorder. Site specific table SS23 gives provides an analysis of patterns of high intensity treatments for the four most common diagnostic groups.

Which patients received higher intensity treatments?

Overall, of those study group patients with finished episodes, who received any treatment, 46.0% received a high intensity intervention of some sort. There is no suggestion that this was a randomised process, sites assessed each patient and provided for them as they considered appropriate and feasible within resource constraints. Thus it is illuminating to ask which patients received high intensity treatment.

In relation to the four simple categorical variables used to describe the study group as a whole, women, people of working age or below, White British, and those referred by their GP were more likely to receive high intensity treatments. These patterns are shown in table 34. The initial test scores, or at least awareness of the answers to the test questions, would be expected to have had an influence. Charts 17 to 20 show the profile of initial scores on each of the numerical scales, comparing those who went on to have high intensity treatment with those who did not. A tendency for more severe scores in the former is apparent, and, as a result of the large numbers of cases, is highly statistically significant but the differences are clearly not large. Kruskal-Wallis rank sum tests provide a type of chi square test appropriate given the distributions of the figures. Summary numbers and significance tests are shown in table 35

However, as noted above, several of these factors are already related to each other, and also vary between sites. Sites differed in many ways in term of both the nature of the care provided and the quality and completeness of data. So analyses of the proportions receiving high intensity care in relation to individual variables in isolation are not altogether satisfactory. Logistic regression allows many variables to be considered at the same time.

Logistic regression models were developed for three dependent variables: whether the patient had received any high intensity care, whether they had had low intensity care with no high intensity care and whether they had received only interventions in the 'other interventions' category (mostly advice about benefits and employment). In all cases analysis was confined to patients with completed episodes who had received some treatment and had scores on all relevant predictors.

The predictor variables used were initial PHQ-9, GAD-7, Phobia questions and Work and Social Adjustment scores, gender, broad age, ethnic and referral groups, use of psychotropic drugs at referral and site. We explored also using a variable for whether the patient spoke English, but this was unavailable for a large majority of patients and identified very few patients who did not speak English so it made no effective contribution.

The resulting models were relatively weak, indicating that other factors, for which we had no measures, had an important bearing. However, where variables do emerge in models like this, it is reasonable to conclude that they are effective predictors of the outcome after all the other predictor variables have been allowed for. The results are shown in table 36.

The overall number of patients included in the models was considerably lower than the 38,891 patients with finished episodes who had received some intervention. This was the result of missing data: 11,237 patients had no ethnic category, 2,322 no usable age group, 750 no gender, 437 no referral source, 8,335 no recording of whether they were using psychotropics at the point of referral, and initial symptom scores were missing for PHQ-9 for 186 patients, GAD-7 for 379, Phobia questions for 2,794 and W&SAS for 1,940. In these models, a little over half of the sites emerged as significant predictors, indicating that there were many features particular to individual sites not allowed for by our more specific predictors.

However allowing for this, a number of statistically significant patterns emerged. Patients with higher PHQ-9 or phobia scores are slightly more likely to have a high intensity treatment and less likely to have a low intensity one alone. Those with higher GAD-7 scores are more likely to have a high intensity treatment. Women were more likely to have only low intensity treatments and less likely to have high intensity treatments or counselling. Compared to the reference group, those age 35 to 64, older patients were more likely to have some treatment and less likely to have high intensity treatment or CBT. Younger age groups were also less likely to have high intensity treatments or counselling. Compared to White British, Asians were less likely to have any treatment, Blacks were less likely to have high intensity treatment and both were less likely to have CBT.

Compared to the majority referred by their GP, self-referred patients were more likely to be treated but less likely to receive any high intervention treatment, while those referred by other agencies were more likely to have high intensity treatments, and CBT, but less likely to have counselling.

Diagnoses were viewed alongside the commonest group, patients with depressive episode. In comparison almost all specified diagnoses, including recurrent depressive disorder, agoraphobia, social phobia, specific phobias, OCD, PTSD, somatoform disorder and eating disorders, were associated with greater odds of higher intensity treatment and CBT, and lower odds of low intensity treatment only and counselling. Mixed anxiety depressive disorder followed some of this pattern, though it did not significantly predict any high intensity treatment or low but not high intensity. Generalised anxiety disorder differed slightly in having lower odds of high intensity treatments and higher odds of low intensity only. Two diagnoses stood out as different. Mental and behavioural disorders related to alcohol use were associated with low odds of any high intensity treatment or of CBT, while family loss was associated with increased odds of high intensity and strikingly high odds of having counselling.

All treatments became less likely as the month in which the episode began fell later in the year. This probably simply reflects the fact that dropping out after assessment without treatment can be achieved more quickly than staying on and having treatment; thus recent starters who have not dropped out, are still in the system and thus excluded from the analysis. In treatment efficacy trials it would be common to follow all patients to episode ending, or at least for a uniform time from inception. However the need to observe the newly set up services and to report relatively quickly meant that we had to use a less neat dataset.

Hosmer-Lemeshow tests for goodness of fit test were performed. They were non-significant (indicating no reason to suspect lack of fit) for the models of having any treatment (H-L 14.93, 8df, $p = 0.0606$) and having CBT (H-L 12.78, 8df, $p = 0.1196$). However they were significant for the models of having high intensity treatment (H-L 65.27, 8df, $p < 0.001$), low but not high intensity treatment (H-L 40.06, 8df, $p < 0.001$) and counselling (H-L 22.29, 8df, $p = 0.0044$), suggesting these models should be viewed with more caution.

How successful was treatment?

The success of interventions is measurable only for patients who have had treatment, and have been seen on more than one occasion, so that a progression is visible in their symptom scores. This subgroup of the overall study population numbered 26,780, one third of the total study population and 64.2% of those whose episodes were completed. Recovery rates are influenced by which patients are included or excluded from calculations. We attempted to follow the principles of 'intention to treat' analysis [9], omitting as far as possible no-one. All patients who attended more than once were included in analyses. In many cases rating measurements, particularly final measurements were missing. We handled this by doing two analyses, one using only cases with complete data ('complete data only' method), the other including all cases with available first ratings, and assuming no change where a second rating was missing ('second unknown – no change' method).

This overall approach to analysis makes it likely that outcomes will appear less successful than those reported in the Doncaster and Newham pilot studies, where missing data seem to have been much less of a problem. This is also likely because of the different handling of the group of patients deemed 'not suitable'. In Clarke's report of the pilot studies, progress charts show these patients dropping out of consideration as if before treatment[4]. In the present study it was clear that many whose eventual ending was classified as 'not suitable' had received treatment, in some cases several treatment sessions. Following 'intention to treat principles,' it was considered appropriate to include them here. In the pilot study, this group comprised 8.5% of referrals in Doncaster and 22.1% in Newham. In the present study they comprised 8.9% of the whole group of 137,285 for whom there was any recorded contact, and 6.1% of the study group. We do not have reasons why these people were considered unsuitable, however we have symptom scores for most of the 4,835 of them in the study group. Comparison of the profiles of both PHQ-9 and GAD-7 scores with the rest of the study group shows a small excess at the lowest score levels but a larger excess at the top of the scale. This may suggest that in many cases the individuals were 'unsuitable' in that they required more intensive care than the services were designed to provide.

The high level performance indicators for the project suggest a measure of success termed 'movement towards recovery'. This phrase was used in the reports of the pilot studies and

describes the situation of patient who were at or above a case threshold at initial assessment and below all caseness thresholds at follow up. We have followed this principle, though we have been able to apply it in more detail as we had an additional marker of outcome, the Phobia questions. We also applied the approach to the scales individually and to the benefit receipt and employment data giving a total of seven 'MTR' markers, definitions for which are set out in box 5.

Box 5 Definitions of Movement towards recovery markers.		
	Starting condition	Requirement for movement towards recovery – always requires that starting condition is met
MTR1	Either First PHQ-9 score of 10 or more OR first GAD-7 score of 8 or more	Latest PHQ-9 less than 10 AND latest GAD-7 score less than 8
MTR2	Either First PHQ-9 score of 10 or more OR first GAD-7 score of 8 OR score of 4 or more on at least one of the phobia questions	Latest PHQ-9 less than 10 AND latest GAD-7 score less than 8 AND no phobia question score greater than 3
MTRDep	Either First PHQ-9 score of 10 or more OR first GAD-7 score of 8 OR score of 4 or more on at least one of the phobia questions	Latest PHQ-9 less than 10
MTRAnx	Either First PHQ-9 score of 10 or more OR first GAD-7 score of 8 OR score of 4 or more on at least one of the phobia questions	Latest GAD-7 score less than 8
MTRPhob	Either First PHQ-9 score of 10 or more OR first GAD-7 score of 8 OR score of 4 or more on at least one of the phobia questions	Latest phobia questions - no score greater than 3
MTRben	Receiving either social security benefits or statutory sick pay at initial assessment	Not receiving social security benefits or statutory sick pay at latest assessment
MTRemp	Economically active and unemployed at initial assessment, and economically active at latest	Economically active and employed full or part time at latest assessment

The simplest way to interpret these markers is the calculation of recovery rates (number of recovered cases divided by number of initial cases). However this is complicated in three ways. First, individuals may get ill as well as well, so the parallel 'incidence rate' (number becoming cases divided by number not at case level initially) is also important. Second, some conditions are more widespread than others, hence smaller recovery rates may indicate larger numbers of people benefitting. These two issues can be addressed by presenting a net change in prevalence rate. The third issue specifically affects the employment marker. In this case, as well as gaining or losing employment, individuals can move in and out of the labour market (by changing their status to or

from the economically inactive categories of home-maker / carer, retired or full time student). 738, 4.2% of those initially in the labour market, did this. 152 (43.4% of them formerly unemployed) became full time students and 586 (61.6% of them initially unemployed) became homemakers / carers or retired.

Table 37 shows the figures for the outcome measures. The upper half of the table shows analysis using complete data only, the lower half assumes that patients with second ratings missing remained as they were at initial assessment. The true figure is likely to be somewhere between these two. Recovery rates for individual rating scales are better than for the compound measures MTR1 and MTR2. This is scarcely surprising as the task of producing a recovery becomes greater as more scales are added to the marker. However the 'incidence' figures suggest other possible explanations: patients may either develop new symptoms as their presenting problem remits, or the process of being asked in their assessment about a wide range of possible experiences may lead them to identify as symptoms phenomena to which they had previously given little thought. The size of the increases for patients crossing from non-cases to cases are not small: mean increases are 7.0 (95% CI 6.7 to 7.2) for the PHQ-9 in those becoming cases on this variable, and 6.3 (95% CI 6.15 to 6.54) for the GAD-7, suggesting that the score increases of those becoming cases cannot be dismissed as random fluctuation.

The net change in prevalence figures suggest that the programme is associated with a highly significant and substantial fall in caseness levels for all the clinical markers. Among the single marker measures, the net change impact on phobic states was notably less than that on depression and anxiety, mainly because these conditions were less common in the study group at the outset. Recovery rates from phobic states were similar to those from depression and anxiety on the basis of the more optimistic 'complete data only' analysis, but slightly, though significantly, poorer using the less optimistic approach. There was a small, though statistically significant fall in benefit claimancy, though not in unemployment.

Site specific tables SS24 and SS25 show corresponding site specific figures, and a selection of these, case-recovery and net change in prevalence, for the MTR1 and MTR2 markers, are shown graphically in charts 21 to 24. The pattern these seem to show is of a substantial variation in effectiveness between sites: in all four graphs, the 95% confidence intervals for the upper and lower thirds have very little overlap.

Recovery rates for the three types of symptom were highly correlated between sites. Using the 'second unknown no change' assumption the depression/anxiety, depression/phobia and anxiety/phobia recovery rates had pair-wise correlation coefficients of 0.96, 0.78 and 0.80, all significant at the $p < 0.001$ level. Incidence rates showed less correlation –corresponding coefficients were 0.77 ($p < 0.001$), 0.34 ($p = 0.069$) and 0.41 ($p = 0.024$). The 'complete data only' approach produced similar results. The difference is interesting. The close correlation in both directions between anxiety and depressive symptoms presumably reflects the fact that they commonly occur together. However the much closer correlation of recovery than incidence for phobia with the other two symptoms may indicate that symptom recovery is the result of a purposeful process at which some sites are more successful than others, while symptom onset is a more random process.

Employment 'recovery' (moving from unemployed to employed) was significantly correlated with phobia recovery (coefficient using 'second unknown no change' = 0.59, $p < 0.001$). It was also correlated with anxiety recovery (0.43, $p = 0.18$) and both compound indicators (MTR1: 0.40, $p = 0.03$,

MTR2: 0.51, $p=0.004$), however these three findings should be treated with caution since the parallel calculations using the 'complete data only' approach produced non-significant findings. The only significant association for 'benefit recovery' – moving away from being in receipt of benefit – was the predictable positive correlation with employment 'recovery'.

Changes in employment and benefits status in the pilot studies were studied by Clark [4]. These authors reported a 'net increase' of 4% in the larger Doncaster pilot and 10% in the smaller Newham pilot 'in the number of people in work and not receiving statutory sick pay'. They helpfully provide detailed figures which seem to show that it was in fact the proportion of patients in work and not claiming benefits that rose by these figures (Doncaster – 39.3% to 43.1%, Newham 42.2% to 51.9%)⁴. Site specific table SS25a shows detailed figures for the study sites. Overall the proportion in work and not claiming benefits rose by 2% (95% CI of difference in proportions = +1.0% to +3.0%). Individual sites showed rises in 23 of the 30 sites where this could be calculated, with four falls and three unchanged, however none of the site level figures was statistically significant in isolation.

The association between primary diagnosis and outcome is shown in more detail in table 38. This table extends over 8 pages and sets out recovery rates and net changes in prevalence for the MTR1 and MTR2 markers, tabulated for the common primary diagnoses and the common combinations of high intensity treatment. The calculations are done for both approaches to handling missing second ratings. Overall results as measured by the MTR2 marker are substantially less favourable. However, given the poor performance of the phobia questions (the only difference between these result markers) in respect of what we are able to infer about their positive predictive value in this sample group from our attempts at validation described above, we would regard the MTR1 findings as the more reliable. The MTR2 findings are shown for the sake of completeness only.

It is important to stress that this cannot be seen as a test of the comparative efficacy of the different treatment approaches, as patients were not assigned randomly, but to the approach which appeared most suitable in the light of initial assessment and locally available resources. The table makes it clear that the different approaches were used selectively for different problems. As the programmes included a substantial element of training for CBT therapists, it is also likely that a substantial proportion of the staff providing CBT were inexperienced or trainees, whilst those employed to provide counselling were probably mainly already trained and experienced.

We attempted to explore further what factors were most influential in predicting positive outcomes (movement towards recovery) on each of these markers using multivariate analysis (logistic regression). The results are presented in table 38a; most statically significant influences predicted lower likelihood of recovery.

In almost all cases higher scores on all the symptom rating scales were associated with reduced odds of recovery. Use of psychotropics was associated with reduced odds of recovery on the PHQ-9 or phobia questions. Compared to those aged 35 to 64 year, people of 18 to 34 were less likely to recover on the GAD-7, but both younger and older were more likely to recover on the phobia scales. Compared to White British, Asians were less likely to recover on any scale, and Blacks less likely to recover on the PHQ-9. Compared to those referred by their GP, self-referred patients were more

⁴ The actual numbers in the sites rose by more than this – Doncaster: 175 to 192 – 9.7%, Newham: 57 to 70: 22.8%.

likely to recover on the PHQ-9. In comparison to those with primary diagnoses of depressive episode, those with mental or behavioural disorders from use of alcohol, recurrent depressive disorder and family loss had lower odds of recovery on the PHQ-9 scale, while those with generalised anxiety disorder did better. Interestingly, treatment of those with specific phobias appeared to be associated with lower odds of recovery as measured by the phobia questions.

Hosmer-Lemeshow goodness of fit tests were satisfactory (non-significant) for all five models (MTR1 H-L 2.81, 8df, $p = 0.9457$; MTR2 H-L 8.45, 8df, $p = 0.3905$; MTRDep H-L 6.52, 8df, $p = 0.5891$; MTRAnx H-L 3.24, 8df, $p = 0.9182$, and MTRPhob H-L 10.76, 8df, $p = 0.216$).

The finding in respect of the poorer outcomes for Asians differs from the finding of Clark and his colleagues in the Newham pilot study [4]. Asians there had a recovery rate 32% better than White British, although this difference was not statistically significant given the sample size (roughly 134 White British and 67 Asians). However it is significantly at variance with our observation that after allowance for other factors, Asians were only 0.7x as likely to recover on the MTR1 measure.⁵

An alternative approach to examining the success of the interventions is to look at the numerical changes in scale scores. The distribution of score changes is substantially closer to a Gaussian normal distribution than the initial score profiles (charts 25 to 28), though in all cases the distributions deviate significantly from a pure normal pattern. The most noticeable aspect of this is that they all show a zero (no-change) peak. Table 39 shows how changes overall vary across the categories used to describe the study group earlier. Using oneway analysis of variance, it is apparent that women showed a slightly greater fall than men in all three symptom scores, though no difference in their reduction in W&SAS. The youngest and oldest age groups showed the largest falls in PHQ-9 scores but the largest in W&SAS; however in view of the numbers involved the significance of these observations was marginal, and the GAD-7 and phobia change scores showed no association with age group. There were significant inter-ethnic differences in the change in PHQ-9 and GAD-7 scores; in each case the largest fall was recorded for White British, with Black people following closely. Mixed and Asian groups recorded lower reductions. Symptom score changes did not vary significantly between broad referral groups, though W&SAS scores reduced more for self referred and less for those referred from others not referred by their GP.

Table 40 shows a similar analysis in relation to primary diagnosis. Ignoring the residual categories that we cannot characterise, patients with depressive episodes and 'mixed anxiety depression' showed the greatest fall in PHQ-9, followed by those with PTSD and with recurrent depression. GAD-7 scores reduced most in patients with Agoraphobia and with PTSD, followed by those with Generalised anxiety disorder and OCD. Phobia scores fell most notably in patients with Agoraphobia, specific phobias, social phobias and PTSD. In relation to treatment approaches, amongst those patients receiving high intensity treatments, GAD-7 and phobia scores fell more in patients given CBT. PHQ-9 scores fell more in patients given counselling than CBT alone, but more still in those given the two in combination. However it should be repeated here, that this is in no

⁵ In the Newham findings, the recovery rate for White British was 50%, 95% CI using Wilson's method, 41.7% to 58.3%; the rate for Asians was 66%, 95% CI 54.1% to 76.2%; had the rate for Asians been 0.7x the rate for White British, it would have been 35%, 95% CI 24.7% to 47.9%. A similar calculation for the Blacks suggests that the difference we observed in the level of improvement in the PHQ-9 scale would not have been statistically discernible.

sense a comparison of the efficacy of the treatments as individuals were assigned to treatment packages on the basis of clinical judgements about what seemed most appropriate. Site specific table SS26 gives similar figures for the four most common diagnostic groups, for patients treated with CBT (irrespective of what else they received).

We explored the way all these factors come together using multiple regression to explore significant predictors for PHQ-9 change, GAD-7 change and W&SAS change (table 40a). In these models it is important to remember that the factors tending to produce desired outcomes (greater falls in symptom score), have a negative signs.

Higher initial PHQ-9 scores were associated with greater PHQ-9 falls, but lesser falls in the other two ratings. Higher initial GAD-7 scores were associated with greater GAD-7 falls, but lower falls in the other two scales. Higher phobia scores were associated with lower falls in all three scales. Initial psychotropic use was associated with lower reductions in PHQ-9. Females had lower reductions in PHQ-9 and W&SAS, people over 65 had greater falls in all three scales. People of mixed race and Asians achieved lower reductions in both PHQ-9 and GAD-7, Asians had lower reductions in W&SAS and minority white groups had lower reductions in GAD-7. In comparison to those with depressive episodes, people with mental or behavioural disorders due to use of alcohol, and those with recurrent depressive disorder did less well having lower reductions in both PHQ-9 and GAD-7. People with generalised anxiety disorder and the three phobic groups had greater reductions in PHQ-9, those with specific phobias had greater reductions in W&SAS and those with agoraphobia on all three scales. Eating disorder was associated with lower falls in PHQ-9 and OCD, somatoform disorder and family loss lower falls in GAD-7. Ten sites emerged as predictors of outcome in the PHQ-9 and W&SAS change models, eight in the GAD-7 model. In all three cases the models were weak, predicting 22% to 24% of the variance in outcomes.

We attempted to present a more direct comparison with the pilot work. Table 41 presents symptom change scores in a directly comparable way. Richards and Suckling [10] provide the more detailed account of symptom outcomes in the Doncaster site. They found the mean PHQ-9 score fell from 15.96 to 8.09 (effect size 1.09) while mean GAD-7 scores fell from 13.98 to 7.22 (effect size 1.07). Clark et al [4] report similar findings for the Newham pilot (PHQ-9 15.3 to 8.2, effect size 0.99 and GAD-7 13.7 to 6.8, effect size 1.19). The comparable overall effect sizes for the 30 sites we can report here were more modest: 0.69 for the PHQ-9 and 0.72 for the GAD-7. Individual site effect sizes ranged for the PHQ-9 from 0.38 to 0.95 (median 0.68, IQR 0.59 to 0.76) and for the GAD-7 from 0.41 to 1.09 (median 0.70, IQR 0.62 to 0.82). Detailed figures are in site specific table in table SS28.

Richards and Suckling went on to show the extent of symptom reduction was related to how care episodes ended. Our findings for the sites combined show a similar picture. Patients whose episodes ended with treatment being considered complete (53.5% of those with finished episodes of two or more contacts) showed effect sizes of 0.97 for PHQ-9 score reduction and 1.04 for GAD-7 score reductions. Those who declined or dropped out, or who were considered unsuitable had higher initial ratings and showed much smaller average reductions.

Associations between outcomes and service level factors

We undertook a simple rank correlation analysis of the outcomes of services and a small number of salient service level variables. The results are shown in table 42. The MTR1, PHQ-9, GAD-7 and overall phobia questions recovery rates (proportion of patients who score at case level at initial

assessment but not at final assessment) are familiar markers from the previous section. We used the 'second unknown no change' approach results here. Service variables added were the total study group size, the proportion of the study group still in the system (as a measure of how effectively patients were being moved through the system, though possibly also identifying newer services), the proportion of therapist sessions contributed by high intensity therapists, the proportion of patients receiving high intensity therapy who had CBT, the proportion of patients having only low intensity therapy, and the 'step fraction' (the proportion of patients receiving any high intensity care who also received low intensity care).

The symptom recovery rates were all highly correlated with each other. The proportion of therapist sessions contributed by high intensity therapists was the most consistently significant of the service variables. This was moderately correlated with all four recovery measures, most strongly with the MTR1 measure, also with the proportion of patients still in the system and negatively with the proportion of patients having low intensity therapy only. The proportion of patients receiving high intensity therapy who had CBT was moderately negatively correlated with overall patient group size. The proportion of patients getting high intensity care who 'stepped' was moderately negatively correlated with the proportion of patients still in the system; the interpretation of this is not obvious. Stepping was also negatively correlated with the proportion of therapist sessions by high intensity therapists, and positively correlated with the proportion of patients having low intensity treatment only. There were no other significant correlations. The figures from which these correlations were calculated are set out in site specific table SS28.

Outcome in relation to the phobia questions

Table 43 explores the issue of the performance of the phobia questions in relation to treatment outcomes for patients assigned the corresponding conditions as primary or secondary diagnoses. We undertook the analysis on both a 'complete data only' and a 'second unknown – no change' basis. In each case patients assigned the diagnosis were divided into groups by treatment approach – low intensity treatments only, CBT, counselling or both CBT and counselling. Only low intensity and CBT treatment results are presented as in all other cases ten or fewer patients were involved in some of the calculations. There were no significant differences in recovery or incidence rates or in the net prevalence change between low and high intensity treatments for individual conditions, and only one (of dubious interpretability) between conditions⁶.

⁶ The incidence of new cases of specific phobias in people receiving CBT was significantly greater than that of social phobias.

Discussion

Data collection

A central feature of the IAPT programme is on-going evaluation of the work of the new services. The detailed specification for the clinical monitoring dataset was developed prior to new services starting work and all new services were expected to adhere to it. For the new services this was a substantial undertaking. In many cases the services were designed to sit outside routine secondary care mental health services. This meant they could not necessarily have access to an NHS Trust IT service, and the practical support this could provide. Even where they did, it is quite likely that some work would have been undertaken outside Trust premises. The speed of introduction of the new services meant that even where they were administratively located in conventional secondary care trusts, it was usually unrealistic to expect existing routine data collection mechanisms to be modified to accommodate the complex new requirements from their inception. The use of specialist information system suppliers, operating remotely through the NHSnet, was innovative and appears to have been remarkably successful in supporting a high level of completeness of information collection from the start of service operation.

The level of completeness of data items (field completeness) was high in most areas from most sites. Exceptions fell into two categories. One comprised items which are commonly poorly completed in administrative databases, most notably the disability ratings, ethnic categories and diagnoses. The other related to issues of administrative process, notably dates of treatment milestones: referral, first contact, assessment, and start and end of treatment. Here, the widely different patterns of data between sites suggests that there are different administrative arrangements about how and when referred patients are entered onto the system. Standardising these is likely to be extremely difficult as services are always likely to be organised in subtly different ways. Administrative data are always likely to reflect this. However in addition to this there were some clear internal contradictions in the data, most notably in respect of the three types of contact counts. It is likely that with more time available, or with a routine of repeated quarterly or annual data transfers, these issues should be amenable to correction.

The other aspect of completeness relevant to evaluations of this type is record completeness – the question of whether all individuals having relevant interactions with services were documented. By its nature this is hard to assess in this type of situation. Where all services studies are wholly responsible for performing a task of predictable volume, high or low figures can reasonably bear interpretation. However in this case, most sites were in the early months of their life and it is unlikely that referral routes will have been thoroughly established, or knowledge of the availability of the service universal. Hence we can only note here that there is a wide range (more than an order of magnitude) of population base rates of actual service use (having an initial assessment with whatever follows), and also a wide disparity in the proportion of those people who appeared at least in a fragmentary way in the records supplied to us, who went on to have an initial assessment. Working through the explanations for these differences between sites will be an important task in establishing a minimum dataset for the service in the long term.

Beyond the completeness of the data we collected, we would recommend that in future the dataset should be developed to identify well crystallised dates for 'steps' from one treatment level to another in the 'stepped' conceptualisation of the care process. This issue was not explicitly

conceptualised in the initial dataset specification and we were therefore not able to request it. A key omission by us was that as a result of a design oversight, we did not request from sites the duration of patients' symptoms prior to presentation. This variable is amongst those that we understand is routinely collected and would probably have improved our models of site treatment outcomes.

Usability of the rating scales

In most sites, rating scale data achieved very high level of recording completeness. These rating scales are evidently collectable in this type of high-volume, routine setting. This simple fact has profound evidential implications for the use of this type of approach in outcome monitoring more generally in mental health care.

Given its large size, the resulting dataset probably offers a substantial resource for further focussed work on the psychometric properties of the scales used. The present authors are not experts in this field, however it is clear that the dataset assembled is of a different order of magnitude from, for example that collected by Mundt and his colleagues to evaluate clinical thresholds in the Work and Social Adjustment scale [3] (see above) and thus offers wider analytic possibilities. Rating scales are not good or bad, valid or invalid in absolute terms, they are more or less suitable for specific purposes. The performance of the rating scales used here showed a number of specific issues which deserve consideration. First several of them showed apparent 'floor' and 'ceiling' effects. This was most important in the GAD-7 measure at initial assessment, where there seems to have been a substantial number of patients for whom further divisions at the top of the scale would have provided useful separation, enhancing the chances of identifying score changes during treatment. This is important for analyses using Richards and Sucklings [10] definition of recovery, where a patient halving their score on one of the scales is one of way of defining having recovered. The Work and Social Adjustment scale showed a clear 'floor' as defined, though if scores below 10 on this measure are considered clinically irrelevant this may matter less.

Both the phobia questions and the Work and Social Adjustment scale seemed to show some sawtooth patterning (alternating higher and lower frequencies for successive scores). This was clearest on the total score from the three phobia questions and, as noted, may reflect the provision of nine rating points but only five textual guides for them. The phobia questions were the least satisfactory of the rating instruments. They were developed for the programme and have not, as far as we know, been extensively tested. Their performance has been discussed extensively above. Unfortunately our dataset was not a good testbed for a new scale. Their sensitivity and specificity needs to be tested in a patient group properly characterised by well established diagnostic measures for the phobic conditions they are intended to identify. As described, they performed poorly when judged against the primary and secondary diagnoses assigned in our study; however these diagnostic data are not of sufficient quality for this task. These diagnoses in the present study should be seen as probably less complete and accurate even than ordinary routine clinical diagnoses. We have no evidence about whether referrers or IAPT assessors assigned these, though in the case of self referred patients it must have been the latter. Assessors may in some cases have been relatively junior trainee therapists, working at low intensity levels. It is clear from our further analysis of the use of the 'mixed anxiety and depressive disorder' category that their use of ICD10 categories was not accurate. It still less clear that it was complete (in other words that patients were assigned all

the diagnoses which were applicable). In any case the dataset did not permit them to assign more than two diagnoses.

The phobia questions have a slightly unusual conceptual structure in that they are intended to establish whether an individual has any one of three possible symptoms. In analytic terms this is not difficult to operationalise: any single case level score establishes caseness. It has the slightly odd corollary that a person may have recovered from the type of phobia that established their initial caseness but still be at phobia case level because of worsening symptoms in another area⁷. For the purpose of measuring changes in symptom scores, we considered the three types of phobic symptom could be grouped to give a single score representing the overall burden of phobic symptoms. We understand that this may be considered controversial.

Assignment of diagnoses

Diagnostic data are of interest here primarily because they reflect the categorisation of patients in relation to NICE treatment guidelines. The idea of categorical, as opposed to dimensional, classifications in the area of common mental disorders is inherently problematic, since it is by crossing thresholds on continuously scored symptom scales that these conditions are operationally defined. The use of diagnostic categories is not favoured by many psychologically oriented therapists. Thus the overall relatively poor level of diagnostic coding is not surprising. However it is notable that the level of coding differed very substantially between sites. This therefore needs to be seen as an issue of locally operating policy. If the goal of establishing that psychological treatments provided by the new services are evidence-based in the sense of being compliant with NICE guidelines for defined groups of patients, completeness will need to be improved.

Other than the overall level of entering diagnoses, there were three specific issues in this area needing attention. The first was in the dataset coding frame which currently specifies a small subset of diagnoses likely to be referred to the services. At present this does not include panic disorder. The second relates to the apparent confusion surrounding the appropriate assignment for individuals with both anxiety and depressive symptoms. The apparent misuse of the term 'Mixed Anxiety Depressive Disorder' is discussed above. It seems likely that the usage of this term in the data submitted was wider than that covered by the ICD10 code F41.2, and the applicability of NICE guidelines to these patients is thus not clear. The third is the use of non-specific codings. While technically a valid ICD10 code, in the present context this needs to be considered as missing data.

Patients receiving assessment and treatment

The data made it clear that the very substantial number of patients reaching the sites and engaging with assessment – nearly 80,000 in all, were in most cases clearly suffering with the types of disorder envisaged. Overall nearly 83.6% of patients reached caseness on one of the well established scales and over 75% were significantly impaired or worse in terms of work and social adjustment. Over three quarters were economically active, and of these 30% were currently unemployed. These levels of caseness were slightly lower than in the two pilot studies: they found caseness rates of 90% in both sites using the GAD-7 and the PHQ-9.

⁷ In fact this was the case for 5.8% of the 13,212 Study Group patients who scored at case level for any phobia at both first and last assessments.

The multivariate analyses suggested that in relation to diagnoses and symptom scores, treatments were assigned as expected. More severe depressive or anxiety symptoms, or diagnoses of recurrent depression, phobias, OCD, post traumatic stress disorder, somatoform disorders and family loss were strongly associated with greater likelihood of high intensity treatment. In most cases the same disorders also predicted receiving CBT and not receiving counselling, though family loss strongly predicted counselling and not CBT. Interestingly, the diagnosis of mixed anxiety depression did not strongly predict high or low intensity treatment, and while not significantly predicting CBT, significantly predicted not getting counselling.

Patterns of treatment received

An aim of the programme is to provide 'evidence-based' treatments. We explored the current NICE treatment guidelines for details of recommended treatments for the types of mental health problem prominent in study group subjects. Four guidelines seemed relevant, the two most important numerically are those covering depression and generalised anxiety disorder. These two guidelines differ considerably in the detail in which they recommend approaches, the depression guideline recommending a wider range of approaches.

However the patterns of treatment patients received in different sites varied widely. At the low intensity level, structured exercise and computerised CBT, both recommended treatments for depression, were used extensively in a third to a half of sites respectively. This seems surprising since these would be amongst the easier programme components to establish. At the high intensity level, counselling, recommended by NICE as a fall-back treatment for patients with depression who are unwilling to agree to other approaches, was used for around 95% of patients receiving high intensity care in two sites and none in two more (median 41.3%, IQR 10.6% to 56.6%). It seems more likely that this variation reflects differences in the resources that could be mobilised at short notice than variations in the needs and expressed wishes of patients. Given focus of the programme on seeking to develop efficient structures for delivering evidenced based treatments to large volumes of previously un-served patients, it seems surprising that the dataset does not ask staff to record at assessments which NICE guideline is applicable, and what pointers they have identified to the care path point identified in the guidelines.

In addition to variations in the types of treatment given, the numbers of clinical sessions recorded for patients fell substantially below what NICE guidelines indicate is appropriate for almost all types of care. For example, the Depression guideline, recommends that for low intensity care, patients should have six to eight contacts for guided self-help or around thirty sessions of physical exercise. For patients treated with CBT the guideline recommends sixteen to twenty sessions [ref]. The median number of contacts for patients having just low intensity treatments was two, for those receiving high intensity it was three. There were some exceptions to this; the top two sites had median attendance numbers for high intensity treatments in excess of 8, while for low intensity work, but overall, it seems that the pattern of treatments as reported in these data are, in most places, some distance from NICE guidelines.

Outcomes

It was not to be expected that clinical outcomes in the roll-out sites would match those in the pilot sites. Widespread programme implementations results seldom match the effectiveness achieved in demonstration centres with the most able and committed staff, the élan of pioneers, and usually

relatively generous resources. The pilot sites demonstrated that treatments of established efficacy could be introduced quickly, and on a large scale by the development of specialist centres using a stepped care model and undertaking extensive staff training to develop a workforce. Our study was intended to monitor the progress of ordinary sites following them in the first year or so of their work. It was intended to identify what was working well and what less well to guide the programme's managers in their direction of subsequent phases of work.

While the sites we studied did not match the success of the pilots either jointly or in any individual case, nevertheless they achieved substantial clinical success. Our principal analyses were as conservative as would be plausible. They included all patients who attended more than once irrespective of whether they completed their treatment or were subsequently considered unsuitable, they allowed for the development as well as the remission of symptoms, and patients with missing final ratings were assumed unchanged. With these assumptions there was a net fall of almost 30% in the rate of caseness on the PHQ-9 or the GAD-7 combined, and of 27% in caseness for each individually.

However there was a notable variation in the effectiveness of different sites. The 95% confidence intervals for outcomes in the most, and least effective thirds of the sites scarcely overlapped. We explored the question of whether outcomes were related to any obvious service characteristics. The only significant association we found was to the proportion of staff employed at higher levels, and thus presumably inversely related to the proportion of trainee therapists undertaking treatment sessions. This issue needs further exploration.

We explored a number of different ways to characterise the outcome of the interventions. Our conclusion was that the most satisfactory was the overall net change in the caseness rate between initial assessment and final contact, using the GAD-7 and the PHQ-9 combined. We would not recommend the use of the phobia questions in their present form as discussed elsewhere. It is important that the outcome ratings should not be confined to whether initially present symptoms remit, as we demonstrated that a significant number of individuals acquire symptoms (or at least positive symptom ratings) during the course of treatment. There also needs to be an explicit method of handling missing outcome data. In pilot work, it is just about possible to achieve near complete datasets; in routine practice this is unlikely. Some assumption about what happens to those missing final ratings is needed. The only feasible possibilities are to exclude them from analysis (as if to assume they had never been treated), to assume that they all became cases, all got better, fared identically to the other patients, or stayed as they were. We would recommend the last of these. When spelt out, the first three seem absurd. The fourth would probably be complacent, as our evidence, in line with Richards and Sucklings findings, was that those ending their treatment prematurely have substantially poorer outcomes.

The relevance of work and benefit outcomes is clear in relation to the history of the commissioning of the programme. It was very unfortunate for the sites work that the study period coincided with a severe economic recession making it exceptionally difficult for people seeking work. However we consider that this outcome, as measured was in any case unlikely give a good reflection of how the programme's actual success in getting people back to work and off benefits. The use of the last clinical contact event as the end point for monitoring is pragmatic, and from the point of view of symptom recovery may be reasonable. However for these two (related) outcomes it seems likely to

underestimate effectiveness if the programme is successful. Patients may do well in treatment but may not start looking for work until their treatment is complete and they have regained some confidence. There is no obvious solution to this problem. If services were asked to follow up patients at some interval after the end of treatment, this would be burdensome and likely to achieve low response rates. The likelihood of patients responding could well also be related to their satisfaction with their treatment and its outcome.

In addition to symptomatic relief, our findings suggest that one process measure would be useful for monitoring purposes. This would be the proportion of patients starting treatment who finishing by completing. We would interpret this as a measure of the effectiveness of patient selection. The association with outcomes is clear. A small proportion of those not completing treatment appeared to have benefitted, however the net fall in prevalence of caseness in those dropping out was less than a third of the figure it was for those who completed treatment, while for those subsequently considered unsuitable it was close to nothing. It is not realistic to expect sites to get selection decisions right all the time, however there was a substantial range among our sites (median 42.6%, IQR 35.1% to 50.8%, inter-decile range 22.1% to 71.8%). This suggests that selection can be done more or less effectively; this is evidently critical both to targeting limited resources where they will have the greatest beneficial effect and to ensuring that individuals who need more support than is feasible in these services are referred on to more suitable care as early as possible.

Equality issues

In terms of equality of access, there appeared to be some concerns. Older people and men appeared under-represented in relation to expectation based on the patterns of morbidity shown by the psychiatric morbidity survey. The position for people with disabilities is not recorded at all in most sites, making it difficult to see how commissioners and providers can discharge their responsibilities to promote access to services for disabled people under Disability Discrimination legislation.

Asian, Black and 'Other' minority ethnic groups and male members of white minority ethnic groups appeared under-represented amongst those reaching initial assessment. The most authoritative study of patterns of common mental disorders in minority ethnic groups in England suggests that the prevalence of common mental disorders in these groups are overall fairly similar to those of White British people [11], however the numbers coming to treatment are unrepresentatively small. Particularly in the case of Asian minority groups, the pattern appears to be carried through from overall numbers in the study group to the likelihood of receiving high intensity treatments. After allowing for all other relevant factors for which data were available, Black people were significantly less likely to receive any treatment or to recover on either the two scale or the three scale makers, Asians were less likely to receive high intensity treatment (CBT or counselling), and both were significantly less likely to receive CBT. However, it seems important to stress at this point that this is probably the first time this issue has been systematically analysed for most sites. It would have been surprising if access and treatment responses had been uniform from the outset. These figures provide a first indication of issues sites will need to address.

Conclusion

In concluding, it seems important to note that despite the limitations and shortcomings identified in this report, the fact of the collection of such a large outcome dataset is in itself a remarkable achievement for the services. Introducing outcome measures into mental health services more generally has been the subject of extensive work over the last fifteen years but has repeatedly proved elusive [12-14]. The data presented here were collected in the context of ordinary routine practice in evidently busy, new clinical units. As far as we can see, they provides a rich and largely representative view of the problems and progress of the patients, and, more widely of the new service.

Health service developments on the scale of the IAPT programme are in themselves unusual. Outside the controlled environment of clinical studies, the collection of routine statistical data about health service operation is a difficult task in itself. Integrating clinical rating data to an extent that provides a plausible basis for service evaluation based on individual clinical outcomes is particularly difficult, requiring, as it does, active participation from a large number of clinical professionals working for many different organisations. This has been largely achieved here provides a very detailed picture not only of the detailed issues likely to need attention in forthcoming months, but also of the very substantial extent of the clinical success achieved by the programme.

References

1. DH, *IAPT Implementation Plan: National Guidelines for Regional Delivery*, Department of Health, Editor. 2008.
2. DH, *Technical Guidance for IAPT Key Performance Indicators, March 2009, Version 3.*, D.o. Health, Editor. 2009: London.
3. Mundt, J.C., et al., *The Work and Social Adjustment Scale: a simple measure of impairment in functioning*. Br J Psychiatry, 2002. **180**: p. 461-4.
4. Clark, D.M., et al., *Improving access to psychological therapy: Initial evaluation of two UK demonstration sites*. Behav Res Ther, 2009. **47**(11): p. 910-20.
5. McManus, S., et al., *Adult psychiatric morbidity in England, 2007. Results of a household survey*. 2009, The NHS Information Centre for health and social care: Leeds.
6. Glover, G. and F. Evison, *Use of mental health services by minority ethnic groups in England*. 2010, North East Public Health Observatory: Stockton-on-Tees.
7. Tyrer, P., *The case for cothymia: mixed anxiety and depression as a single diagnosis*. Br J Psychiatry, 2001. **179**: p. 191-3.
8. Meltzer, H., et al., *The prevalence of psychiatric morbidity among adults living in private households.*, in *OPCS Surveys of Psychiatric Morbidity in Great Britain*. 1995, HMSO: London.
9. Hollis, S. and F. Campbell, *What is meant by intention to treat analysis? Survey of published randomised controlled trials*. BMJ, 1999. **319**(7211): p. 670-4.
10. Richards, D.A. and R. Suckling, *Improving access to psychological therapies: Phase IV prospective cohort study*. British Journal of Clinical Psychology, 2009. **48**: p. 377-396.
11. Sproston, K. and J. Nazroo, eds. *Ethnic minority psychiatric illness rates in the community*. 2002, The Stationery Office: London.
12. Fonagy, P., et al., *The mental health outcome measurement initiative. Report from the chair of the Outcomes Reference Group*. 2004, National Collaborating Centre for Mental Health, University College, London. .
13. National Institute for Mental Health in England, *Mental Health Outcomes Compendium*, D.o. Health, Editor. 2008.
14. Wing, J., A. Beevor, and R.H. Curtis, *HoNOS: Health of the Nation Outcome Scales. Report on Research and Development*. 1996, Royal College of Psychiatrists Research Unit, London.

Tables

Table 1. Summary of episode records and study group inclusion

Start date	Initial symptom scores	Study group	Records
In period	At least one	Included	79,310 (57%)
In period	None	Not included	57,975 (42%)
Missing, earlier or erroneous	At least one	Not included	710 (1%)
Missing, earlier or erroneous	None	Not included	546 (0%)
Total episodes			138,541

Table 2. Completeness of recording in data submitted.

Item	Overall completeness	Maximum	90th	75th	Median	25th	10th	Minimum
Gender	97.5%	100.0%	100.0%	99.8%	99.0%	97.3%	72.6%	82.7%
Age	91.7%	100.0%	91.7%	99.9%	99.9%	99.2%	41.7%	0.0%
Ethnicity	69.5%	99.4%	71.7%	87.2%	74.5%	49.5%	0.0%	0.0%
Visual disability	10.9%	100.0%	71.6%	0.0%	0.0%	0.0%	0.0%	0.0%
Speech disability	10.8%	100.0%	71.4%	0.0%	0.0%	0.0%	0.0%	0.0%
Hearing disability	10.9%	100.0%	75.4%	0.0%	0.0%	0.0%	0.0%	0.0%
Mobility disability	19.3%	100.0%	87.5%	30.9%	5.6%	0.4%	0.0%	0.0%
Spoken English	14.0%	97.7%	100.0%	1.0%	0.0%	0.0%	96.9%	0.0%
Referral date	97.2%	100.0%	100.0%	100.0%	100.0%	100.0%	72.6%	0.0%
Assessment date	86.6%	100.0%	98.4%	99.2%	96.9%	93.4%	34.5%	0.0%
First treatment date	74.1%	99.9%	69.9%	93.9%	84.0%	69.4%	30.3%	0.0%
Ending date	50.1%	72.1%	100.0%	63.1%	51.6%	37.2%	94.3%	0.0%
Source of referral	98.3%	100.0%	70.7%	100.0%	99.8%	98.7%	30.3%	82.7%
Reason for ending	52.2%	100.0%	99.8%	66.1%	55.7%	38.7%	86.6%	0.0%
First employment status	96.9%	100.0%	100.0%	99.1%	98.3%	95.0%	91.2%	71.3%
First sick pay status	97.5%	100.0%	100.0%	99.8%	98.8%	97.7%	90.9%	71.3%
First benefit status	97.2%	100.0%	100.0%	99.8%	98.6%	97.1%	80.4%	71.4%
First psychotropic drug status	95.3%	100.0%	73.1%	99.9%	99.9%	99.5%	32.8%	0.0%
Last employment status	55.6%	99.9%	99.7%	65.0%	56.1%	45.1%	38.7%	21.0%
Last sick pay status	58.9%	100.0%	95.5%	69.4%	59.6%	45.1%	33.0%	23.3%
Last benefit status	56.6%	100.0%	84.1%	67.2%	56.7%	44.5%	40.5%	21.0%
Last psychotropic drug status	63.0%	100.0%	93.0%	76.4%	64.6%	52.9%	1.8%	0.0%
Primary diagnosis	67.8%	97.5%	100.0%	80.8%	64.0%	26.2%	99.3%	0.0%
First PHQ-9 rating	99.6%	100.0%	100.0%	100.0%	99.8%	99.7%	98.9%	93.6%
First GAD-7 rating	98.8%	100.0%	99.8%	99.8%	99.5%	99.1%	79.4%	71.8%
First phobia questions	93.5%	100.0%	99.8%	98.5%	96.3%	92.3%	86.7%	70.3%
First W&SAS rating	94.2%	100.0%	78.0%	98.9%	97.5%	94.8%	40.2%	38.4%
Last PHQ-9 rating	58.4%	100.0%	77.9%	65.9%	56.3%	51.3%	39.3%	21.7%
Last GAD-7 rating	58.1%	100.0%	77.4%	65.8%	57.2%	51.1%	27.1%	21.4%
Last phobia questions	52.6%	100.0%	77.3%	61.7%	52.6%	41.8%	32.7%	18.7%
Last WASAS rating	53.2%	100.0%	100.0%	61.9%	53.6%	43.1%	75.2%	12.1%
Contacts by intervention	91.5%	100.0%	100.0%	98.7%	92.4%	85.7%	79.2%	0.0%
Contacts by session purpose	97.1%	100.0%	99.5%	99.8%	99.3%	98.4%	12.4%	0.0%
Contacts by therapist grade	71.9%	100.0%	100.0%	97.7%	84.3%	41.7%	3.7%	0.0%

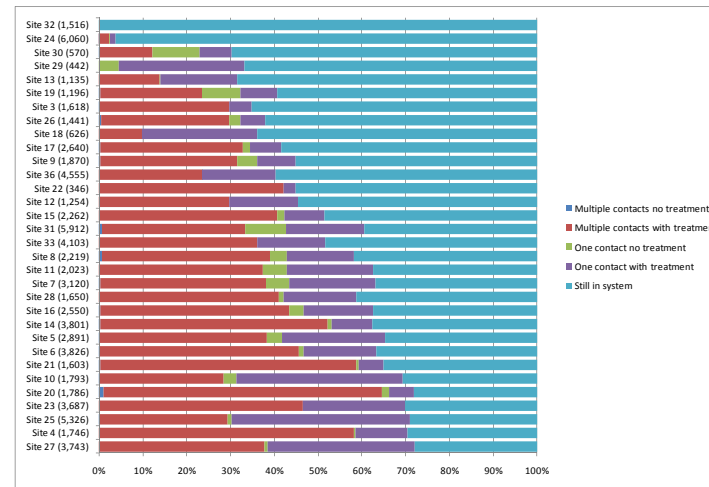
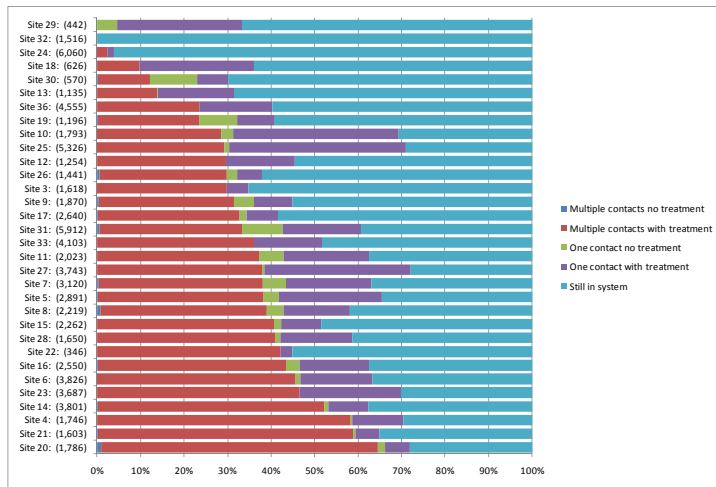
The table shows the overall proportion of study group records with valid codes data in each field, and the percentile points in the distribution of this figure for all 32 sites. These differ from the proportions with 'usable' codes in some cases where list of codes that are technically valid includes some that are uninformative. The most important example is the diagnostic code F99 – Mental illness not otherwise specified, which was assigned as the primary diagnosis for 13.5% of patients.

Table 3. Data availability for simple comparison of patients proceeding from initial contact or referral to initial assessment.

Characteristic	Number or proportion
Total records	137,285
In Study Group	79,310
Not Study Group	57,975
Proportion in study group	57.8%
Usable age group	
Study group	91.7%
Others	99.3%
Usable gender	
Study group	97.5%
Others	92.4%
Usable ethnic category	
Study group	69.5%
Others	28.3%

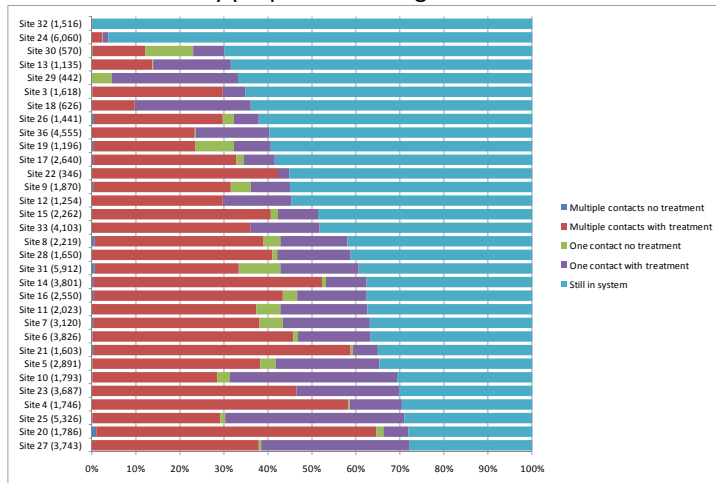
Patients in the 'Study group' had an initial assessment. Others have no record of this.

Charts 1 to 3: Illustration of the variation between sites in proportion of patients at each pathway stage.



1. Ordered by proportion having more than one contact and concluded

2. Ordered by proportion 'treated'



3. Ordered by proportion concluded

Table 4. Age and gender profile of study group patients

Age group	Female	Male	Gender 'Not Specified'	Gender code missing or unrecognised	Total	Gender ratio (F/M)
5-17	1%	1%	~	~	1.0%	2.8
18-24	14%	11%	13%	13%	12.9%	2.4
25-34	23%	22%	20%	19%	22.7%	2.1
35-44	23%	25%	33%	22%	23.9%	1.8
45-54	17%	19%	21%	16%	17.6%	1.7
55-64	9%	10%	9%	10%	9.6%	1.8
65-74	3%	3%	2%	2%	2.9%	2.3
75-84	1%	1%	~	~	1.1%	2.4
Age group missing	8%	8%	~	~	8.3%	1.9
Total	51109	26226	196	1779	79310	1.9

Values in cells marked '~' suppressed to conceal potentially disclosive numbers.

Table 5 and chart 4. Age profile of study group patients. 95% confidence intervals are shown in parentheses in table and as error bars in chart. Chart 5. shows age and sex profiles of rates of common mental disorders from the 2007 Adult Psychiatric Morbidity survey for comparison.

Age group	Females	Males
05 to 17	0.69 (0.64 to 0.76)	0.24 (0.21 to 0.28)
18 to 24	13.83 (13.50 to 14.17)	5.54 (5.33 to 5.75)
25 to 34	16.33 (16.03 to 16.63)	7.69 (7.48 to 7.90)
35 to 44	13.98 (13.72 to 14.24)	7.66 (7.47 to 7.86)
45 to 54	11.82 (11.56 to 12.09)	7.10 (6.89 to 7.30)
55 to 64	7.19 (6.98 to 7.41)	4.20 (4.03 to 4.37)
65 to 74	3.27 (3.10 to 3.44)	1.54 (1.42 to 1.66)
75 to 84	1.45 (1.32 to 1.58)	0.88 (0.76 to 1.00)
85 plus	0.52 (0.41 to 0.65)	0.42 (0.28 to 0.60)
All ages	8.37 (8.29 to 8.45)	4.42 (4.37 to 4.48)

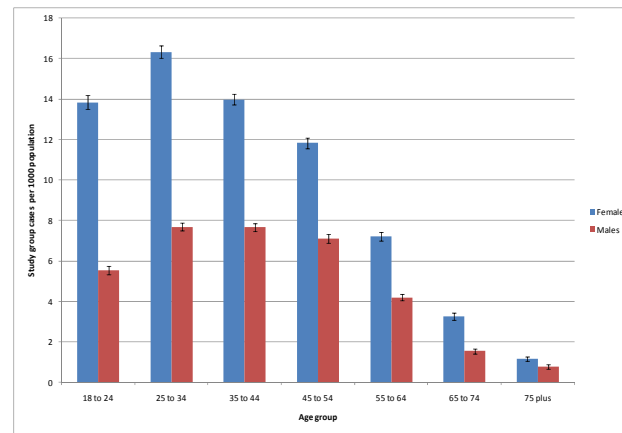


Chart 4. Profile of age specific rates, study group members per 1000 population by gender

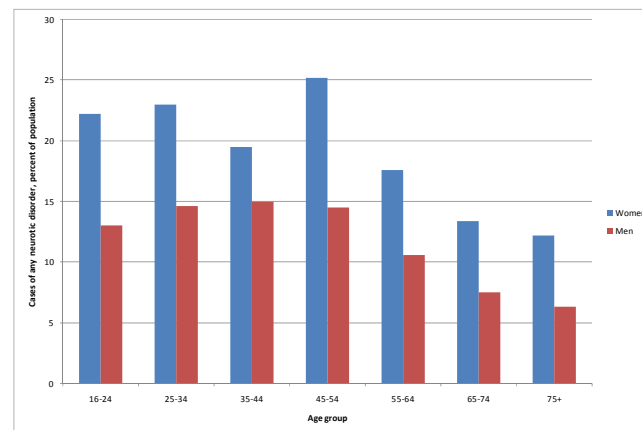


Chart 5. Profile of age specific prevalence of any neurotic disorder, Adult Psychiatric Morbidity Survey 2007

Table 6. Ethnic profile of study group patients by broad age group.

Ethnic category groupings	18 to 34	35 to 64	65 and over
White British	15037 (83.9%)	21897 (86.7%)	1841 (90.3%)
Minority White groups	1207 (6.7%)	1606 (6.4%)	118 (5.8%)
Mixed groups	412 (2.3%)	304 (1.2%)	8 (0.4%)
Asian groups	558 (3.1%)	549 (2.2%)	24 (1.2%)
Black groups	422 (2.4%)	571 (2.3%)	39 (1.9%)
Other groups	285 (1.6%)	324 (1.3%)	8 (0.4%)
Any minority ethnic group	2884 (16.1%)	3354 (13.3%)	197 (9.7%)
Total with usable ethnic code	17921 (70.9%)	25251 (70.8%)	2038 (70.7%)
Total patients	25290	35676	2882

Includes data from all but four sites (see text for explanation). Percentages are by column; for ethnic groupings these are of total with usable code, for penultimate row they are percentages of total patients.

Table 7 Indirectly age –standardised use ratios (comparison of observed/expected) for broad ethnic groups by gender.

Ethnic category	Standardised service use ratio (95% confidence interval)			
	Females		Males	
White British	105.2%	(103.9% to 106.4%)	105.8%	(104.1% to 107.6%)
Minority White	99.9%	(95.7% to 104.3%)	86.9%	(81.3% to 92.8%)
Mixed	109.0%	(99.8% to 119.0%)	105.7%	(92.6% to 120.0%)
Asian	50.6%	(46.9% to 54.5%)	59.4%	(54.0% to 65.2%)
Black	68.1%	(63.3% to 73.2%)	58.5%	(52.1% to 65.5%)
Other	55.1%	(49.8% to 60.8%)	66.2%	(58.0% to 75.3%)

Includes data from all but four sites (see text for explanation).

Table 8. Standardised service use rates by ethnic category with 95% confidence intervals.

Ethnic category	Standardised service use ratio (95% confidence interval)	
	Females	Males
White British	105.1% (103.9% to 106.4%)	105.8% (104.0% to 107.6%)
White Irish	86.1% (77.0% to 96.0%)	90.8% (78.1% to 105.1%)
White Other	102.9% (98.1% to 107.8%)	86.0% (79.8% to 92.6%)
Mixed White and Black Caribbean	135.2% (115.7% to 157.0%)	126.4% (98.7% to 159.4%)
Mixed White and Black African	103.2% (80.8% to 130.0%)	81.8% (54.8% to 117.5%)
Mixed White and Asian	61.1% (48.6% to 75.7%)	79.8% (60.1% to 103.8%)
Mixed Other	136.2% (117.0% to 157.5%)	129.0% (102.4% to 160.3%)
Indian	46.1% (41.2% to 51.4%)	53.2% (46.1% to 61.2%)
Pakistani	42.8% (35.7% to 50.9%)	58.9% (47.6% to 72.1%)
Bangladeshi	41.5% (32.6% to 52.0%)	44.9% (32.5% to 60.4%)
Other Asian	89.8% (76.5% to 104.7%)	91.4% (74.9% to 110.5%)
Black Caribbean	85.0% (76.7% to 94.0%)	73.1% (61.5% to 86.2%)
Black African	43.8% (38.5% to 49.7%)	40.9% (33.8% to 49.1%)
Black Other	135.1% (111.5% to 162.2%)	115.0% (83.9% to 153.9%)
Chinese	20.9% (16.4% to 26.4%)	12.1% (7.5% to 18.5%)
Other	86.8% (77.6% to 96.9%)	119.6% (104.0% to 136.9%)

Includes data from all but four sites (see text for explanation).

Table 9. Sources of referral for members of study group: totals and percentages for all sites combined, and percentile points on distribution for proportion referred from each type of source by site

Referral source	Total for all sites	Site percentiles						
		Maximum	90th	75th	Median	25th	10th	Minimum
General Medical Practitioner	66,297 (83.6%)	99.9%	97.5%	92.0%	87.2%	72.4%	59.0%	-
Self	6,811 (8.6%)	99.8%	27.1%	8.5%	2.1%	0.3%	0.1%	-
Other clinical specialty	2,217 (2.8%)	47.1%	12.5%	4.4%	1.6%	0.1%	0.0%	-
Community/practice nurse/health visitor	991 (1.3%)	10.6%	3.3%	1.2%	0.7%	0.2%	0.0%	-
Local Authority Social Services	88 (0.1%)	0.7%	0.3%	0.1%	0.1%	-	-	-
Police	38 (0.0%)	0.9%	0.2%	-	-	-	-	-
A&E Department	37 (0.0%)	0.5%	0.1%	0.1%	0.0%	-	-	-
Voluntary sector organisation	36 (0.0%)	0.4%	0.1%	0.0%	-	-	-	-
Carer	18 (0.0%)	0.2%	0.0%	-	-	-	-	-
Employer	16 (0.0%)	0.2%	0.1%	0.0%	-	-	-	-
Job centre plus	8 (0.0%)	0.3%	0.0%	-	-	-	-	-
Education Service	6 (0.0%)	0.1%	0.0%	-	-	-	-	-
Courts	4 (0.0%)	0.1%	0.0%	-	-	-	-	-
Probation Service	3 (0.0%)	0.1%	-	-	-	-	-	-
Other	1,399 (1.8%)	11.0%	5.3%	2.3%	1.3%	0.3%	0.1%	-
Usable data	98%	100.0%	100.0%	100.0%	99.8%	98.7%	94.3%	82.7%

Table 10. Economic position and benefit status for members of study group: totals and percentages for all sites combined, and percentile points on distribution for each grouping by site.

	Total for all sites		Percentiles						
			Maximum	90th	0.75	0.5	0.25	10th	Minimum
Employed full-time	30960	39.0%	48.3%	44.0%	42.5%	37.6%	35.7%	31.8%	24.6%
Employed part-time	11648	14.7%	19.0%	18.0%	16.9%	14.9%	11.9%	10.3%	5.8%
<i>All employed</i>	<i>42608</i>	<i>53.7%</i>	<i>62.4%</i>	<i>60.8%</i>	<i>55.9%</i>	<i>51.8%</i>	<i>48.9%</i>	<i>45.7%</i>	<i>37.5%</i>
Unemployed	18873	23.8%	45.9%	32.1%	29.8%	22.4%	19.5%	17.7%	15.0%
<i>All economically active</i>	<i>61481</i>	<i>77.5%</i>	<i>85.3%</i>	<i>83.2%</i>	<i>80.3%</i>	<i>78.5%</i>	<i>74.6%</i>	<i>69.6%</i>	<i>56.6%</i>
Unemployed as % of economically active		30.7%	55.0%	40.2%	37.4%	30.4%	26.2%	22.2%	21.2%
Full-time student	3485	4.4%	8.8%	6.0%	4.9%	3.9%	3.3%	2.8%	1.9%
Full-time homemaker or carer	6442	8.1%	11.2%	10.3%	9.8%	8.2%	6.4%	5.7%	0.0%
Retired	5415	6.8%	10.1%	9.5%	8.2%	6.6%	4.6%	3.7%	2.8%
<i>All economically inactive</i>	<i>15342</i>	<i>19.3%</i>	<i>24.7%</i>	<i>22.1%</i>	<i>21.1%</i>	<i>19.8%</i>	<i>16.1%</i>	<i>13.7%</i>	<i>7.6%</i>
Missing data	2487	3.1%	28.7%	13.4%	5.0%	1.7%	0.9%	0.2%	0.0%
Receiving benefits	23723	29.9%	51.0%	37.6%	32.0%	29.1%	25.5%	21.2%	20.4%
No benefits	50822	64.1%	78.8%	74.8%	67.8%	63.7%	58.5%	51.8%	40.2%
Benefit status not known	4765	6.0%	29.2%	12.1%	9.2%	4.6%	2.5%	1.5%	0.3%
Total cases	79310								

Rows in italics are sub-totals.

Table 11. Primary diagnoses. Numbers and proportions with usable primary diagnosis, proportion with each diagnosis, all study group members. Percentile points for corresponding figures for sites.

Diagnosis	Percentage/ numbers for all sites combined.	Percentiles						
		Maximum	90th	75th	Median	25th	10th	Minimum
Depressive episode	29.4%	52.3%	39.2%	32.4%	27.3%	21.2%	15.7%	11.1%
Mixed anxiety and depressive disorder	29.3%	49.7%	40.2%	36.8%	29.8%	23.0%	16.2%	0.0%
Generalized anxiety disorder	17.5%	36.8%	23.8%	21.2%	17.4%	10.6%	7.1%	0.0%
Recurrent depressive disorder	6.9%	13.3%	11.2%	9.5%	7.9%	5.1%	3.3%	0.0%
Obsessive-compulsive disorder	2.2%	8.1%	4.1%	3.5%	2.1%	1.2%	0.5%	0.0%
Posttraumatic stress disorder	2.1%	7.8%	4.6%	2.9%	1.5%	1.0%	0.6%	0.0%
Disappearance and death of family member	2.0%	8.3%	3.3%	2.5%	1.2%	0.2%	0.0%	0.0%
Agoraphobia	1.5%	15.3%	4.3%	2.5%	1.5%	0.6%	0.0%	0.0%
Social phobias	1.5%	11.1%	4.5%	2.3%	1.4%	0.5%	0.2%	0.0%
Specific phobias	1.2%	22.2%	2.5%	1.8%	1.4%	0.6%	0.2%	0.0%
Eating disorders	0.6%	1.8%	1.0%	0.7%	0.4%	0.2%	0.0%	0.0%
Mental or behavioural disorders due to alcohol	0.5%	1.4%	0.8%	0.6%	0.5%	0.2%	0.0%	0.0%
Somatoform disorders	0.4%	3.9%	1.0%	0.7%	0.4%	0.1%	0.0%	0.0%
Bipolar affective disorder	0.3%	1.7%	0.6%	0.4%	0.2%	0.1%	0.0%	0.0%
Other specified mental disorder	4.2%	63.6%	6.2%	2.2%	1.4%	0.5%	0.0%	0.0%
Other diagnosis	0.4%	7.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
% with usable diagnosis	54.4%	97.5%	93.1%	80.8%	64.1%	26.3%	1.7%	0.0%
Records with usable diagnosis	43111							
Total records	79310							

Chart 5a Diagnostic profiles to show variation between sites. Bars show the proportion of cases with a usable diagnosis in each category. Numbers in the row labels indicate the number and proportion of cases from each site for which usably coded diagnoses were available.

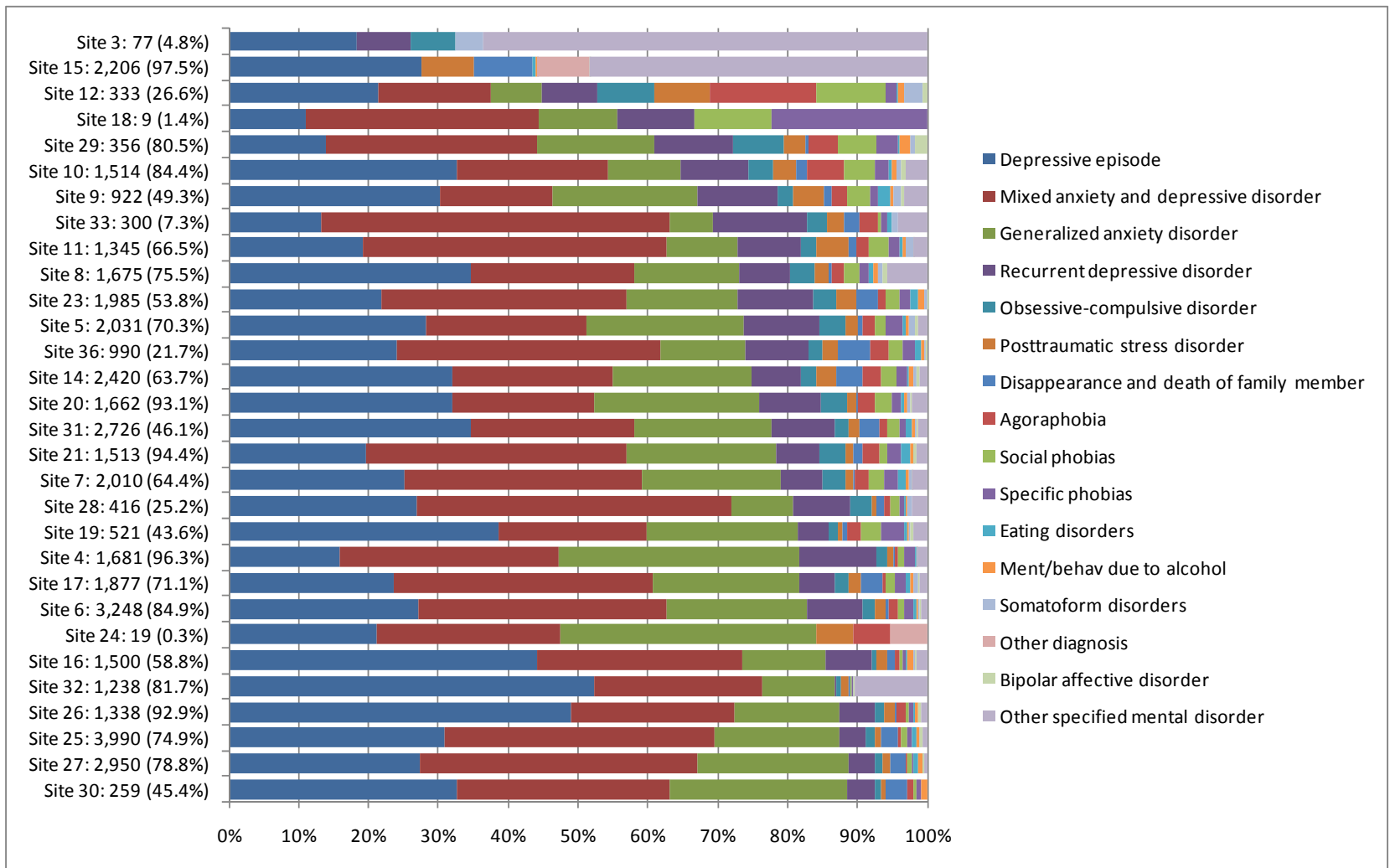


Chart 5b and 5c. Profiles of initial PHQ-9 and GAD-7 for study group patients by primary diagnosis.

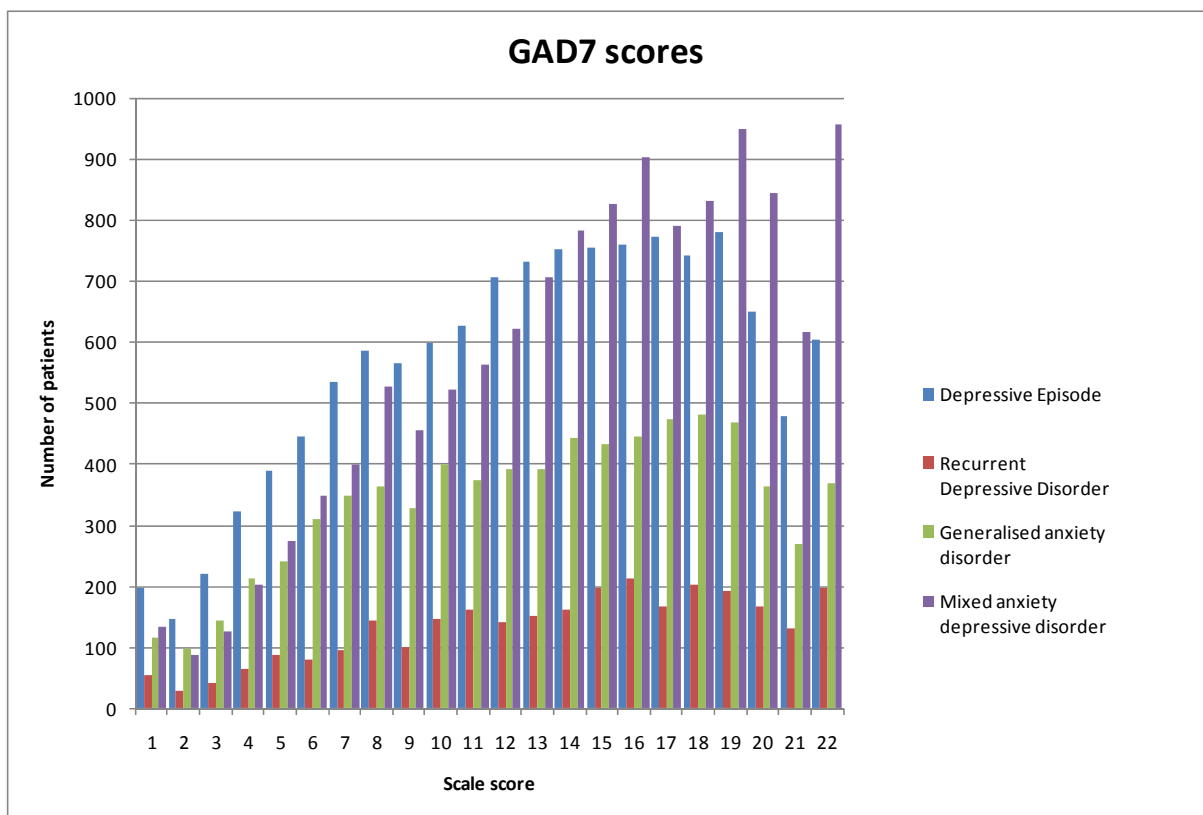
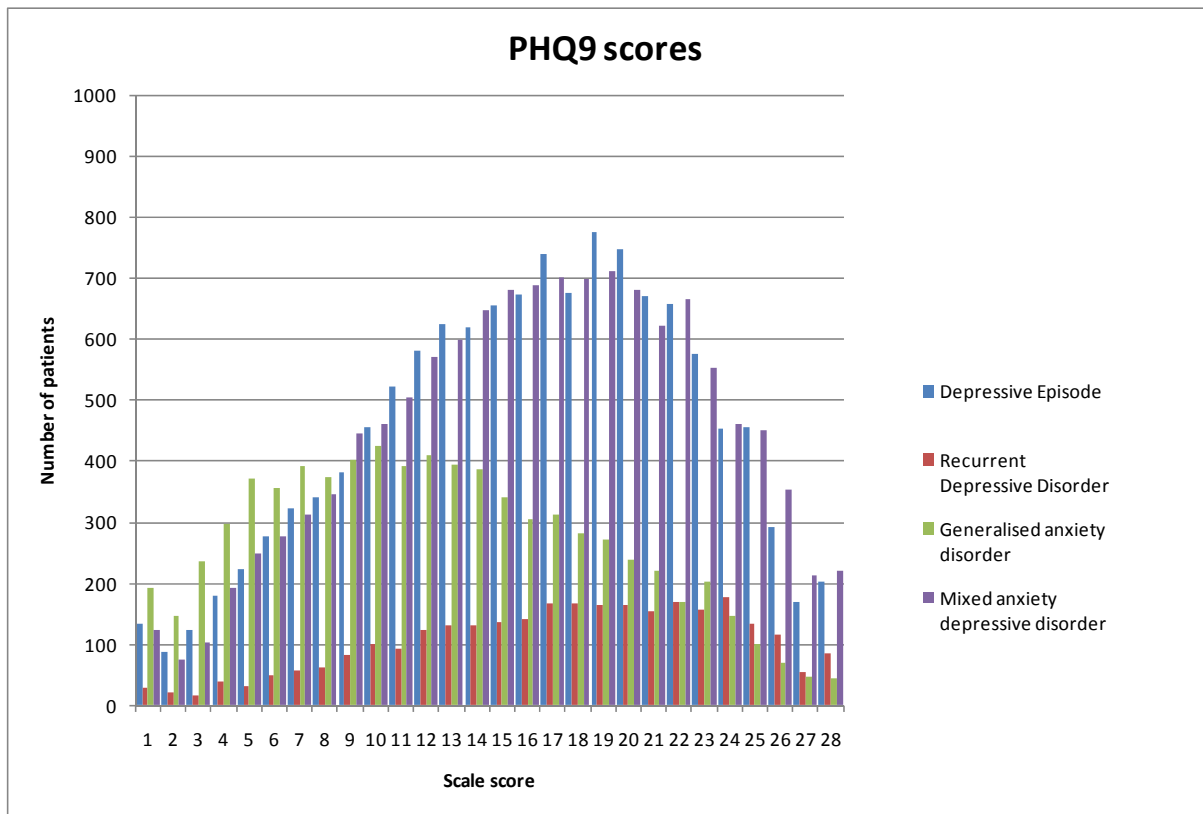


Table 11a Frequency of secondary diagnoses for study group patients assigned primary diagnoses of depressive or anxiety conditions; average PHQ-9 and GAD-7 scores for these groupings.

Secondary diagnosis	Primary diagnosis											
	Depressive episode			Recurrent depressive disorder			Generalized anxiety disorder			Mixed anxiety and depressive disorder		
	Average PHQ-9	Average GAD-7	Cases	Average PHQ-9	Average GAD-7	Cases	Average PHQ-9	Average GAD-7	Cases	Average PHQ-9	Average GAD-7	Cases
Generalized anxiety disorder	14.50	12.46	48	16.22	13.24	68	9.75	11.38	8	14.17	13.83	6
Depressive episode	17.00	12.00	12	14.25	7.25	4	14.97	13.05	60	13.85	12.69	13
Recurrent depressive disorder	16.67	10.83	6	0.00	1.00	2	14.00	14.87	23	14.75	14.50	4
Mixed anxiety and depressive disorder	16.69	14.31	36	17.00	12.50	18	15.00	16.00	17	14.63	12.75	8
Obsessive-compulsive disorder	15.35	13.78	23	17.46	14.15	13	10.24	12.50	38	17.89	15.64	36
Social phobias	17.33	13.38	24	15.94	11.76	17	11.14	12.00	22	17.07	14.90	29
Family Loss	15.44	11.62	34	14.00	12.42	12	13.45	14.27	11	16.50	13.77	26
Other specified mental disorder	21.08	16.08	13	19.50	17.75	4	14.91	14.74	23	16.77	15.15	26
Agoraphobia	17.29	10.43	7	21.50	17.13	8	18.08	16.77	13	17.00	15.31	29
Eating disorders	17.44	12.22	18	18.75	13.67	12	11.90	9.30	10	16.94	14.06	17
Posttraumatic stress disorder	16.87	15.09	23	15.00	13.00	4	13.63	13.50	8	18.12	15.41	17
Ment/behav due to alcohol	15.92	12.42	12	20.00	13.50	6				16.13	13.88	16
Specific phobias	8.67	9.33	3	12.50	11.50	2	9.85	14.38	13	13.13	12.50	8
Somatoform disorders	15.80	10.20	5	15.00	7.00	1	15.86	14.00	7	17.40	16.80	5
Bipolar affective disorder	18.00	12.50	2	22.00	15.00	1	13.50	15.00	2			
Other diagnosis	16.64	13.21	229				19.00	7.00	1			
Mental disorder NOS	15.59	11.56	34	18.85	14.46	13	13.70	12.40	10	14.30	12.35	20
No secondary code	15.00	11.82	12,155	16.19	12.57	2,783	11.33	11.84	7,288	15.11	13.18	12,367
% no secondary code			95.8%			93.8%			96.5%			97.9%

Table 11b

Measures	Diagnostic grouping - primary and secondary diagnoses				
	Anxiety	Anxiety and Depression	Depression	Mixed anxiety and depressive disorder	None of these
Patients	7,627	199	16,291	12,712	42,481
Valid PHQ-9	7,615	199	16,202	12,691	42,273
Mean (s.d.) first PHQ-9	11.38 (6.50)	15.17 (6.17)	15.37 (6.41)	15.14 (6.37)	13.48 (6.93)
Median (IQR)	11 (6 to 16)	16 (11 to 20)	16 (11 to 20)	16 (11 to 20)	14 (8 to 19)
PHQ-9 case level	57.4% (4,373)	80.9% (161)	81.1% (13,140)	79.5% (10,093)	69.5% (29,370)
Valid GAD-7	7,544	199	15,921	12,560	42,128
Mean (s.d.) first GAD-7	11.90 (5.69)	13.18 (5.02)	12.13 (5.80)	13.21 (5.49)	12.02 (5.83)
Median (IQR)	13 (8 to 17)	14 (10 to 17)	13 (8 to 17)	14 (10 to 18)	13 (8 to 17)
GAD-7 case level	75.4% (5,690)	85.4% (170)	78.2% (12,449)	83.2% (10,450)	75.5% (31,789)
Valid Social phobia Q1	7,218	192	14,973	11,859	40,113
Mean (s.d.) first rating Social phobia Q1	2.41 (2.55)	2.90 (2.52)	2.70 (2.73)	2.88 (2.74)	2.62 (2.69)
Social phobia Q1 case level	30.5% (2,199)	38.5% (74)	38.9% (5,818)	40.0% (4,740)	34.8% (13,964)
Valid Agoraphobia Q2	7,207	192	14,954	11,836	40,039
Mean (s.d.) first rating Agoraphobia Q2	2.34 (2.73)	2.36 (2.60)	1.86 (2.60)	2.36 (2.75)	2.20 (2.75)
Agoraphobia Q2 case level	32.9% (2,374)	30.7% (59)	27.6% (4,120)	35.0% (4,141)	31.4% (12,576)
Valid Specific phobia Q3	7,204	192	14,944	11,823	40,023
Mean (s.d.) first rating Specific phobia Q3	1.98 (2.65)	1.86 (2.60)	1.62 (2.56)	1.99 (2.72)	1.97 (2.73)
Specific phobia Q3 case level	27.1% (1,951)	25.5% (49)	23.1% (3,455)	27.6% (3,260)	26.9% (10,763)

Table 12. Proportions of study group patients with complete sets of ratings on each of the symptom rating scales and the Work and Social Adjustment Scale (W&SAS); percentile points for the distribution of complete scores between sites. (95% confidence intervals for overall percentage scores are given in parentheses).

Rating	Overall score, all sites combined	Percentiles for site scores						
		Maximum	90th	75th	Median	25th	10th	Minimum
Availability of initial ratings								
GAD-7	98.8% (98.7% to 98.9%)	100.0%	100.0%	99.8%	99.5%	99.1%	98.9%	71.8%
PHQ-9	99.6% (99.5% to 99.6%)	100.0%	100.0%	100.0%	99.8%	99.7%	99.3%	93.6%
Phobia Qs	93.5% (93.3% to 93.6%)	100.0%	99.8%	98.5%	96.3%	92.3%	79.4%	70.3%
W&SAS	94.2% (94.0% to 94.3%)	100.0%	99.8%	98.9%	97.5%	94.8%	86.7%	38.4%
GAD-7 and PHQ	98.6% (98.5% to 98.7%)	100.0%	100.0%	99.7%	99.3%	99.0%	98.0%	71.6%
GAD-7, PHQ-9 and Phobia Qs	93.0% (92.8% to 93.2%)	100.0%	99.8%	98.3%	95.9%	90.9%	78.9%	70.1%
Caseness where rateable								
First GAD-7 8+	77.3% (77.0% to 77.6%)	89.5%	79.6%	78.7%	77.6%	75.1%	73.8%	58.6%
First PHQ-9 10+	72.3% (72.0% to 72.7%)	84.6%	77.3%	74.1%	71.8%	68.8%	66.4%	54.6%
First phobia Qs Positive	53.3% (53.0% to 53.7%)	70.5%	62.0%	56.2%	52.7%	50.4%	48.2%	39.4%
MTR2 (any of GAD-7, PHQ-9 or Phobia Qs)	88.1% (87.8% to 88.3%)	96.1%	91.2%	89.5%	88.0%	86.5%	84.5%	68.4%
MTR1 (either of GAD-7 or PHQ)	83.6% (83.3% to 83.9%)	93.2%	86.5%	84.9%	83.9%	81.5%	79.7%	63.8%
Rating of W&SAS where available								
Any problems sub-clinical	23.7% (23.4% to 24.0%)	41.8%	28.1%	26.6%	23.7%	20.3%	17.0%	15.0%
Significant functional impairment	40.7% (40.4% to 41.1%)	45.3%	44.7%	43.0%	40.8%	38.1%	33.7%	26.2%
Moderately severe problems or worse	35.6% (35.3% to 36.0%)	58.7%	48.7%	38.2%	34.1%	31.2%	29.5%	20.4%

Table 12a. Social phobia rating scale scores (mean with s.d. and proportion reaching case level) by primary diagnosis; numbers in each category with social phobia as secondary diagnosis.

Primary diagnosis	Social phobia rating (Phobia question 1)			
	Patients	Mean (sd)	Case level for Social phobia	Secondary diagnosis social phobia
Social phobias	650	4.872 (2.427)	440 (67.7%)	1
Agoraphobia	635	4.090 (2.940)	342 (53.9%)	10
Specific phobias	519	2.139 (2.509)	125 (24.1%)	0
Depressive episode	11,600	2.873 (2.567)	4,230 (36.5%)	20
Mixed anxiety and depressive disorder	11,777	3.149 (2.619)	4,692 (39.8%)	29
Generalized anxiety disorder	7,148	2.590 (2.473)	2,162 (30.2%)	21
Recurrent depressive disorder	2,779	3.390 (2.607)	1,208 (43.5%)	17
Obsessive-compulsive disorder	901	2.690 (2.578)	280 (31.1%)	4
Posttraumatic stress disorder	830	3.370 (2.942)	362 (43.6%)	3
Family Loss	795	2.566 (2.634)	261 (32.8%)	0
Eating disorders	244	2.951 (2.515)	85 (34.8%)	0
Ment/behav due to alcohol	193	3.166 (2.786)	86 (44.6%)	0
Somatoform disorders	170	2.141 (2.357)	42 (24.7%)	1
Bipolar affective disorder	107	3.187 (2.778)	44 (41.1%)	0
Other specified mental disorder	1,728	3.367 (2.873)	823 (47.6%)	3
Other diagnosis	147	2.429 (2.730)	51 (34.7%)	0
Mental disorder NOS	10,034	2.775 (2.572)	3,386 (33.7%)	1
Missing or illegal code	24,098	2.789 (2.578)	8,176 (33.9%)	2
Diagnoses combined	74,355	2.901 (2.606)	26,795 (36.0%)	112

Table 12 b Agoraphobia rating scale scores (mean with s.d. and proportion reaching case level) by primary diagnosis; numbers in each category with agoraphobia as secondary diagnosis.

Primary diagnosis	Agoraphobia rating (Phobia question 2)			
	Patients	Mean (sd) Agoraphobia	Case level for Agoraphobia	Secondary Diagnosis Agoraphobia
Social phobias	646	3.670 (2.798)	334 (51.7%)	7
Agoraphobia	633	5.498 (2.621)	487 (76.9%)	0
Specific phobias	518	3.977 (3.086)	282 (54.4%)	0
Depressive episode	11,592	1.949 (2.474)	2,895 (25.0%)	7
Mixed anxiety and depressive disorder	11,754	2.604 (2.691)	4,098 (34.9%)	29
Generalized anxiety disorder	7,137	2.517 (2.667)	2,323 (32.5%)	12
Recurrent depressive disorder	2,768	2.421 (2.618)	885 (32.0%)	8
Obsessive-compulsive disorder	897	2.414 (2.721)	280 (31.2%)	2
Posttraumatic stress disorder	829	3.476 (3.026)	400 (48.3%)	3
Family Loss	793	2.048 (2.602)	214 (27.0%)	1
Eating disorders	243	1.778 (2.303)	52 (21.4%)	1
Ment/behav due to alcohol	191	2.801 (2.869)	77 (40.3%)	1
Somatoform disorders	169	2.272 (2.478)	50 (29.6%)	0
Bipolar affective disorder	107	2.355 (2.813)	35 (32.7%)	0
Other specified mental disorder	1,728	3.739 (3.033)	915 (53.0%)	21
Other diagnosis	147	2.075 (2.723)	46 (31.3%)	0
Mental disorder NOS	9,998	2.210 (2.611)	2,848 (28.5%)	0
Missing or illegal code	24,078	2.256 (2.636)	7,049 (29.3%)	0
Diagnoses combined	74,228	2.388 (2.680)	23,270 (31.3%)	92

Table 12c Specific phobias rating scale scores (mean with s.d. and proportion reaching case level) by primary diagnosis; numbers in each category with specific phobia as secondary diagnosis.

Primary diagnosis	Specific phobias rating (Phobia question 3)			
	Patients	Mean (sd) Specific phobia	Case level for Specific phobia	Secondary Diagnosis Specific phobia
Social phobias	646	2.567 (2.742)	213 (33.0%)	5
Agoraphobia	632	4.253 (3.075)	368 (58.2%)	5
Specific phobias	518	4.820 (3.018)	328 (63.3%)	0
Depressive episode	11,589	1.744 (2.472)	2,482 (21.4%)	2
Mixed anxiety and depressive disorder	11,741	2.202 (2.681)	3,225 (27.5%)	7
Generalized anxiety disorder	7,134	2.134 (2.613)	1,905 (26.7%)	13
Recurrent depressive disorder	2,763	2.095 (2.633)	724 (26.2%)	2
Obsessive-compulsive disorder	898	2.821 (2.864)	322 (35.9%)	6
Posttraumatic stress disorder	829	3.250 (3.103)	368 (44.4%)	1
Family Loss	793	1.831 (2.512)	185 (23.3%)	0
Eating disorders	244	1.807 (2.454)	54 (22.1%)	2
Ment/behav due to alcohol	191	2.136 (2.707)	55 (28.8%)	0
Somatoform disorders	169	2.101 (2.449)	42 (24.9%)	0
Bipolar affective disorder	106	1.868 (2.651)	24 (22.6%)	0
Other specified mental disorder	1,726	2.841 (3.108)	674 (39.0%)	6
Other diagnosis	147	1.871 (2.778)	34 (23.1%)	0
Mental disorder NOS	9,996	1.923 (2.566)	2,349 (23.5%)	3
Missing or illegal code	24,064	2.061 (2.647)	6,126 (25.5%)	0
Diagnoses combined	74,186	2.103 (2.666)	19,478 (26.3%)	52

Table 12d Sensitivity, specificity, and positive and negative predictive values (PPV and NPV) of phobia questions measured against primary or secondary diagnosis indicating corresponding condition for all possible values of phobia question scores. Study group patients with usable diagnoses and valid phobia ratings only.

Social phobia question				
Cut point	Sensitivity	Specificity	PPV	NPV
0	1	0	1.0%	-
1	93.8%	27.0%	1.3%	99.8%
2	91.5%	34.3%	1.4%	99.7%
3	80.2%	54.8%	1.8%	99.6%
4	69.0%	64.3%	2.0%	99.5%
5	57.4%	74.1%	2.2%	99.4%
6	51.6%	78.2%	2.4%	99.4%
7	33.4%	87.2%	2.6%	99.2%
8	17.1%	92.1%	2.2%	99.1%

Agoraphobia question				
Cut point	Sensitivity	Specificity	PPV	NPV
0	1	0	1.0%	-
1	91.7%	41.3%	1.5%	99.8%
2	89.7%	49.8%	1.7%	99.8%
3	81.8%	62.9%	2.1%	99.7%
4	77.0%	69.1%	2.4%	99.7%
5	64.8%	78.0%	2.8%	99.6%
6	60.6%	81.6%	3.1%	99.5%
7	48.0%	88.7%	4.0%	99.4%
8	35.7%	93.0%	4.8%	99.3%

Specific phobia question				
Cut point	Sensitivity	Specificity	PPV	NPV
0	1	0	0.8%	-
1	85.1%	47.2%	1.2%	99.8%
2	81.1%	56.4%	1.4%	99.7%
3	68.9%	68.7%	1.7%	99.7%
4	63.2%	74.0%	1.8%	99.6%
5	55.4%	80.9%	2.2%	99.6%
6	52.3%	83.4%	2.4%	99.6%
7	42.8%	88.9%	2.9%	99.5%
8	32.5%	92.2%	3.1%	99.4%

Table 12e. Phobia questions as 'backstops'. The table shows the number of patients with a primary diagnosis indicating one of the phobia groups, the number (and percentage) failing to reach case

level on either PHQ-9 or GAD-7, and the number (and percentage) of these rating positive on each of the phobia questions. Study group patients with valid PHQ-9, GAD-7 and phobia ratings only.

Primary diagnosis	Patients	Not PHQ-9 or GAD-7 case	But positive rating on:		
			Social phobia Question 1	Agoraphobia Question 2	Specific phobia Question 3
Social phobias	644	143 (22.2%)	74 (51.7%)	45 (31.5%)	37 (25.9%)
Agoraphobia	632	102 (16.1%)	19 (18.6%)	51 (50.0%)	43 (42.2%)
Specific phobias	514	225 (43.8%)	29 (12.9%)	93 (41.3%)	134 (59.6%)

Table 13. Pattern of co-morbidity based on caseness measures using GAD-7, PHQ-9 and the Phobia questions, study group members with complete initial ratings on all three.

Pattern of caseness	% of patients
Anxiety alone	6.5%
Depression alone	3.9%
Phobia alone	4.4%
Anxiety and depression	24.3%
Anxiety and phobia	4.9%
Depression and phobia	2.4%
Anxiety, depression and phobia	41.6%
No case level rating	11.9%
Patients included	73,750

Table 15. Proportion of study group patients reaching case-level ratings for each scale and of having no case level rating by gender.

	Female	Male	Total	N	P
PHQ-9	72.9	71.7	72.5	77,011	0.001
GAD-7	78.0	76.1	77.4	76388	0.000
Phobia Qs	53.8	52.9	53.5	72267	0.016
MTR2 case	88.8	87.0	88.2	71906	0.000
MTR1 case	84.3	82.5	83.7	76250	0.000
No case level rating	11.2	13.0	11.8	71906	0.000
W&SAS					
At least significant functional impairment	76.2	76.7	76.4	72819	ns
Moderately severe problems or worse	34.9	37.2	35.7	72819	0.000

Significance ratings are from Pearson Chi Square tests. Omits patients without gender coding. To make comparisons interpretable, for the W&SAS groupings the lower level grouping incorporates the higher level, to give the proportion with scores of 10 or more.

Table 16. Patient-level co-morbidity by gender. Table shows proportion of patients reaching case-level scores on 0, 1, 2 and 3 of the rating instruments.

Number of individual scale ratings at case level	Male	Female	Total
0	13.01	11.2	11.82
1	14.5	14.97	14.81
2	31.57	31.77	31.7
3	40.93	42.05	41.67
n	24,370	47,536	71,906

Pearson Chi Square = 51.9, df = 3 Pr = 0.000 Omits patients without gender coding.

Tables 17. Proportion of study group patients reaching case-level ratings for each scale and of having no case level rating by broad age group.

	Under18	18to34	35to64	65Plus	Total	n	P
PHQ-9	71.5	73.3	73.2	56.7	72.5	72,368	0.000
GAD-7	77.7	79.6	77.1	63.0	77.4	71,775	0.000
Phobia Qs	55.4	54.2	54.3	41.3	53.8	67,769	0.000
MTR2 case	89.3	89.7	88.0	77.7	88.3	67,415	0.000
MTR1 case	83.9	85.3	83.6	71.4	83.8	71,635	0.000
No case level rating	10.8	10.3	12.0	22.3	11.8	67,415	0.000
W&SAS							
At least Significant functional impairment	74.7	78.7	77.1	55.1	76.8	68176	0.000
Moderately severe problems or worse	25.8	35.8	38.0	17.5	36.1	68176	0.000

Significance ratings are from Pearson Chi Square tests. Omits patients without age coding. To make comparisons interpretable, for the W&SAS groupings the lower level grouping incorporates the higher level, to give the proportion with scores of 10 or more.

Table 18. Patient-level co-morbidity by broad age group. Table shows proportion of patients reaching case-level scores on 0, 1, 2 and 3 of the rating instruments.

Number of individual scale ratings at case level	Under18	18to34	35to64	65Plus	Total
0	10.75	10.33	11.97	22.26	11.75
1	15.73	14.7	14.43	20.74	14.82
2	32.8	32.65	30.81	31.04	31.56
3	40.73	42.31	42.79	25.96	41.87
N	100	100	100	100	100

Pearson chi square 583.0 df = 2 Pr = 0.000. Omits patients without age coding.

Tables 19. Proportion of study group patients reaching case-level ratings for each scale and of having no case level rating by broad ethnic group.

	White British	Minority White	Mixed	Asian	Black	Other	Total	n	P
PHQ-9	72.4	71.2	77.0	79.2	77.6	74.4	72.7	54,839	0.000
GAD-7	77.5	78.1	80.6	80.9	78.6	78.7	77.7	54,471	0.017
Phobia Qs	53.5	57.5	58.8	60.4	62.8	59.2	54.3	51,580	0.000
MTR2	88.5	89.0	90.6	91.3	90.2	87.8	88.6	51,294	0.007
MTR1	83.9	83.9	87.1	87.5	86.0	83.8	84.1	54,378	0.001
No case level rating	11.6	11.0	9.4	8.7	9.8	12.2	11.4	51,294	0.007
W&SAS									
At least significant functional impairment	76.2	79.9	82.3	81.5	79.9	77.6	76.7	51,850	0.000
Moderately severe problems or worse	34.6	42.1	42.1	47.7	45.9	40.7	35.8	51,850	0.000

Significance ratings are from Pearson Chi Square tests. Omits patients without ethnic category coded. To make comparisons interpretable, for the W&SAS groupings the lower level grouping incorporates the higher level, to give the proportion with scores of 10 or more.

Table 20. Patient-level co-morbidity by broad ethnic group. Table shows proportion of patients reaching case-level scores on 0, 1, 2 and 3 of the rating instruments.

Number of individual scale ratings at case level	White British	Minority White	Mixed	Asian	Black	Other	Total
0	11.55	11.04	9.44	8.68	9.84	12.2	11.4
1	14.99	16.02	12.45	12.11	12.8	12.52	14.87
2	32.01	28.03	31.06	29.55	26.65	26.78	31.54
3	41.45	44.91	47.05	49.66	50.72	48.49	42.2
N	44,681	3,008	763	1,164	1,047	631	51,294

Pearson chi square = 115., df=15 p< 0.0001. Omits patients without ethnic category coded.

Tables 21. Proportion of study group patients reaching case-level ratings for each scale and of having no case level rating by broad referral source.

	GP	Self	Other	Total	n	P
PHQ-9	72.6	70.3	73.3	72.4	77,642	0.000
GAD-7	77.5	75.3	77.7	77.4	77,017	0.000
Phobia Qs	53.0	53.5	60.3	53.5	72,850	0.000
MTR2	88.3	86.3	89.0	88.2	72,486	0.000
MTR1	83.9	81.3	83.8	83.7	76,877	0.000
No case level rating	11.7	13.8	11.0	11.9	72,486	0.000
W&SAS						
At least significant functional impairment	76.2	77.7	77.0	76.4	73,403	0.016
Moderately severe problems or worse	35.1	37.4	40.0	35.6	26,159	0.000

Significance ratings are from Pearson chi square tests. Omits patients without valid coding for source of referral. To make comparisons interpretable, for the W&SAS groupings the lower level grouping incorporates the higher level, to give the proportion with scores of 10 or more.

Table 22. Patient-level co-morbidity by broad referral source. Table shows proportion of patients reaching case-level scores on 0, 1, 2 and 3 of the rating instruments.

Number of individual scale ratings at case level	GP	Self	Other	Total
0	11.73	13.75	11.01	11.85
1	14.98	14.5	13.55	14.85
2	31.95	30.6	28.76	31.64
3	41.34	41.15	46.68	41.66
N	62,087	5,847	4,552	72,486

Pearson chi square = 73.0149, df = 6, p = 0.000. Omits patients without valid coding for source of referral.

Charts 6 to 9. Histograms to show the distributions of each of the symptom scales and the Work and social Adjustment Scale at first assessments.

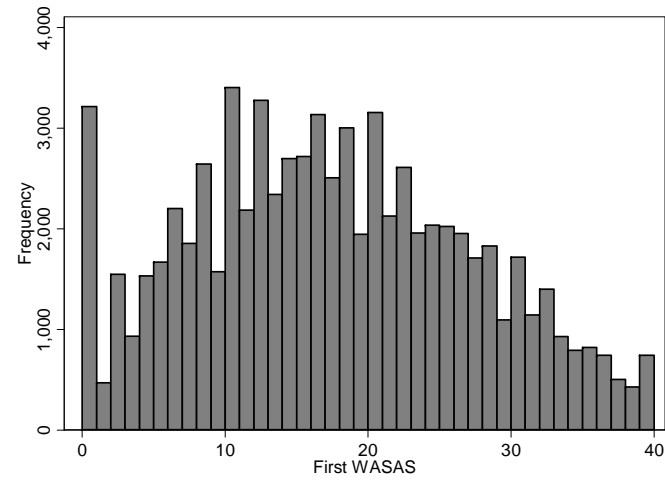
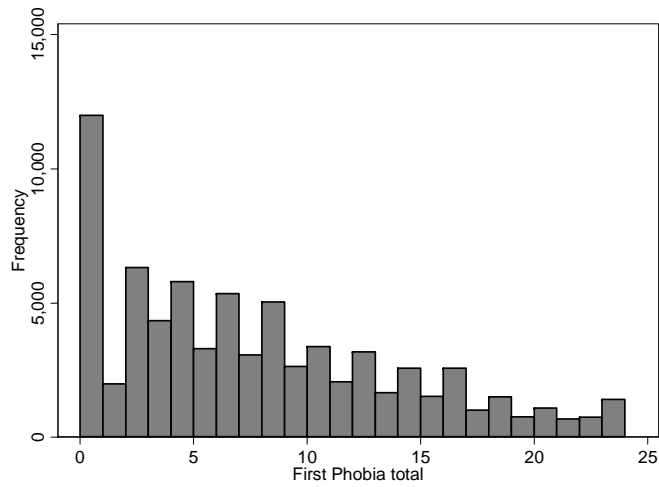
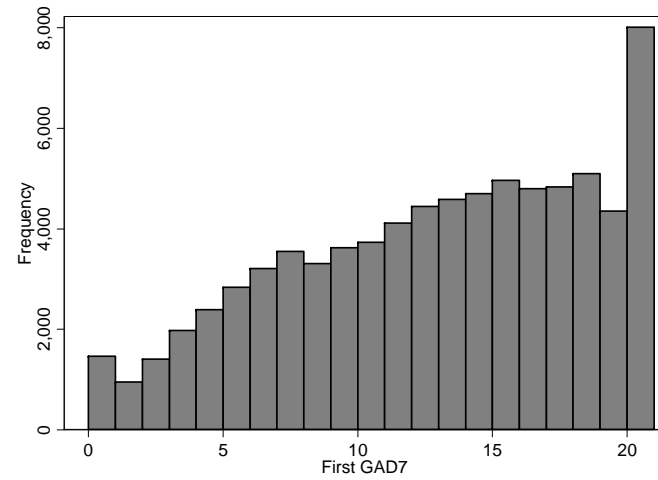
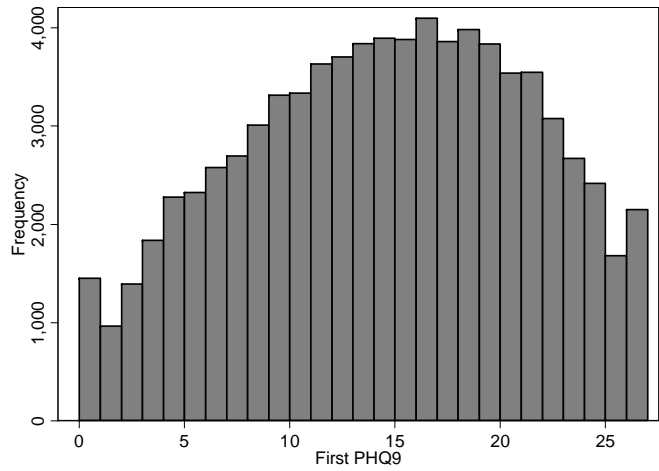


Table 23. Percentile points on distribution of PHQ-9 scores between patient subgroups by gender, age, and broad ethnicity and referral source.

Percentile	Male	Female	Percentile	Under18	18to34	35to64	65Plus
0	0 (0 - 0)	0 (0 - 0)	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)
5	2 (2 - 3)	3 (3 - 3)	5	3 (2 - 3)	3 (3 - 3)	3 (3 - 3)	1 (0 - 1)
25	9 (9 - 9)	9 (9 - 9)	25	9 (8 - 9)	9 (9 - 9)	9 (9 - 9)	6 (6 - 6)
50	14 (14 - 14)	14 (14 - 14)	50	14 (13 - 15)	14 (14 - 14)	15 (15 - 15)	11 (10 - 11)
75	19 (19 - 19)	19 (19 - 19)	75	19 (18 - 19)	19 (19 - 19)	20 (20 - 20)	16 (15 - 16)
95	24 (24 - 24)	25 (24 - 25)	95	24 (23 - 24)	24 (24 - 24)	25 (25 - 25)	23 (22 - 23)
100	27 (27 - 27)	27 (27 - 27)	100	27 (27 - 27)	27 (27 - 27)	27 (27 - 27)	27 (27 - 27)
N	26109	50902	n	797	28144	40253	3174
Kruskal-Wallis Chi Square = 14.6, df = 1, p=0.0001			Kruskal-Wallis Chi Square = 678.6, df = 3, p=0.0000				

Percentile	GP	Self	Other	Percentile	White British	Minority White	Mixed	Asian	Black	Other
0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)
5	3 (3 - 3)	2 (2 - 2)	2 (2 - 3)	5	3 (3 - 3)	2.5 (2 - 3)	3 (2 - 4)	4 (3 - 4)	3 (2 - 3)	3 (2 - 4)
25	9 (9 - 9)	8 (8 - 9)	9 (9 - 9)	25	9 (9 - 9)	9 (8 - 9)	10 (9 - 11)	11 (10 - 11)	10 (10 - 11)	9 (9 - 10)
50	14 (14 - 14)	14 (14 - 14)	15 (15 - 15)	50	14 (14 - 14)	15 (14 - 15)	16 (15 - 16)	16 (16 - 17)	16 (15 - 16)	15 (14 - 16)
75	19 (19 - 19)	19 (19 - 19)	20 (20 - 20)	75	19 (19 - 19)	19 (19 - 20)	20 (19 - 20)	21 (20 - 21)	20 (20 - 21)	20 (20 - 21)
95	24 (24 - 25)	24 (24 - 24)	25 (25 - 25)	95	24 (24 - 24)	25 (24 - 25)	25 (24 - 25)	25 (25 - 26)	25 (24 - 25)	25 (25 - 26)
100	27 (27 - 27)	27 (27 - 27)	27 (27 - 27)	100	27 (27 - 27)	27 (27 - 27)	27 (27 - 27)	27 (27 - 27)	27 (27 - 27)	27 (27 - 27)
n	66008	6797	4837	n	47826	3149	816	1251	1113	684
Kruskal-Wallis Chi Square = 29.7, df = 2, p=0.0000				Kruskal-Wallis Chi Square = 100.6, df = 5, p=0.0000						

Figures in parentheses are 95% confidence intervals

Table 24. Percentile points on distribution of GAD-7 scores between patient subgroups by gender, age, and broad ethnicity and referral source.

Percentile	Male	Female	Percentile	Under18	18to34	35to64	65Plus			
0	0 (0 - 0)	0 (0 - 0)	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)			
5	2 (2 - 3)	3 (3 - 3)	5	3 (2 - 3)	3 (3 - 3)	3 (2 - 3)	1 (0 - 1)			
25	8 (8 - 8)	8 (8 - 8)	25	8 (7 - 9)	9 (8 - 9)	8 (8 - 8)	5 (5 - 6)			
50	13 (13 - 13)	13 (13 - 13)	50	12 (12 - 13)	13 (13 - 13)	13 (13 - 13)	10 (10 - 10)			
75	17 (17 - 17)	17 (17 - 17)	75	16 (16 - 17)	17 (17 - 17)	17 (17 - 17)	15 (15 - 15)			
95	21 (21 - 21)	21 (21 - 21)	95	19 (19 - 20)	21 (21 - 21)	21 (21 - 21)	20 (19 - 20)			
100	21 (21 - 21)	21 (21 - 21)	100	21 (21 - 21)	21 (21 - 21)	21 (21 - 21)	21 (21 - 21)			
n	25913	50475	n	793	27916	39939	3127			
Kruskal-Wallis Chi Square = 78.0, df = 1, p=0.0000			Kruskal-Wallis Chi Square = 512.7, df = 3, p=0.0000							
Percentile	GP	Self	Other	Percentile	White British	Minority White	Mixed	Asian	Black	Other
0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)
5	3 (3 - 3)	2 (2 - 2)	2 (2 - 3)	5	3 (3 - 3)	3 (2 - 3)	3 (2 - 4)	3 (3 - 4)	3 (2 - 3)	3 (1.9 - 3)
25	8 (8 - 8)	8 (7 - 8)	8 (8 - 9)	25	8 (8 - 8)	8 (8 - 9)	9 (8 - 10)	9 (9 - 10)	8 (8 - 9)	9 (8 - 9)
50	13 (13 - 13)	13 (12 - 13)	13 (13 - 14)	50	13 (13 - 13)	13 (13 - 14)	14 (13 - 14.8)	15 (14 - 15)	13 (13 - 14)	14 (13 - 15)
75	17 (17 - 17)	17 (17 - 17)	17 (17 - 18)	75	17 (17 - 17)	17 (17 - 17)	17 (17 - 18)	18 (18 - 18)	17 (17 - 17)	18 (17 - 18)
95	21 (21 - 21)	21 (21 - 21)	21 (21 - 21)	95	21 (21 - 21)	21 (21 - 21)	21 (20 - 21)	21 (21 - 21)	21 (20 - 21)	21 (21 - 21)
100	21 (21 - 21)	21 (21 - 21)	21 (21 - 21)	100	21 (21 - 21)	21 (21 - 21)	21 (21 - 21)	21 (21 - 21)	21 (21 - 21)	21 (21 - 21)
n	65415	6777	4825	n	47495	3136	805	1243	1110	682
Kruskal-Wallis Chi Square = 28.2, df = 2, p=0.0000				Kruskal-Wallis Chi Square = 67.6, df = 5, p=0.0000						

Figures in parentheses are 95% confidence intervals

Table 25. Percentile points on distribution of Phobia scores between patient subgroups by gender, age, and broad ethnicity and referral source.

Percentile	Male	Female	Percentile	Under18	18to34	35to64	65Plus			
0	0 (0 - 0)	0 (0 - 0)	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)			
5	0 (0 - 0)	0 (0 - 0)	5	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)			
25	2 (2 - 2)	2 (2 - 2)	25	2.25 (2 - 3)	2 (2 - 3)	2 (2 - 2)	0 (0 - 0)			
50	6 (6 - 6)	6 (6 - 6)	50	6 (6 - 7)	6 (6 - 6)	6 (6 - 6)	4 (4 - 4)			
75	11 (11 - 11)	12 (12 - 12)	75	12 (11 - 12)	12 (11 - 12)	12 (12 - 12)	8 (8 - 9)			
95	19 (19 - 20)	20 (20 - 20)	95	19 (18 - 20)	19 (19 - 19.8)	20 (20 - 20)	16 (16 - 18)			
100	24 (24 - 24)	24 (24 - 24)	100	24 (24 - 24)	24 (24 - 24)	24 (24 - 24)	24 (24 - 24)			
n	24489	47778	n	748	26537	37625	2859			
Kruskal-Wallis Chi Square = 63.0, df = 1, p=0.0000			Kruskal-Wallis Chi Square = 353.4, df = 3, p=0.0000							
Percentile	GP	Self	Other	Percentile	White British	Minority White	Mixed	Asian	Black	Other
0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)
5	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	5	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)
25	2 (2 - 2)	2 (2 - 3)	3 (3 - 3)	25	2 (2 - 2)	2 (2 - 3)	3 (2 - 3)	3 (2 - 3)	2 (2 - 3)	2 (2 - 3)
50	6 (6 - 6)	6 (6 - 6)	8 (7 - 8)	50	6 (6 - 6)	6 (6 - 7)	6 (6 - 7)	8 (7 - 8)	8 (7 - 8)	7 (6 - 8)
75	12 (11 - 12)	11 (11 - 12)	13 (13 - 13)	75	12 (11 - 12)	12 (12 - 12)	12 (11 - 12)	13 (12 - 14)	14 (13 - 14)	12 (12 - 14)
95	20 (20 - 20)	20 (19 - 20)	21 (20 - 21)	95	20 (19 - 20)	20 (20 - 21)	19 (18 - 20.4)	22 (21 - 22)	22 (21 - 22)	21 (20 - 22)
100	24 (24 - 24)	24 (24 - 24)	24 (24 - 24)	100	24 (24 - 24)	24 (24 - 24)	24 (24 - 24)	24 (24 - 24)	24 (24 - 24)	24 (24 - 24)
n	62409	5860	4581	n	44946	3012	772	1166	1050	634
Kruskal-Wallis Chi Square = 129.4, df = 2, p=0.0000				Kruskal-Wallis Chi Square = 56.9, df = 5, p=0.0000						

Figures in parentheses are 95% confidence intervals

Table 26. Percentile points on distribution of Work and Social Adjustment Schedule scores between patient subgroups by gender, age, and broad ethnicity and referral source.

Percentile	Male	Female	Percentile	Under 18	18 to 34	35 to 64	65 Plus
0	0 (0 - 0)	0 (0 - 0)	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)
5	2 (1 - 2)	2 (1 - 2)	5	1 (0 - 2)	2 (2 - 3)	2 (2 - 2)	0 (0 - 0)
25	10 (10 - 10)	10 (10 - 10)	25	9 (8 - 10)	10 (10 - 10)	10 (10 - 10)	5 (4 - 5)
50	17 (17 - 17)	16 (16 - 16)	50	15 (14 - 16)	17 (17 - 17)	17 (17 - 17)	10 (10 - 10)
75	25 (24 - 25)	24 (24 - 24)	75	21 (20 - 21)	24 (24 - 24)	25 (25 - 25)	17 (17 - 18)
95	34 (34 - 34)	34 (34 - 34)	95	30 (28.7 - 31)	33 (33 - 33)	35 (35 - 35)	30 (28 - 30)
100	40 (40 - 40)	40 (40 - 40)	100	40 (40 - 40)	40 (40 - 40)	40 (40 - 40)	40 (40 - 40)
N	24608	48211	n	736	26639	37922	2879
Kruskal-Wallis Chi Square = 25.5, df = 1, p=0.0000			Kruskal-Wallis Chi Square = 1026.6, df = 3, p=0.0000				

Percentile	GP	Self	Other	Percentile	White British	Minority White	Mixed	Asian	Black	Other
0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)
5	2 (2 - 2)	1 (1 - 2)	0 (0 - 0)	5	2 (2 - 2)	2 (2 - 3)	2 (2 - 4)	2 (2 - 3)	1 (0 - 2)	2 (0.4 - 3)
25	10 (10 - 10)	10 (10 - 10)	10 (10 - 10)	25	10 (10 - 10)	11 (10 - 11)	11 (10 - 12)	12 (11 - 12)	11 (10 - 12)	10 (9 - 11)
50	16 (16 - 16)	17 (17 - 18)	18 (17 - 18)	50	16 (16 - 16)	18 (18 - 19)	19 (18 - 20)	20 (19 - 21)	20 (18 - 20)	18 (16 - 19)
75	24 (24 - 24)	25 (24 - 25)	26 (25 - 26)	75	24 (24 - 24)	26 (26 - 27)	26 (25 - 27)	28 (27 - 29)	28 (27 - 29)	26 (25 - 27)
95	34 (34 - 34)	35 (34 - 35)	35 (34 - 35)	95	33 (33 - 34)	35 (35 - 36)	35 (34 - 36)	38 (37 - 39)	36 (36 - 37)	36 (35 - 37)
100	40 (40 - 40)	40 (40 - 40)	40 (40 - 40)	100	40 (40 - 40)	40 (40 - 40)	40 (40 - 40)	40 (40 - 40)	40 (40 - 40)	40 (40 - 40)
n	62192	6664	4547	n	45134	3049	775	1178	1067	647
Kruskal-Wallis Chi Square = 41.8, df = 0, p=0.0000				Kruskal-Wallis Chi Square = 249.8, df = 5, p=0.0000						

Figures in parentheses are 95% confidence intervals

Table 27. Totals for counts of contacts by the three methods, average contacts per patient.

Contacts:	Counts	Average per patient
By therapist type	191451	2.4
By treatment type	281514	3.5
By purpose of session	255500	3.2
By purpose where treatment included	174576	2.2
Total patients	79310	

Note that per-patient averages in this table would be expected to be lower than rates shown later in the report which report activity only for patients with finished episodes.

Table 28. Total numbers of contact sessions reported, proportions by broad and detailed intensity level.

Intensity level	Proportion of sessions
All low intensity (AFC 1 to 5)	48.1%
All high intensity (AFC 6+)	51.9%
AFC grade 1	0.4%
AFC grade 2	0.1%
AFC grade 3	0.5%
AFC grade 4	18.0%
AFC grade 5	29.2%
AFC grade 6	36.4%
AFC grade 7	11.9%
AFC grade 8a	2.2%
AFC grade 8b	1.0%
AFC grade 8c	0.2%
AFC grade 8d	0.1%
Total reported therapist sessions	191,451

Chart 10. Staff grade profiles for reported clinical sessions by site

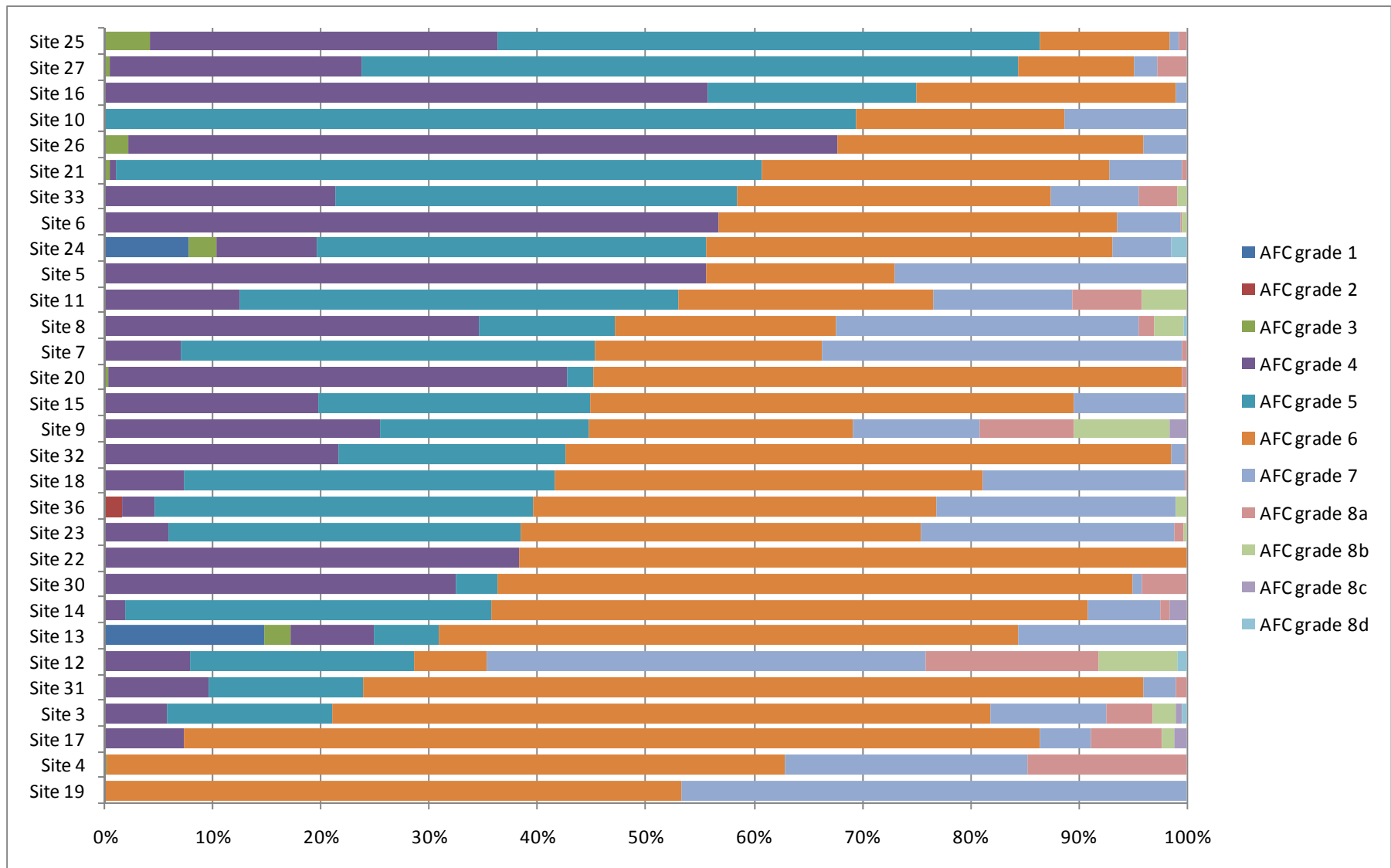


Table 29. Total session counts by types of therapy.

Type of intervention	Total reported sessions
Low intensity	
Computerised CBT	8355
Pure self help	28017
Guided self help	48233
Behavioural activation	9695
Structured exercise	4184
Psycho-educational group	18925
High intensity (% of high intensity)	
CBT	89,215 (66.4%)
Interpersonal therapy	1,340 (1.%)
Counselling	43,615 (32.4%)
Couple therapy	252 (.2%)
Other interventions	29683

Note, it seems likely that the low intensity treatment session counts include some multiple counting where more than one type of intervention was used in a session.

Table 30. Pattern of treatments for study group patients with finished episodes of care who received at least some treatment.

Pattern of treatment	Numbers / proportions
Low no high	
Number (% of total)	16,858 (42.3%)
1 contact	32%
2+ contacts	68%
High no low	
Number (% of total)	10,934 (27.5%)
1 contact	27%
2+ contacts	73%
Both low and high	
Number (% of total)	7,374 (18.5%)
1 contact	19%
2+ contacts	81%
Other treatment only	
Number (% of total)	3,725 (9.4%)
1 contact	72%
2+ contacts	28%
Treated - type not recorded	
Number (% of total)	928 (2.3%)
1 contact	72%
2+ contacts	28%
Total treated	39819

The table shows numbers and proportions receiving treatment in each of the patterns indicated. In addition it shows the proportions with only one or more than one reported contact – the requirement for outcome evaluation.

Chart 11. Site variation in treatment pattern for patients with finished episodes who received some treatment.

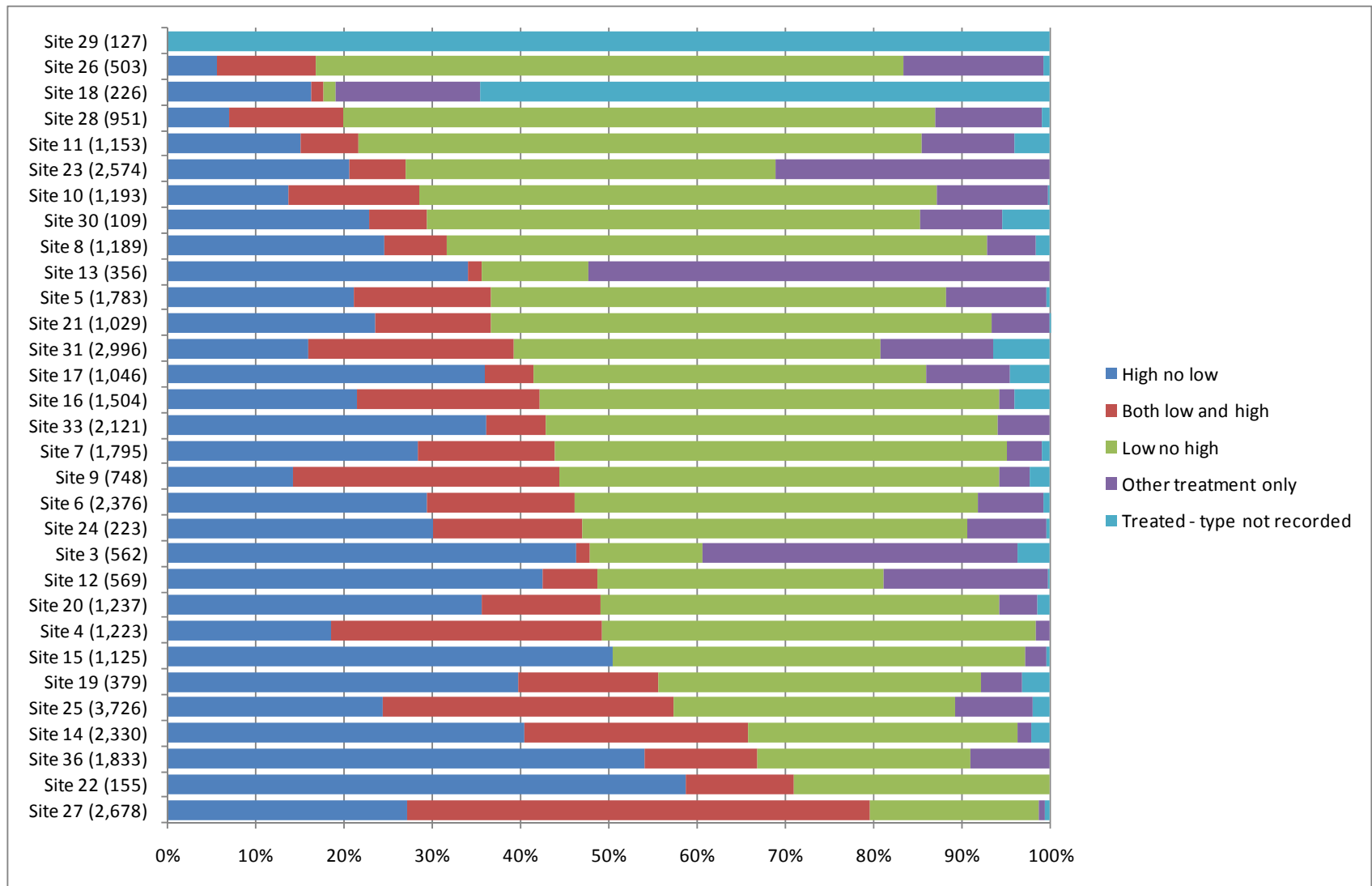


Table 31. Frequency of pair-wise combinations of low intensity treatment approaches

Combination	Total
Pure self help and Guided self help	3014
Guided self help and Behavioural activation	1397
Pure self help and Behavioural activation	1355
Pure self help and Psycho-educational group	1013
Computerised CBT and Pure self help	756
Pure self help and Structured exercise	745
Guided self help and Structured exercise	558
Guided self help and Psycho-educational group	456
Computerised CBT and Guided self help	423
Behavioural activation and Structured exercise	409
Computerised CBT and Psycho-educational group	282
Computerised CBT and Behavioural activation	209
Behavioural activation and Psycho-educational group	123
Structured exercise and Psycho-educational group	80
Computerised CBT and Structured exercise	73

Note that the 1574 individuals who received three or more approaches are counted more than once in this table.

Chart 12. Proportion of study group patients with finished episodes who received high intensity care who received CBT, Counselling or both, to show site variation.

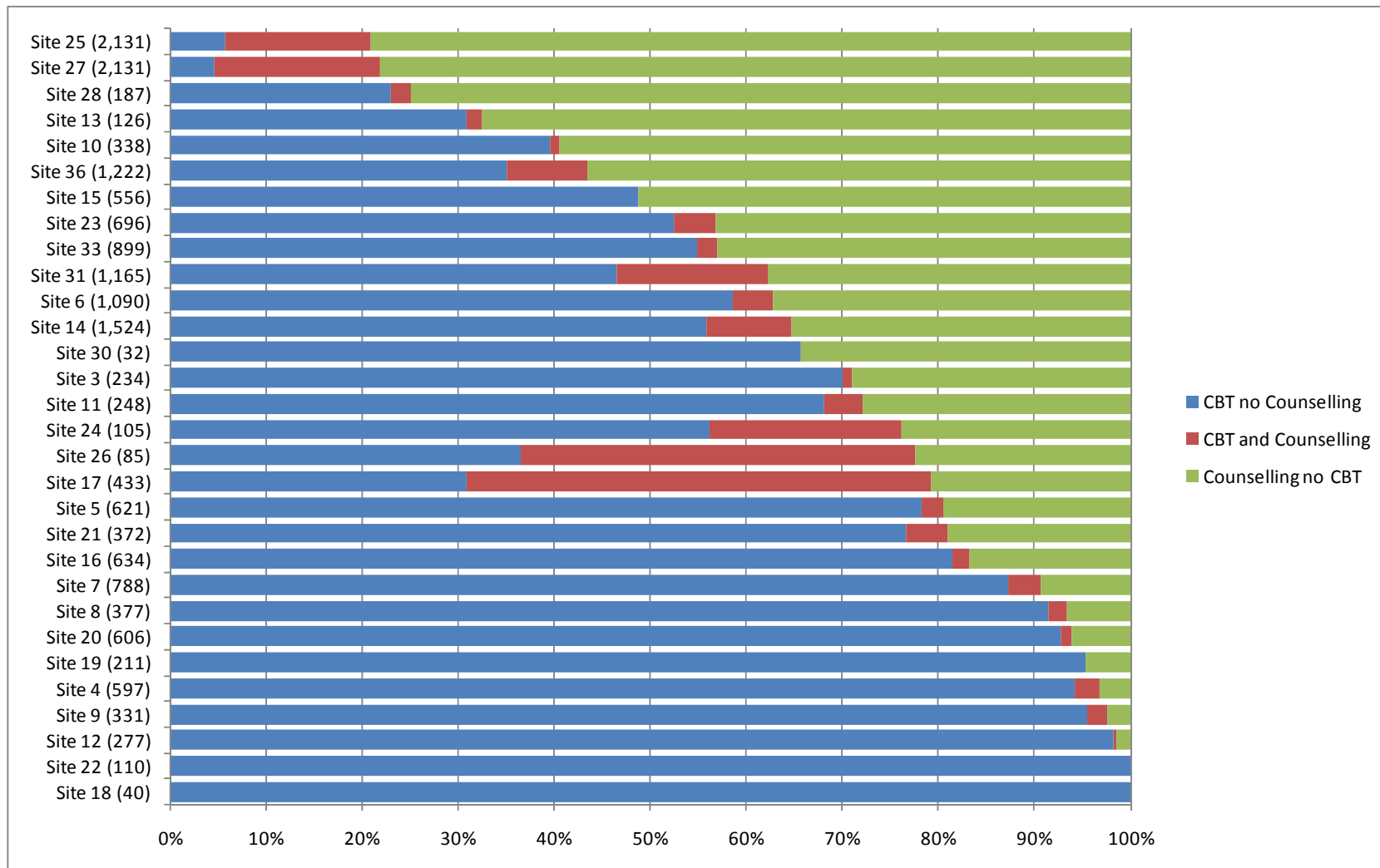


Chart 13. Proportions of study group patients with finished episodes by reason for ending of episode by site.

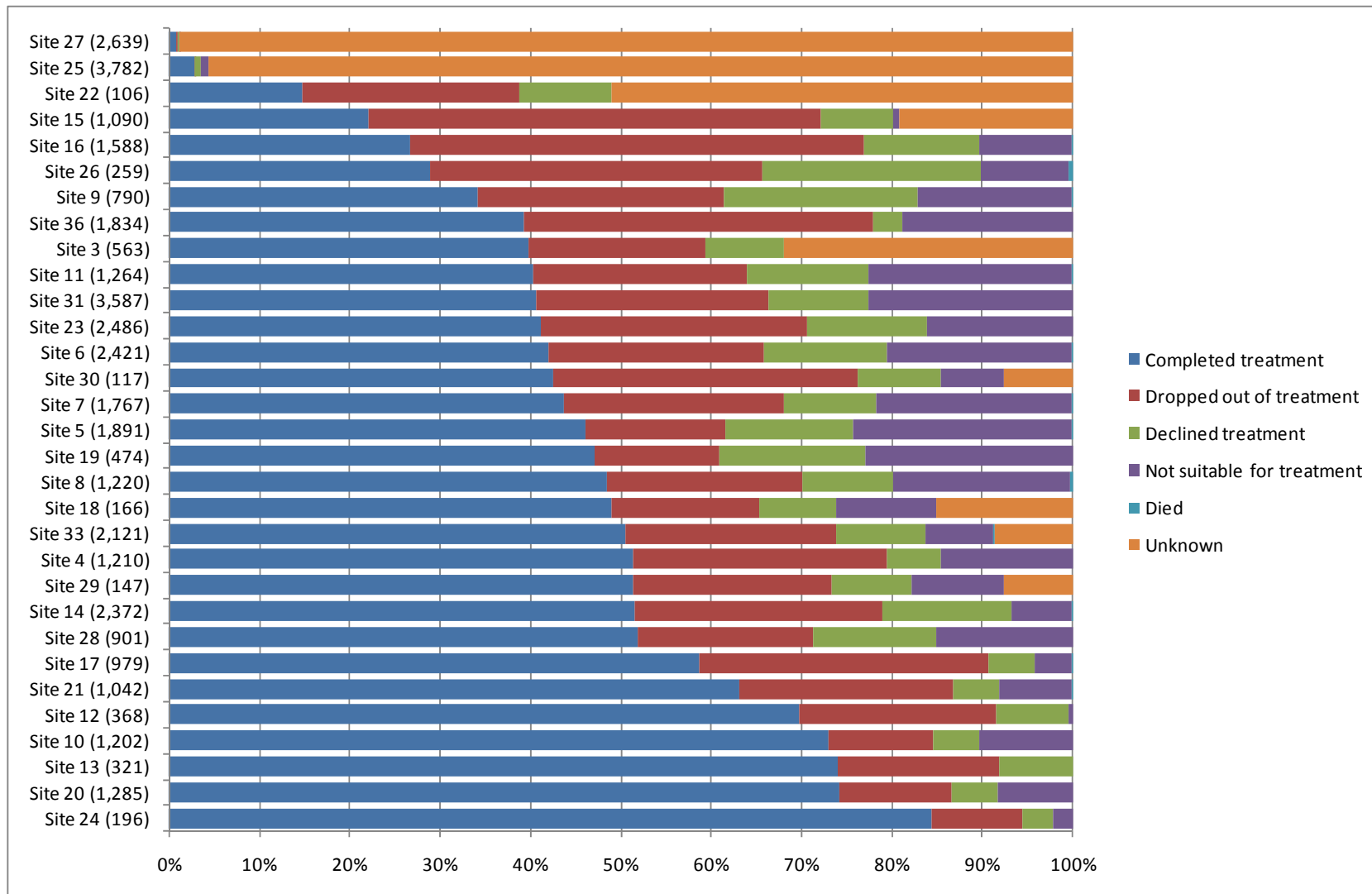


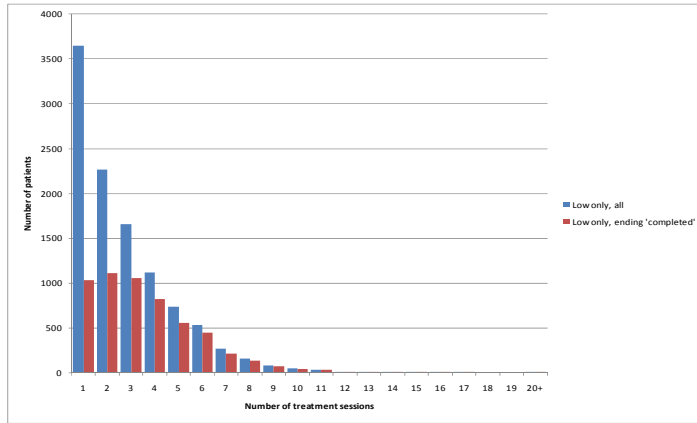
Table 31a Multivariate analysis of factors associated with different types of ending.

Predicting:	Unsuitable		Declined or dropped out	
Observations	29066		22178	
Predictor	Odds Ratio	p	Odds Ratio	p
PHQ-9 first score	1.26 (1.16 to 1.37)	**	1.08 (1.03 to 1.13)	**
GAD-7 first score	0.92 (0.85 to 0.99)	***		
Phobia questions total				
W&SAS first score	1.14 (1.07 to 1.21)	***	0.96 (0.91 to 1.00)	*
Start Month	0.96 (0.95 to 0.98)	***	0.83 (0.82 to 0.84)	
Whether using psychotropics				
Compared with Male				
Female	1.19 (1.07 to 1.32)	***		
Compared with Age 35 to 64				
Under 18	1.95 (1.16 to 3.28)	**		
Age 18to34			1.43 (1.33 to 1.54)	***
Age 65Plus			0.57 (0.45 to 0.73)	***
Compared with White British				
Minority White				
Mixed	1.38 (0.98 to 1.94)	*		
Asian				
Black	1.79 (1.37 to 2.35)	***		
Other				
Compared with Referred by GP				
Self referred				
Referred by other source				

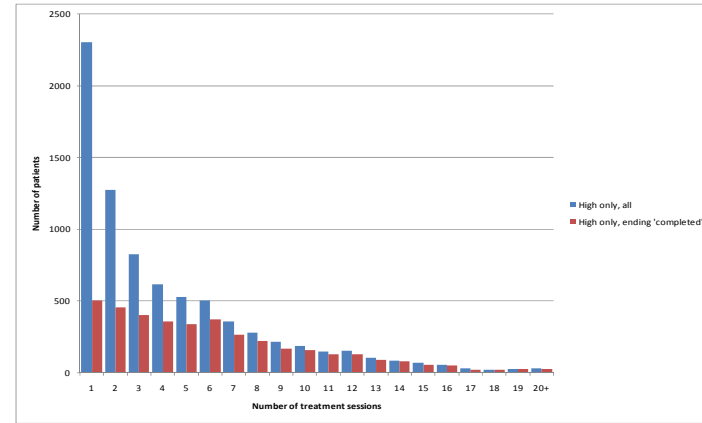
Table 31a (cont) Multivariate analysis of factors associated with different types of ending.

Predicting:	Unsuitable	Declined or dropped out
Primary Diagnosis comparison with Depressive Disorders		
Alcohol problems	6.77 (4.50 to 10.17) ***	
Bipolar disorder	3.03 (1.57 to 5.85) ***	
Recurrent depressive disorder		0.84 (0.74 to 0.97) **
Generalised anxiety disorder	0.76 (0.66 to 0.89) ***	
Mixed anxiety and depressing disorder		
Agoraphobia	0.48 (0.29 to 0.80) **	
Social phobia	0.32 (0.18 to 0.58) ***	0.63 (0.48 to 0.84) ***
Specific phobias	0.41 (0.22 to 0.78) **	
OCD	0.58 (0.39 to 0.86) **	0.51 (0.39 to 0.67) ***
PTSD		0.63 (0.47 to 0.84) ***
Somatoform disorder		0.50 (0.28 to 0.90) **
Eating disorder	2.81 (1.82 to 4.33) ***	
Family loss	1.92 (1.32 to 2.78) ***	0.46 (0.31 to 0.69) ***
Sites adding significantly to the model	14	24

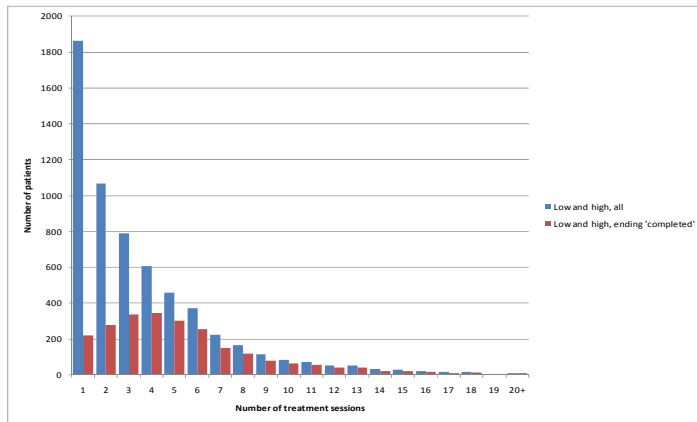
Charts 14 to 16. Numbers of treatment sessions for study group patients with completed episodes who received some treatment, by broad treatment pattern, overall and only those with endings coded as 'completed'.



Low intensity treatments only



High intensity treatments only



Both low and high intensity treatments

Table 32. Choice of treatment in relation to diagnosis, study group patients with finished episodes who received at least some treatment. Treatment columns are exclusive (each patient appears in only one); percentages are by row.

Diagnosis	CBT no counselling	Counselling no CBT	CBT and counselling	Other high intensity interventions	Low intensity interventions alone	Treatment type not recorded	Total
Mixed Anxiety Depressive Disorder	1,494 (20.8%)	1,738 (24.2%)	378 (5.3%)	17 (0.2%)	3,015 (42.0%)	545 (7.6%)	7,187
Depressive Episode	1,373 (19.1%)	1,554 (21.7%)	383 (5.3%)	27 (0.4%)	3,223 (44.9%)	612 (8.5%)	7,172
Generalised Anxiety Disorder	1,026 (22.7%)	871 (19.2%)	184 (4.1%)	9 (0.2%)	2,144 (47.4%)	291 (6.4%)	4,525
Recurrent Depressive Disorder	545 (32.5%)	216 (12.9%)	71 (4.2%)	9 (0.5%)	671 (40.0%)	164 (9.8%)	1,676
Family Loss	33 (7.2%)	250 (54.3%)	14 (3.0%)	3 (0.7%)	109 (23.7%)	51 (11.1%)	460
OCD	286 (64.6%)	16 (3.6%)	19 (4.3%)		90 (20.3%)	32 (7.2%)	443
PTSD	232 (57.3%)	60 (14.8%)	13 (3.2%)		67 (16.5%)	33 (8.1%)	405
Agoraphobia	187 (51.8%)	14 (3.9%)	3 (0.8%)		140 (38.8%)	17 (4.7%)	361
Social Phobia	171 (52.3%)	17 (5.2%)	7 (2.1%)	1 (0.3%)	111 (33.9%)	20 (6.1%)	327
Specific Phobia	154 (53.7%)	13 (4.5%)	10 (3.5%)		95 (33.1%)	15 (5.2%)	287
Eating Disorder	51 (34.2%)	23 (15.4%)	8 (5.4%)		47 (31.5%)	20 (13.4%)	149
Mental and behavioural disorders due to misuse of Alcohol	19 (14.2%)	22 (16.4%)	5 (3.7%)		52 (38.8%)	36 (26.9%)	134
Somatoform disorder	51 (47.7%)	5 (4.7%)	2 (1.9%)		40 (37.4%)	9 (8.4%)	107
Bipolar Disorder	16 (26.7%)	13 (21.7%)	2 (3.3%)	1 (1.7%)	19 (31.7%)	9 (15.0%)	60
Other Mental and behavioural Disorders	234 (28.6%)	102 (12.5%)	7 (0.9%)	6 (0.7%)	446 (54.5%)	24 (2.9%)	819
Other Diagnoses	10 (12.3%)	41 (50.6%)		3 (3.7%)	25 (30.9%)	2 (2.5%)	81
Mental disorder not otherwise specified	1,411 (25.2%)	876 (15.7%)	147 (2.6%)	55 (1.0%)	2,343 (41.9%)	760 (13.6%)	5,592
Missing or illegal data	1,693 (16.9%)	1,748 (17.4%)	348 (3.5%)	11 (0.1%)	4,221 (42.1%)	2,013 (20.1%)	10,034
Total	8,986 (22.6%)	7,579 (19.0%)	1,601 (4.0%)	142 (0.4%)	16,858 (42.3%)	4,653 (11.7%)	39,819

Table 33. Proportions of patients in each primary diagnosis category receiving specified types of treatment.

	Mixed anxiety and depressive disorder	Depressive episode	Generalized anxiety disorder	Recurrent depressive disorder	Obsessive-compulsive disorder	Post-traumatic stress disorder	Agora-phobia	Social phobias	Specific phobias
High intensity:									
CBT	26.0%	24.5%	26.7%	36.8%	68.8%	60.5%	52.6%	54.4%	57.1%
IPT	0.4%	0.6%	0.2%	0.7%	0.0%	0.0%	0.0%	0.3%	0.0%
Counselling	29.4%	27.0%	23.3%	17.1%	7.9%	18.0%	4.7%	7.3%	8.0%
Couple therapy	0.3%	0.2%	0.3%	0.2%	0.0%	0.2%	0.0%	0.0%	0.0%
No high intensity	49.5%	53.5%	53.8%	49.8%	27.5%	24.7%	43.5%	40.1%	38.3%
Low intensity									
Computerised CBT	6.4%	5.5%	6.9%	6.0%	0.9%	1.0%	3.3%	7.0%	4.2%
Pure self help	35.5%	34.8%	40.2%	34.0%	25.7%	22.5%	31.3%	30.6%	30.7%
Guided self help	25.1%	27.1%	28.6%	26.6%	16.7%	16.0%	32.4%	26.6%	27.9%
Behavioural activation	7.9%	9.9%	7.2%	11.4%	2.7%	3.5%	3.6%	4.3%	2.8%
Structured exercise	4.9%	2.6%	5.5%	3.8%	1.4%	1.7%	5.5%	4.6%	4.2%
Psycho-educational group	16.6%	13.5%	19.4%	9.5%	4.1%	3.7%	5.0%	8.3%	4.9%
No low intensity	32.1%	34.0%	26.4%	37.5%	58.9%	64.2%	44.0%	44.0%	42.9%
Other interventions	29.7%	28.8%	27.1%	33.9%	25.1%	25.4%	19.9%	21.7%	22.6%
Patients	7187	7172	4525	1676	443	405	361	327	287

Table 34. Receipt of high intensity care and the simple categorical variables. The table shows the proportions of study group members in each category who had high intensity interventions by simple categorical variables; study group members with finished episodes and receiving some treatment only.

Categories	Proportion receiving high intensity care
Gender	
Male	44.5%
Female	46.6%
n=39,053, chi square=15.0, df=1, p<.0001	
Broad Age group	
Under18	46.9%
18 to 34	44.9%
35 to 64	47.9%
65 Plus	43.9%
n=37,208, chi square=36.9, df=3, p<.0001	
Broad ethnic category	
White British	49.7%
Minority White	36.0%
Mixed	40.4%
Asian	33.5%
Black	32.2%
Other	38.1%
n=28,272, chi square=273.0, df=5, p<.0001	
Broad referral source	
GP	48.7%
Self	24.8%
Other	43.9%
n=39,368, chi square=777.8, df=2, p<.0001	

Total numbers vary between category groupings because of differing frequencies of missing or unusable category data. Total number with finished episodes receiving some treatment was 39,819.

Charts 17 to 20. Profile of initial test scores to compare patients who had, and did not have, high intensity care; study group patients with completed episodes who received some treatment only.

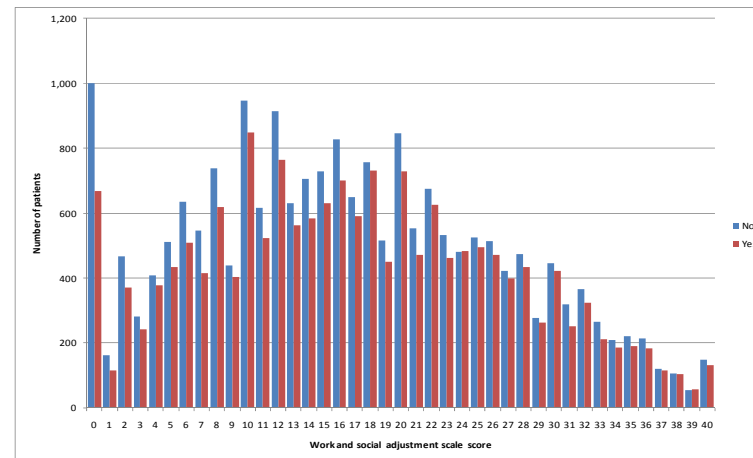
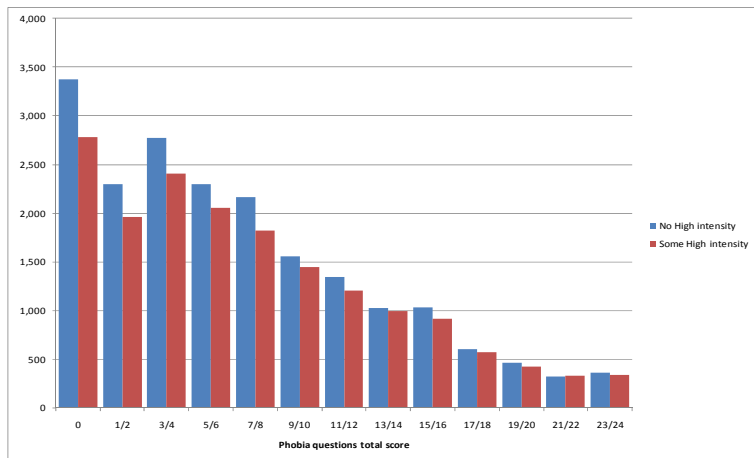
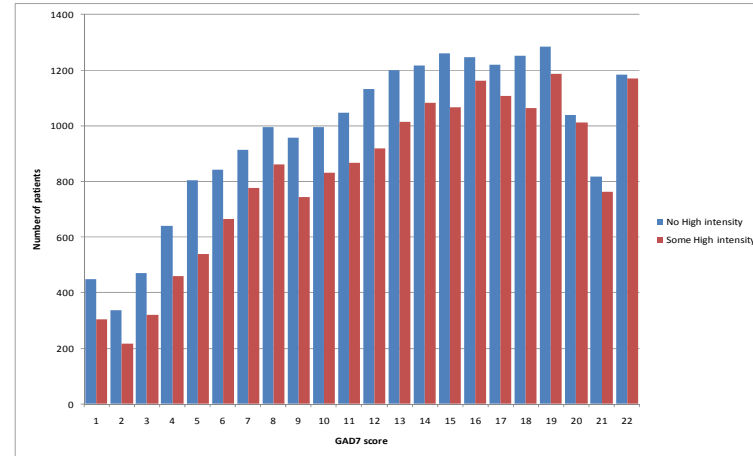
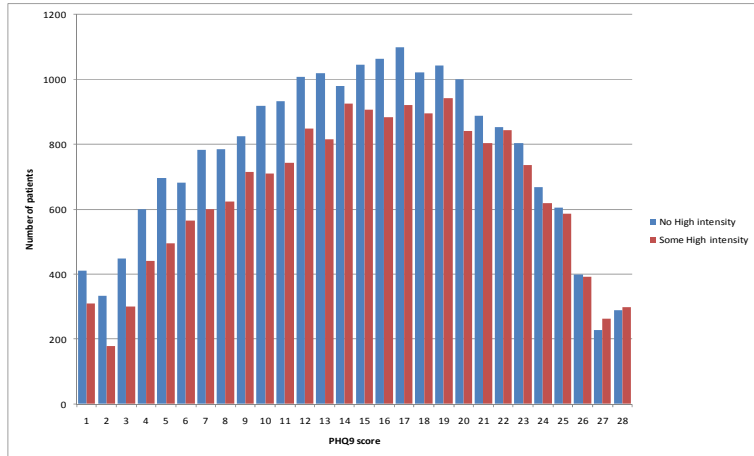


Table 35. Comparison of initial test scores in patients who received or did not receive high intensity treatment care; study group patients with completed episodes who received some treatment only.

Rating scale	Any high intensity treatment?	
	No	Yes
PHQ-9	Median (IQR)	14 (8 to 19) 14 (9 to 19)
	Mean	13.5 14.1
	n	21423 18205
	Kruskall Wallis chi square=75.3 df=1 p<0.0001	
GAD-7	Median (IQR)	12 (7 to 17) 13 (8 to 17)
	Mean	11.9 12.4
	n	21301 18132
	Kruskall Wallis chi square=73.6 df=1 p<0.0001	
Phobia questions	Median (IQR)	6 (2 to 11) 6 (2 to 11)
	Mean	7.1 7.3
	n	19617 17260
	Kruskall Wallis chi square=15.4 df=1 p<0.0001	
Work and social adjustment scale	Median (IQR)	16 (9 to 24) 17 (10 to 24)
	Mean	16.6 17.1
	n	20259 17548
	Kruskall Wallis chi square=21.4 df=1 p<0.0001	

Table 36. Multivariate analysis of factors predicting which patients receive any, and any high intensity, treatment, and type of high intensity treatment.

Predicting: Observations	Any Treatment Marker 20613		Any high intensity treatment 22157		Low intensity but no high intensity treatment 22157		Any CBT 22157		Any Counselling 22150	
Predictor	Odds Ratio	p	Odds Ratio	p	Odds Ratio	p	Odds Ratio	p	Odds Ratio	p
PHQ-9 first score			1.09 (1.04 to 1.14)	***	0.92 (0.88 to 0.96)	**			1.14 (1.07 to 1.20)	***
GAD-7 first score			1.07 (1.03 to 1.12)	***	0.94 (0.90 to 0.99)	**	1.09 (1.05 to 1.13)	***		
Phobia questions total					0.97 (0.94 to 1.00)	*	1.06 (1.03 to 1.10)	***	0.95 (0.91 to 1.00)	*
W&SAS first score							1.08 (1.04 to 1.12)	***	0.93 (0.88 to 0.98)	**
Start Month	0.87 (0.85 to 0.89)	***	0.89 (0.89 to 0.90)	**	1.08 (1.07 to 1.09)	***	0.89 (0.88 to 0.90)	***	0.95 (0.93 to 0.96)	***
Whether using psychotropics									0.82 (0.75 to 0.90)	***
Compared with Male										
Female	1.15 (0.99 to 1.33)	*	0.91 (0.86 to 0.97)	**	1.08 (1.02 to 1.15)	**			0.83 (0.76 to 0.91)	***
Compared with Age 35 to 64										
Under 18			0.65 (0.45 to 0.92)	**	1.65 (1.17 to 2.34)	**			0.40 (0.25 to 0.65)	***
Age 18to34			0.85 (0.80 to 0.90)	***	1.14 (1.08 to 1.21)	***	0.89 (0.84 to 0.95)	***	0.80 (0.73 to 0.87)	***
Age 65Plus	1.72 (1.10 to 2.67)	**	0.85 (0.73 to 0.99)	**	1.17 (1.01 to 1.36)	**	0.79 (0.67 to 0.93)	**		
Compared with White British										
Minority White	1.33 (1.01 to 1.74)	**					0.88 (0.78 to 1.00)	**		
Mixed										
Asian	0.72 (0.52 to 1.00)	**			1.17 (0.98 to 1.40)	*	0.81 (0.66 to 1.00)	**		
Black			0.81 (0.66 to 1.00)	*			0.73 (0.58 to 0.91)	**		
Other										
Compared with Referred by GP										
Self referred	2.22 (1.62 to 3.02)	***	0.84 (0.76 to 0.93)	***	1.16 (1.05 to 1.28)	**				
Referred by other source			1.18 (1.05 to 1.33)	**	0.80 (0.71 to 0.90)	***	1.38 (1.23 to 1.55)	***	0.66 (0.54 to 0.80)	***

Table 36 (cont). Multivariate analysis of factors predicting which patients receive any, and any high intensity, treatment, and type of high intensity treatment.

Predicting:	Any Treatment Marker		Any high intensity treatment		Low intensity but no high intensity treatment		Any CBT		Any Counselling	
Primary Diagnosis comparison with Depressive Episodes										
Alcohol problems			0.66 (0.44 to 1.01)	*			0.67 (0.42 to 1.08)	*		
Bipolar disorder					0.45 (0.25 to 0.83)	***	1.80 (1.05 to 3.08)	**		
Recurrent depressive disorder			1.53 (1.37 to 1.70)	***	0.70 (0.63 to 0.78)	***	1.75 (1.57 to 1.95)	***		
Generalised anxiety depressive disorder	1.58 (1.29 to 0.87)	***	0.87 (0.80 to 0.94)	***	1.28 (1.18 to 1.38)	***			0.71 (0.62 to 0.80)	***
Mixed anxiety depression	1.31 (1.12 to 0.83)	***					1.22 (1.14 to 1.31)	***	0.85 (0.77 to 0.94)	**
Agoraphobia			1.75 (1.41 to 2.16)	***	0.70 (0.56 to 0.87)	***	2.72 (2.20 to 3.37)	***	0.16 (0.09 to 0.28)	***
Social phobia	1.72 (1.02 to 2.05)	**	2.27 (1.84 to 2.80)	***	0.56 (0.45 to 0.69)	***	3.35 (2.72 to 4.12)	***	0.26 (0.17 to 0.42)	***
Specific phobias	2.56 (1.19 to 1.94)	**	2.71 (2.13 to 3.45)	***	0.45 (0.35 to 0.57)	***	4.22 (3.32 to 5.36)	***	0.36 (0.22 to 0.57)	***
OCD	1.47 (0.93 to 3.42)	*	5.08 (4.10 to 6.28)	***	0.24 (0.19 to 0.30)	***	8.51 (6.92 to 10.47)	***	0.22 (0.15 to 0.33)	***
PTSD			4.24 (3.36 to 5.35)	***	0.20 (0.16 to 0.27)	***	5.07 (4.09 to 6.27)	***	0.60 (0.43 to 0.83)	**
Somatoform disorder			2.21 (1.48 to 3.30)	***	0.53 (0.35 to 0.80)	**	3.51 (2.37 to 5.21)	***	0.18 (0.07 to 0.45)	***
Eating disorders	3.05 (0.95 to 0.50)	*	1.61 (1.13 to 2.30)	**	0.49 (0.34 to 0.71)	***	2.50 (1.78 to 3.50)	***	0.37 (0.21 to 0.66)	***
Family loss			1.78 (1.41 to 2.26)	***	0.42 (0.33 to 0.55)	***	0.34 (0.24 to 0.49)	***	3.91 (3.03 to 5.04)	***
Sites adding significantly to the model	17		16		18		17		19	

Seventeen of the thirty two sites add significant information to the model for any treatment marker. The number of covariate patterns is close to the number of observations so although there is a non-significant Pearson

Table 37. Three outcome measures for study group patients with more than one attendance and some treatment, all seven outcome markers.

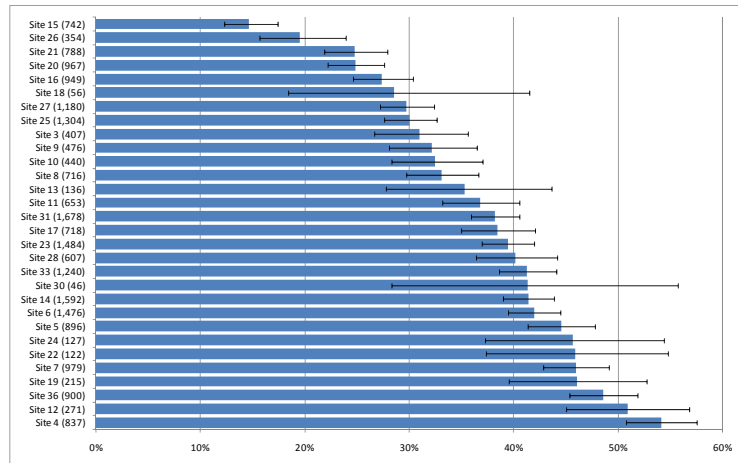
	Recovery	Incidence	Net change
Complete data only			
MTRDep	46.9% (46.1% to 47.7%) n=16,767	10.7% (10.0% to 11.5%) n=6,396	-31.0% (-31.9% to -30.1%) n=23,163
MTRAnx	45.8% (45.1% to 46.5%) n=17,843	14.6% (13.7% to 15.6%) n=5,194	-32.2% (-33.0% to -31.3%) n=23,037
MTRPhob	46.1% (45.2% to 47.1%) n=11,086	12.7% (12.1% to 13.4%) n=9,676	-18.7% (-19.6% to -17.8%) n=20,762
MTR1	42.2% (41.6% to 42.9%) n=19,467	13.8% (12.7% to 14.9%) n=3,650	-33.4% (-34.2% to -32.6%) n=23,117
MTR2	34.8% (34.2% to 35.5%) n=19,823	18.7% (17.2% to 20.3%) n=2,359	-29.1% (-29.9% to -28.4%) n=22,182
MTRben	23.1% (22.1% to 24.2%) n=6,259	6.9% (6.5% to 7.3%) n=14,908	-2.0% (-2.9% to -1.1%) n=21,167
MTRemp	9.7% (8.9% to 10.6%) n=4,588	4.8% (4.4% to 5.2%) n=12,314	+0.9% (-0.1% to +1.8%) n=16,902
Second Unknown No Change			
MTRDep	40.8% (40.1% to 41.5%) n=19,266	9.3% (8.6% to 10.0%) n=7,371	-27.0% (-27.8% to -26.2%) n=26,637
MTRAnx	39.8% (39.1% to 40.5%) n=20,543	12.5% (11.7% to 13.4%) n=6,062	-27.9% (-28.7% to -27.1%) n=26,605
MTRPhob	37.5% (36.6% to 38.3%) n=13,652	10.4% (9.8% to 10.9%) n=11,892	-15.2% (-16.0% to -14.3%) n=25,544
MTR1	36.8% (36.2% to 37.4%) n=22,356	11.7% (10.8% to 12.7%) n=4,292	-29.0% (-29.7% to -28.2%) n=26,648
MTR2	29.3% (28.7% to 29.8%) n=23,595	15.0% (13.8% to 16.4%) n=2,936	-24.4% (-25.0% to -23.7%) n=26,531
MTRben	18.5% (17.6% to 19.3%) n=7,842	5.7% (5.4% to 6.1%) n=17,891	-1.6% (-2.4% to -0.9%) n=25,733
MTRemp	7.6% (7.0% to 8.3%) n=5,828	4.1% (3.8% to 4.4%) n=14,502	+0.7% (-0.2% to +1.6%) n=20,330

The table presents the rates for recovery in cases (recovered cases/total initial cases), incidence in non cases (individuals not at case level at initial assessment but cases at final assessment) and the net change in prevalence. Patients with missing initial assessment data are omitted from both sets of analyses, those with missing final assessment data are omitted in the upper section of the table (complete data only) and in the lower section assumed to be unchanged from initial assessment (Second Unknown No Change). Figures in brackets are 95% confidence intervals.

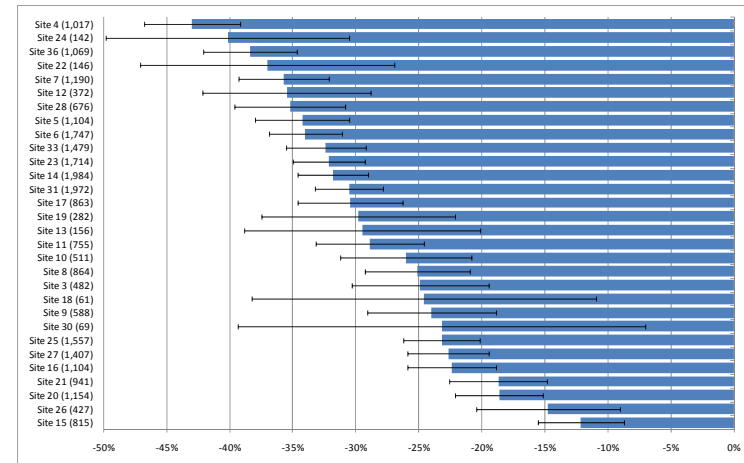
Table 37a Three measures of MTR 1 outcomes (recovery, incidence and net change in prevalence) for study group patients with more than one attendance and some treatment, patients grouped by reason care episode ended.

	Recovery	Incidence	Net change
Complete data only			
Completed	58.7% (57.8% to 59.7%) n=10,023	8.2% (7.2% to 9.4%) n=2,365	-46.0% (-47.0% to -44.9%) n=12,388
Declined or Dropped out	22.7% (21.6% to 23.8%) n=5,304	24.5% (21.4% to 27.8%) n=683	-17.3% (-18.7% to -15.9%) n=5,987
Not suitable	12.3% (10.4% to 14.4%) n=1,068	31.6% (23.8% to 40.6%) n=114	-8.0% (-10.8% to -5.3%) n=1,182
Other	32.6% (31.0% to 34.3%) n=3,072	21.7% (18.3% to 25.6%) n=488	-25.2% (-27.1% to -23.2%) n=3,560
Total	42.2% (41.6% to 42.9%) n=19,467	13.8% (12.7% to 14.9%) n=3,650	-33.4% (-34.2% to -32.6%) n=23,117
Second Unknown No Change			
Completed	56.4% (55.4% to 57.3%) n=10,439	7.6% (6.7% to 8.7%) n=2,541	-43.9% (-44.9% to -42.8%) n=12,980
Declined or Dropped out	18.0% (17.1% to 19.0%) n=6,674	17.9% (15.6% to 20.5%) n=933	-13.6% (-14.9% to -12.4%) n=7,607
Not suitable	8.6% (7.3% to 10.1%) n=1,518	21.1% (15.6% to 27.8%) n=171	-5.6% (-7.9% to -3.4%) n=1,689
Other	26.9% (25.5% to 28.3%) n=3,725	16.4% (13.7% to 19.4%) n=647	-20.5% (-22.3% to -18.7%) n=4,372
Total	36.8% (36.2% to 37.4%) n=22,356	11.7% (10.8% to 12.7%) n=4,292	-29.0% (-29.7% to -28.2%) n=26,648

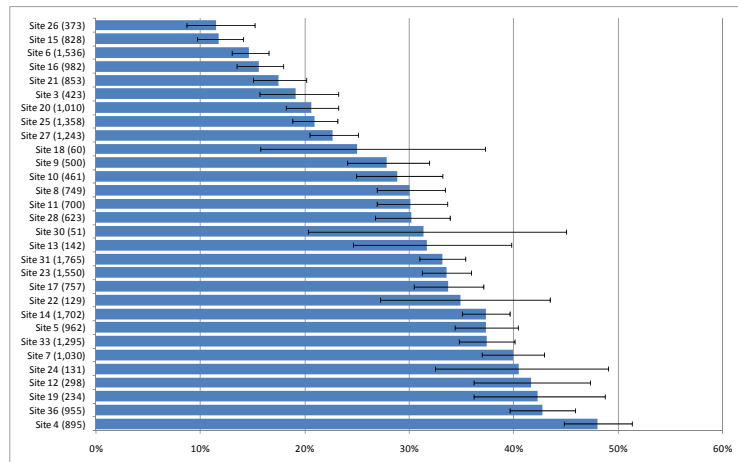
Charts 21 to 24. Site variation in outcomes: recovery rate for cases and net change in prevalence, MTR1 and MTR2 markers, 'second unknown no change' method.



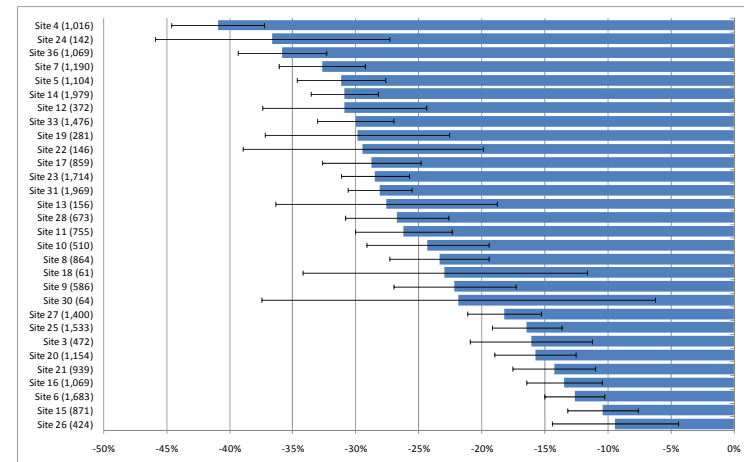
MTR1 – Recovery of cases



MTR1 – Net change in prevalence



MTR2 – Recovery of cases



MTR2 – Net change in prevalence

Table 38. Outcome and change in net prevalence, MTR1 marker, by primary diagnosis and high intensity treatment approach (part 1)

MTR1 Outcomes - Patients with missing second ratings omitted (complete data only method)

Primary diagnosis	CBT, no counselling		Counselling, no CBT	
	MTR1 recovery	MTR1 net prevalence change	MTR1 recovery	MTR1 net prevalence change
Depressive episode	41.3% (38.2% to 44.4%) n=957	-35.7% (-39.2% to -32.2%) n=1,061	41.1% (37.5% to 44.8%) n=689	-34.9% (-39.1% to -30.8%) n=784
Mxd anxiety depressive dis.	40.1% (37.2% to 43.1%) n=1,037	-33.2% (-36.6% to -29.9%) n=1,161	36.0% (32.5% to 39.6%) n=709	-30.2% (-34.2% to -26.1%) n=796
Generalized anxiety disorder	56.2% (52.4% to 59.8%) n=691	-42.6% (-46.7% to -38.4%) n=860	41.1% (35.7% to 46.7%) n=302	-27.0% (-33.4% to -20.5%) n=393
Recurrent depressive disorder	39.0% (34.3% to 43.9%) n=395	-33.5% (-38.9% to -28.1%) n=439	35.4% (26.6% to 45.2%) n=99	-29.5% (-40.3% to -18.6%) n=112
Obsessive-compulsive disorder	44.2% (37.6% to 51.0%) n=208	-35.9% (-43.6% to -28.1%) n=251		
Posttraumatic stress disorder	47.0% (39.1% to 55.0%) n=149	-38.7% (-47.8% to -29.7%) n=173	52.0% (33.5% to 70.0%) n=25	-46.2% (-66.7% to -25.6%) n=26
Agoraphobia	37.8% (30.2% to 45.9%) n=143	-31.7% (-40.9% to -22.5%) n=164		
Social phobias	48.7% (39.8% to 57.7%) n=115	-34.4% (-44.8% to -24.1%) n=154		
Family Loss	22.2% (9.0% to 45.2%) n=18	-18.2% (-43.9% to +7.6%) n=22	43.6% (34.0% to 53.7%) n=94	-33.3% (-44.9% to -21.8%) n=117
Specific phobias	53.8% (42.9% to 64.3%) n=80	-26.7% (-38.1% to -15.2%) n=135		
Other specified categories	42.9% (36.4% to 49.6%) n=210	-35.5% (-43.1% to -27.9%) n=242	40.3% (29.7% to 51.8%) n=72	-32.5% (-45.5% to -19.6%) n=83
Mental disorder NOS	44.3% (41.1% to 47.6%) n=893	-34.5% (-38.2% to -30.7%) n=1,076	49.0% (44.4% to 53.6%) n=447	-36.7% (-42.0% to -31.4%) n=550
Missing or illegal code	42.9% (39.9% to 46.0%) n=1,039	-33.7% (-37.1% to -30.2%) n=1,229	45.9% (42.5% to 49.3%) n=837	-35.3% (-39.1% to -31.4%) n=1,018
Total	43.9% (42.6% to 45.1%) n=5,935	-35.2% (-36.6% to -33.7%) n=6,967	42.2% (40.6% to 43.9%) n=3,301	-33.2% (-35.1% to -31.3%) n=3,915

Table shows recovery and prevalence change rates as percentages, 95% confidence interval in parentheses, and overall number in calculation. Data suppressed when calculations based on less than 10 cases. 'Complete data only' method refers to handling of missing data. It means only patients with valid ratings on all required initial and final scales are included (see text).

Table 38 (cont). Outcome and change in net prevalence, MTR1 marker, by primary diagnosis and high intensity treatment approach (part 2)

MTR1 Outcomes - Patients with missing second ratings omitted (complete data only method)

Primary diagnosis	CBT and counselling		Low intensity only	
	MTR1 recovery	MTR1 net prevalence change	MTR1 recovery	MTR1 net prevalence change
Depressive episode	45.6% (39.1% to 52.3%) n=215	-38.2% (-45.6% to -30.7%) n=241	38.0% (35.6% to 40.4%) n=1,595	-30.8% (-33.6% to -28.1%) n=1,845
Mxd anxiety depressive dis.	33.9% (28.1% to 40.2%) n=233	-29.8% (-36.6% to -23.0%) n=252	39.9% (37.5% to 42.3%) n=1,622	-33.3% (-36.0% to -30.6%) n=1,842
Generalized anxiety disorder	44.4% (35.4% to 53.8%) n=108	-36.3% (-46.9% to -25.7%) n=124	52.1% (49.2% to 54.9%) n=1,156	-38.5% (-41.8% to -35.2%) n=1,473
Recurrent depressive disorder	23.9% (13.9% to 37.9%) n=46	-18.0% (-32.3% to -3.7%) n=50	33.1% (28.4% to 38.2%) n=353	-27.1% (-33.0% to -21.3%) n=409
Obsessive-compulsive disorder	25.0% (8.9% to 53.2%) n=12	-21.4% (-52.5% to +9.7%) n=14	44.7% (30.1% to 60.3%) n=38	-36.2% (-54.3% to -18.0%) n=47
Posttraumatic stress disorder			43.3% (27.4% to 60.8%) n=30	-33.3% (-53.7% to -13.0%) n=36
Agoraphobia			51.4% (40.2% to 62.4%) n=74	-38.1% (-51.0% to -25.3%) n=97
Social phobias			52.5% (40.0% to 64.7%) n=59	-37.7% (-52.1% to -23.2%) n=77
Family Loss	27.3% (9.7% to 56.6%) n=11	-27.3% (-53.6% to -1.0%) n=11	26.9% (13.7% to 46.1%) n=26	-16.7% (-38.5% to +5.2%) n=36
Specific phobias			52.5% (37.5% to 67.1%) n=40	-29.2% (-45.6% to -12.8%) n=65
Other specified categories	23.1% (8.2% to 50.3%) n=13	-17.6% (-48.5% to +13.2%) n=17	37.8% (32.0% to 43.9%) n=249	-30.5% (-37.5% to -23.5%) n=295
Mental disorder NOS	46.8% (37.7% to 56.1%) n=109	-36.4% (-47.1% to -25.8%) n=129	45.5% (42.6% to 48.5%) n=1,101	-33.7% (-37.1% to -30.4%) n=1,375
Missing or illegal code	42.1% (36.1% to 48.2%) n=252	-34.1% (-41.1% to -27.1%) n=296	43.8% (41.6% to 46.1%) n=1,877	-33.8% (-36.4% to -31.3%) n=2,293
Total	40.2% (37.2% to 43.2%) n=1,018	-33.1% (-36.6% to -29.7%) n=1,162	42.8% (41.7% to 43.9%) n=8,220	-33.5% (-34.7% to -32.2%) n=9,890

Table shows recovery and prevalence change rates as percentages, 95% confidence interval in parentheses, and overall number in calculation. Data suppressed when calculations based on less than 10 cases. 'Complete data only' method refers to handling of missing data. It means only patients with valid ratings on all required initial and final scales are included (see text).

Table 38 (cont). Outcome and change in net prevalence, MTR1 marker, by primary diagnosis and high intensity treatment approach (part 3)

MTR1 Outcomes - Patients with missing second ratings handled by 'second unknown, no change method'.

Primary diagnosis	CBT, no counselling		Counselling, no CBT	
	MTR1 recovery	MTR1 net prevalence change	MTR1 recovery	MTR1 net prevalence change
Depressive episode	37.2% (34.3% to 40.1%) n=1,062	-32.0% (-35.3% to -28.7%) n=1,184	35.5% (32.3% to 38.9%) n=797	-30.1% (-33.9% to -26.2%) n=911
Mxd anxiety depressive dis.	36.5% (33.8% to 39.4%) n=1,139	-30.3% (-33.5% to -27.1%) n=1,273	32.5% (29.3% to 35.8%) n=785	-27.0% (-30.8% to -23.2%) n=889
Generalized anxiety disorder	53.4% (49.8% to 57.0%) n=726	-40.4% (-44.5% to -36.3%) n=906	36.6% (31.6% to 41.8%) n=339	-24.1% (-30.2% to -18.0%) n=439
Recurrent depressive disorder	36.4% (32.0% to 41.1%) n=423	-31.3% (-36.5% to -26.1%) n=469	29.9% (22.4% to 38.7%) n=117	-24.8% (-34.7% to -14.9%) n=133
Obsessive-compulsive disorder	41.8% (35.5% to 48.4%) n=220	-34.0% (-41.5% to -26.4%) n=265		
Posttraumatic stress disorder	40.9% (33.8% to 48.4%) n=171	-33.8% (-42.3% to -25.4%) n=198	32.5% (20.1% to 48.0%) n=40	-29.3% (-44.3% to -14.3%) n=41
Agoraphobia	35.8% (28.6% to 43.7%) n=151	-30.2% (-39.1% to -21.4%) n=172		
Social phobias	45.5% (37.0% to 54.3%) n=123	-32.3% (-42.4% to -22.3%) n=164	50.0% (23.7% to 76.3%) n=10	-41.7% (-76.6% to -6.7%) n=12
Family Loss	18.2% (7.3% to 38.5%) n=22	-15.4% (-37.9% to +7.1%) n=26	33.1% (25.4% to 41.7%) n=124	-24.8% (-34.9% to -14.8%) n=157
Specific phobias	52.4% (41.8% to 62.9%) n=82	-25.2% (-36.3% to -14.0%) n=143		
Other specified categories	32.1% (26.9% to 37.8%) n=280	-26.7% (-33.2% to -20.2%) n=322	22.5% (16.1% to 30.4%) n=129	-19.0% (-27.8% to -10.2%) n=142
Mental disorder NOS	40.2% (37.1% to 43.3%) n=986	-31.0% (-34.6% to -27.5%) n=1,196	41.0% (36.9% to 45.2%) n=534	-30.3% (-35.2% to -25.5%) n=666
Missing or illegal code	39.8% (37.0% to 42.7%) n=1,121	-31.2% (-34.5% to -27.9%) n=1,327	43.1% (39.9% to 46.4%) n=891	-33.2% (-37.0% to -29.5%) n=1,081
Total	40.0% (38.8% to 41.2%) n=6,506	-32.1% (-33.4% to -30.7%) n=7,645	36.8% (35.3% to 38.4%) n=3,785	-28.9% (-30.7% to -27.1%) n=4,497

Table shows recovery and prevalence change rates as percentages, 95% confidence interval in parentheses, and overall number in calculation. Data suppressed when calculations based on less than 10 cases. 'Second unknown, no change method' method refers to handling of missing data. It means all patients with valid initial ratings on required scales are included. Where final ratings are missing, patients are assumed to be unchanged from initial rating (see text).

Table 38 (cont). Outcome and change in net prevalence, MTR1 marker, by primary diagnosis and high intensity treatment approach (part 4)

MTR1 Outcomes - Patients with missing second ratings handled by 'second unknown, no change method'.

Primary diagnosis	CBT and counselling		Low intensity only	
	MTR1 recovery	MTR1 net prevalence change	MTR1 recovery	MTR1 net prevalence change
Depressive episode	29.9% (25.2% to 35.0%) n=328	-24.1% (-30.1% to -18.1%) n=382	32.2% (30.1% to 34.3%) n=1,883	-26.2% (-28.7% to -23.7%) n=2,175
Mxd anxiety depressive dis.	22.8% (18.7% to 27.5%) n=347	-19.9% (-25.2% to -14.7%) n=376	35.7% (33.6% to 38.0%) n=1,810	-29.7% (-32.3% to -27.2%) n=2,062
Generalized anxiety disorder	31.0% (24.2% to 38.6%) n=155	-24.5% (-33.3% to -15.6%) n=184	48.0% (45.3% to 50.8%) n=1,254	-34.7% (-37.9% to -31.5%) n=1,634
Recurrent depressive disorder	17.7% (10.2% to 29.0%) n=62	-12.7% (-25.4% to +0.1%) n=71	28.6% (24.4% to 33.2%) n=409	-23.4% (-28.7% to -18.0%) n=475
Obsessive-compulsive disorder	20.0% (7.0% to 45.2%) n=15	-17.6% (-44.2% to +8.9%) n=17	41.5% (27.8% to 56.6%) n=41	-31.5% (-49.0% to -14.0%) n=54
Posttraumatic stress disorder	30.0% (10.8% to 60.3%) n=10	-23.1% (-58.6% to +12.4%) n=13	40.6% (25.5% to 57.7%) n=32	-30.8% (-50.5% to -11.0%) n=39
Agoraphobia			46.3% (36.0% to 57.1%) n=82	-34.3% (-46.6% to -22.0%) n=108
Social phobias			48.4% (36.6% to 60.4%) n=64	-34.5% (-48.5% to -20.6%) n=84
Family Loss	23.1% (8.2% to 50.3%) n=13	-21.4% (-48.7% to +5.8%) n=14	20.6% (10.3% to 36.8%) n=34	-13.3% (-32.3% to +5.6%) n=45
Specific phobias		-20.0% (-62.0% to +22.0%) n=10	45.7% (32.2% to 59.8%) n=46	-25.7% (-41.2% to -10.1%) n=74
Other specified categories	15.0% (5.2% to 36.0%) n=20	-12.5% (-36.0% to +11.0%) n=24	26.5% (22.2% to 31.3%) n=355	-21.4% (-27.2% to -15.7%) n=420
Mental disorder NOS	41.1% (32.9% to 49.9%) n=124	-32.0% (-42.0% to -22.0%) n=147	40.9% (38.1% to 43.6%) n=1,226	-30.2% (-33.4% to -27.0%) n=1,537
Missing or illegal code	36.4% (31.1% to 42.1%) n=291	-29.3% (-35.8% to -22.8%) n=345	38.5% (36.5% to 40.6%) n=2,136	-29.4% (-31.8% to -26.9%) n=2,643
Total	29.7% (27.3% to 32.1%) n=1,378	-24.2% (-27.1% to -21.3%) n=1,593	37.5% (36.6% to 38.5%) n=9,372	-29.2% (-30.3% to -28.0%) n=11,350

Table shows recovery and prevalence change rates as percentages, 95% confidence interval in parentheses, and overall number in calculation. Data suppressed when calculations based on less than 10 cases. 'Second unknown, no change method' method refers to handling of missing data. It means all patients with valid initial ratings on required scales are included. Where final ratings are missing, patients are assumed to be unchanged from initial rating (see text).

Table 38 (cont). Outcome and change in net prevalence, MTR2 marker, by primary diagnosis and high intensity treatment approach (part 5)

MTR2 Outcomes - Second unknown, no change method

Primary diagnosis	CBT, no counselling		Counselling, no CBT	
	MTR2 recovery	MTR2 net prevalence change	MTR2 recovery	MTR2 net prevalence change
Depressive episode	29.9% (27.3% to 32.7%) n=1,089	-26.2% (-29.3% to -23.1%) n=1,187	26.9% (24.0% to 30.1%) n=824	-23.8% (-27.3% to -20.2%) n=905
Mixed anxiety and depressive disorder	29.2% (26.6% to 31.8%) n=1,173	-25.0% (-28.0% to -22.1%) n=1,271	24.2% (21.4% to 27.3%) n=801	-20.0% (-23.6% to -16.4%) n=886
Generalized anxiety disorder	43.3% (39.9% to 46.8%) n=792	-36.0% (-39.9% to -32.1%) n=905	27.0% (22.7% to 31.8%) n=359	-18.7% (-24.4% to -12.9%) n=434
Recurrent depressive disorder	30.3% (26.2% to 34.7%) n=436	-26.9% (-31.8% to -22.1%) n=468	20.3% (14.2% to 28.3%) n=123	-17.4% (-25.9% to -8.9%) n=132
Obsessive-compulsive disorder	36.2% (30.3% to 42.5%) n=235	-30.6% (-37.6% to -23.5%) n=265		
Posttraumatic stress disorder	33.3% (26.9% to 40.4%) n=183	-29.9% (-37.6% to -22.3%) n=197	19.0% (10.0% to 33.3%) n=42	-16.3% (-28.8% to -3.8%) n=43
Agoraphobia	26.5% (20.4% to 33.7%) n=166	-25.7% (-33.0% to -18.5%) n=171		
Social phobias	34.0% (26.9% to 41.9%) n=150	-30.5% (-39.1% to -21.9%) n=164	45.5% (21.3% to 72.0%) n=11	-33.3% (-65.3% to -1.4%) n=12
Disappearance / death of family member	18.2% (7.3% to 38.5%) n=22	-15.4% (-37.9% to +7.1%) n=26	28.6% (21.6% to 36.8%) n=133	-23.4% (-32.9% to -13.9%) n=158
Specific phobias	34.9% (27.2% to 43.6%) n=126	-27.3% (-36.9% to -17.7%) n=143		
Other specified categories	27.0% (22.4% to 32.2%) n=311	-24.2% (-29.9% to -18.5%) n=335	19.3% (13.6% to 26.6%) n=140	-18.4% (-26.0% to -10.7%) n=147
Mental disorder NOS	33.8% (31.0% to 36.7%) n=1,047	-27.9% (-31.2% to -24.6%) n=1,190	36.4% (32.5% to 40.5%) n=558	-27.8% (-32.5% to -23.1%) n=665
Missing or illegal code	32.9% (30.2% to 35.6%) n=1,187	-27.2% (-30.3% to -24.1%) n=1,327	32.0% (29.1% to 35.0%) n=932	-25.8% (-29.4% to -22.3%) n=1,065
Total	32.7% (31.6% to 33.8%) n=6,917	-27.9% (-29.1% to -26.6%) n=7,649	28.4% (27.0% to 29.8%) n=3,946	-23.1% (-24.8% to -21.4%) n=4,473

Table shows recovery and prevalence change rates as percentages, 95% confidence interval in parentheses, and overall number in calculation. Data suppressed when calculations based on less than 10 cases. 'Second unknown, no change method' method refers to handling of missing data. It means all patients with valid initial ratings on required scales are included. Where final ratings are missing, patients are assumed to be unchanged from initial rating (see text).

Table 38 (cont). Outcome and change in net prevalence, MTR2 marker, by primary diagnosis and high intensity treatment approach (part 6)

MTR2 Outcomes - Second unknown, no change method

Primary diagnosis	CBT and counselling		Low intensity only	
	MTR2 recovery	MTR2 net prevalence change	MTR2 recovery	MTR2 net prevalence change
Depressive episode	25.0% (20.7% to 29.9%) n=340	-20.9% (-26.5% to -15.3%) n=378	24.2% (22.4% to 26.2%) n=1,948	-20.5% (-22.8% to -18.1%) n=2,166
Mixed anxiety and depressive disorder	16.4% (12.9% to 20.6%) n=354	-14.8% (-19.4% to -10.2%) n=372	26.5% (24.6% to 28.5%) n=1,895	-23.3% (-25.6% to -21.0%) n=2,050
Generalized anxiety disorder	23.5% (17.7% to 30.5%) n=166	-20.3% (-28.1% to -12.6%) n=182	38.3% (35.8% to 41.0%) n=1,356	-30.1% (-33.1% to -27.1%) n=1,617
Recurrent depressive disorder	15.6% (8.7% to 26.4%) n=64	-10.0% (-21.2% to +1.2%) n=70	24.6% (20.8% to 29.0%) n=422	-20.6% (-25.6% to -15.5%) n=472
Obsessive-compulsive disorder	20.0% (7.0% to 45.2%) n=15	-17.6% (-44.2% to +8.9%) n=17	33.3% (21.0% to 48.4%) n=42	-24.1% (-41.4% to -6.8%) n=54
Posttraumatic stress disorder	30.0% (10.8% to 60.3%) n=10	-23.1% (-58.6% to +12.4%) n=13	32.4% (19.1% to 49.2%) n=34	-23.1% (-41.4% to -4.7%) n=39
Agoraphobia			33.0% (24.3% to 43.0%) n=94	-27.8% (-39.0% to -16.6%) n=108
Social phobias			40.3% (30.0% to 51.4%) n=77	-35.7% (-47.9% to -23.6%) n=84
Disappearance / death of family member	0.0% (0.0% to 22.8%) n=13	0.0% (-19.1% to +19.1%) n=14	11.1% (4.4% to 25.3%) n=36	-6.8% (-24.0% to +10.3%) n=44
Specific phobias		-20.0% (-53.9% to +13.9%) n=10	38.5% (27.6% to 50.6%) n=65	-33.8% (-47.4% to -20.2%) n=74
Other specified categories	14.3% (5.0% to 34.6%) n=21	-8.3% (-29.3% to +12.6%) n=24	19.7% (16.1% to 23.9%) n=396	-17.1% (-22.0% to -12.2%) n=433
Mental disorder NOS	37.0% (29.1% to 45.7%) n=127	-29.3% (-39.0% to -19.5%) n=147	34.1% (31.5% to 36.7%) n=1,301	-27.2% (-30.2% to -24.2%) n=1,522
Missing or illegal code	32.9% (27.9% to 38.3%) n=310	-28.1% (-34.2% to -22.1%) n=345	31.6% (29.8% to 33.6%) n=2,266	-25.5% (-27.8% to -23.2%) n=2,616
Total	24.5% (22.4% to 26.8%) n=1,439	-20.8% (-23.5% to -18.1%) n=1,582	29.7% (28.8% to 30.6%) n=9,932	-24.6% (-25.6% to -23.5%) n=11,279

Table shows recovery and prevalence change rates as percentages, 95% confidence interval in parentheses, and overall number in calculation. Data suppressed when calculations based on less than 10 cases. 'Second unknown, no change method' method refers to handling of missing data. It means all patients with valid initial ratings on required scales are included. Where final ratings are missing, patients are assumed to be unchanged from initial rating (see text).

Table 38. Outcome and change in net prevalence, MTR2 marker, by primary diagnosis and high intensity treatment approach (part 7)

MTR2 Outcomes - Complete data only method

Primary diagnosis	CBT, no counselling		Counselling, no CBT	
	MTR2 recovery	MTR2 net prevalence change	MTR2 recovery	MTR2 net prevalence change
Depressive episode	34.6% (31.6% to 37.7%) n=943	-30.3% (-33.8% to -26.9%) n=1,025	32.9% (29.5% to 36.5%) n=675	-29.3% (-33.3% to -25.3%) n=733
Mixed anxiety and depressive disorder	33.3% (30.5% to 36.2%) n=1,027	-28.6% (-31.8% to -25.3%) n=1,113	27.6% (24.5% to 31.1%) n=702	-23.2% (-27.0% to -19.4%) n=764
Generalized anxiety disorder	47.3% (43.7% to 50.9%) n=725	-39.5% (-43.6% to -35.4%) n=825	31.1% (26.2% to 36.4%) n=312	-21.5% (-27.8% to -15.3%) n=376
Recurrent depressive disorder	33.2% (28.7% to 37.9%) n=398	-29.4% (-34.6% to -24.3%) n=428	26.0% (18.3% to 35.6%) n=96	-21.9% (-32.2% to -11.6%) n=105
Obsessive-compulsive disorder	39.2% (32.9% to 45.8%) n=217	-33.3% (-40.7% to -26.0%) n=243		
Posttraumatic stress disorder	39.6% (32.2% to 47.5%) n=154	-35.8% (-44.2% to -27.3%) n=165	30.8% (16.5% to 50.0%) n=26	-25.9% (-44.6% to -7.3%) n=27
Agoraphobia	28.4% (21.9% to 35.9%) n=155	-27.5% (-35.1% to -19.9%) n=160		
Social phobias	37.0% (29.4% to 45.3%) n=138	-33.1% (-42.2% to -24.1%) n=151	50.0% (23.7% to 76.3%) n=10	-36.4% (-70.3% to -2.4%) n=11
Disappearance / death of family member	22.2% (9.0% to 45.2%) n=18	-18.2% (-43.9% to +7.6%) n=22	38.8% (29.7% to 48.7%) n=98	-32.2% (-43.4% to -21.0%) n=115
Specific phobias	37.0% (28.8% to 45.9%) n=119	-29.3% (-39.1% to -19.5%) n=133		
Other specified categories	37.0% (31.0% to 43.5%) n=227	-33.2% (-40.1% to -26.3%) n=244	34.2% (24.7% to 45.2%) n=79	-32.1% (-43.7% to -20.6%) n=84
Mental disorder NOS	38.0% (35.0% to 41.2%) n=931	-31.6% (-35.1% to -28.0%) n=1,051	44.5% (40.0% to 49.1%) n=456	-34.5% (-39.7% to -29.3%) n=537
Missing or illegal code	36.1% (33.3% to 39.1%) n=1,079	-29.9% (-33.2% to -26.6%) n=1,208	36.3% (33.0% to 39.6%) n=822	-29.3% (-33.1% to -25.5%) n=939
Total	36.9% (35.7% to 38.1%) n=6,131	-31.5% (-32.9% to -30.1%) n=6,768	33.9% (32.3% to 35.6%) n=3,298	-27.8% (-29.6% to -25.9%) n=3,716

Table shows recovery and prevalence change rates as percentages, 95% confidence interval in parentheses, and overall number in calculation. Data suppressed when calculations based on less than 10 cases. 'Complete data only' method refers to handling of missing data. It means only patients with valid ratings on all required initial and final scales are included (see text).

Table 38 (cont). Outcome and change in net prevalence, MTR2 marker, by primary diagnosis and high intensity treatment approach (part 8)

MTR2 Outcomes - Complete data only

Primary diagnosis	CBT and counselling		Low intensity only	
	MTR2 recovery	MTR2 net prevalence change	MTR2 recovery	MTR2 net prevalence change
Depressive episode	38.8% (32.6% to 45.4%) n=219	-33.2% (-40.3% to -26.1%) n=238	30.2% (28.0% to 32.5%) n=1,562	-25.5% (-28.2% to -22.9%) n=1,735
Mixed anxiety and depressive disorder	24.8% (19.7% to 30.7%) n=234	-22.4% (-28.5% to -16.2%) n=246	31.2% (29.0% to 33.5%) n=1,609	-27.6% (-30.1% to -25.0%) n=1,735
Generalized anxiety disorder	35.1% (26.9% to 44.4%) n=111	-30.6% (-40.6% to -20.6%) n=121	43.7% (40.9% to 46.5%) n=1,190	-34.9% (-38.1% to -31.7%) n=1,395
Recurrent depressive disorder	21.7% (12.3% to 35.6%) n=46	-14.3% (-27.4% to -1.2%) n=49	29.5% (24.9% to 34.4%) n=353	-24.7% (-30.3% to -19.1%) n=393
Obsessive-compulsive disorder	25.0% (8.9% to 53.2%) n=12	-21.4% (-52.5% to +9.7%) n=14	38.9% (24.8% to 55.1%) n=36	-30.2% (-48.8% to -11.7%) n=43
Posttraumatic stress disorder		-27.3% (-66.8% to +12.2%) n=11	37.9% (22.7% to 56.0%) n=29	-27.3% (-47.3% to -7.2%) n=33
Agoraphobia			37.8% (28.1% to 48.6%) n=82	-32.3% (-44.3% to -20.2%) n=93
Social phobias			43.7% (32.7% to 55.2%) n=71	-40.5% (-52.7% to -28.4%) n=74
Disappearance / death of family member	0.0% (0.0% to 27.8%) n=10	0.0% (0.0% to 0.0%) n=10	16.0% (6.4% to 34.7%) n=25	-9.4% (-30.9% to +12.1%) n=32
Specific phobias			45.5% (33.0% to 58.5%) n=55	-41.0% (-55.6% to -26.4%) n=61
Other specified categories	21.4% (7.6% to 47.6%) n=14	-11.8% (-40.0% to +16.5%) n=17	29.1% (24.0% to 34.8%) n=268	-25.4% (-31.7% to -19.2%) n=291
Mental disorder NOS	42.3% (33.6% to 51.6%) n=111	-33.9% (-44.3% to -23.4%) n=127	38.7% (35.9% to 41.5%) n=1,145	-31.1% (-34.3% to -27.8%) n=1,333
Missing or illegal code	38.2% (32.6% to 44.2%) n=267	-33.0% (-39.5% to -26.5%) n=294	36.8% (34.7% to 39.0%) n=1,946	-29.9% (-32.4% to -27.4%) n=2,229
Total	33.7% (30.9% to 36.6%) n=1,047	-28.8% (-32.0% to -25.6%) n=1,143	35.3% (34.2% to 36.3%) n=8,371	-29.3% (-30.5% to -28.1%) n=9,447

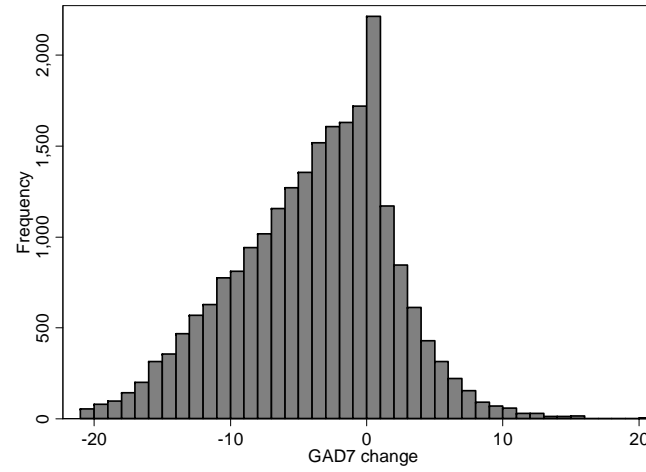
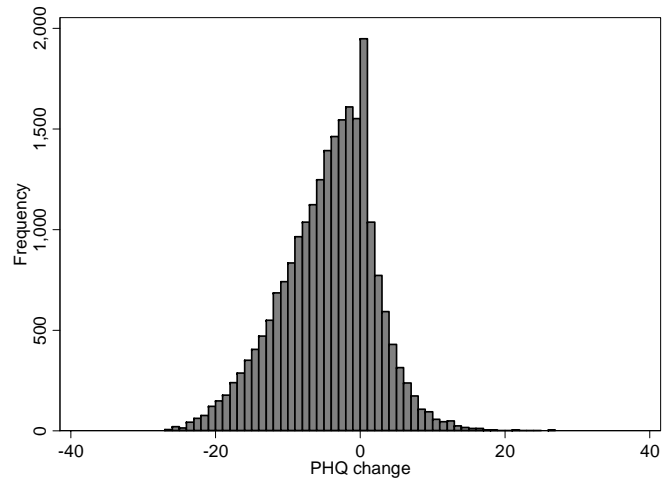
Table shows recovery and prevalence change rates as percentages, 95% confidence interval in parentheses, and overall number in calculation. Data suppressed when calculations based on less than 10 cases. 'Complete data only' method refers to handling of missing data. It means only patients with valid ratings on all required initial and final scales are included (see text).

Table 38a Multivariate analysis of factors associated with movement towards symptomatic recovery using five markers

Predicting:	MTR1		MTR2		MTRDEP		MTRAnx		MTRPhob	
Observations	12856		12259		11112		11915		7707	
Predictor	Odds Ratio	p	Odds Ratio	p	Odds Ratio	p	Odds Ratio	p		
PHQ-9 first score	0.61 (0.58 to 0.65)	***	0.69 (0.64 to 0.73)	***	0.48 (0.44 to 0.52)	***	0.72 (0.68 to 0.77)	***	0.86 (0.79 to 0.93)	***
GAD-7 first score	0.69 (0.65 to 0.73)	***	0.80 (0.76 to 0.85)	***	0.87 (0.82 to 0.92)	***	0.58 (0.54 to 0.62)	***	1.02 (0.95 to 1.10)	
Phobia questions total	0.87 (0.83 to 0.90)	***	0.68 (0.65 to 0.71)	***	0.84 (0.80 to 0.88)	***	0.84 (0.80 to 0.87)	***	0.55 (0.52 to 0.59)	***
W&SAS first score	0.83 (0.78 to 0.87)	***	0.82 (0.78 to 0.87)	***	0.79 (0.75 to 0.83)	***	0.87 (0.82 to 0.91)	***	0.79 (0.74 to 0.85)	***
Start										
Month	0.92 (0.91 to 0.94)	***	0.92 (0.91 to 0.94)	***	0.92 (0.91 to 0.94)	***	0.98 (0.90 to 1.06)	***	0.95 (0.86 to 1.05)	
Whether using psychotropics					0.92 (0.85 to 1.01)	*	0.92 (0.91 to 0.93)		0.94 (0.92 to 0.96)	***
Compared with Male										
Female					0.96 (0.88 to 1.05)		1.00 (0.92 to 1.09)		0.97 (0.88 to 1.07)	
Compared with Age 35 to 64										
Under 18					0.74 (0.45 to 1.22)		0.72 (0.44 to 1.19)		2.00 (1.04 to 3.83)	**
Age 18to34	0.91 (0.84 to 0.99)	**	0.92 (0.85 to 1.00)	*	0.94 (0.86 to 1.02)		0.91 (0.84 to 0.99)	**	1.95 (1.02 to 3.74)	**
Age 65Plus					1.27 (1.01 to 1.59)	**	1.08 (0.88 to 1.34)		2.11 (1.05 to 4.24)	**
Compared with White British										
Minority White			0.85 (0.73 to 1.00)	*	1.03 (0.87 to 1.23)		0.89 (0.75 to 1.05)		0.84 (0.69 to 1.02)	*
Mixed					0.86 (0.62 to 1.18)		0.88 (0.64 to 1.20)		0.94 (0.65 to 1.37)	
Asian	0.69 (0.53 to 0.89)	**	0.74 (0.56 to 0.97)	**	0.71 (0.54 to 0.93)	**	0.63 (0.48 to 0.83)	***	0.73 (0.54 to 0.98)	**
Black			0.78 (0.58 to 1.04)	*	0.74 (0.55 to 0.99)	**	0.83 (0.62 to 1.11)		0.99 (0.72 to 1.37)	
Other	0.70 (0.49 to 1.00)	**	0.73 (0.51 to 1.06)	*	0.82 (0.58 to 1.17)		0.70 (0.49 to 1.00)	**	0.74 (0.50 to 1.11)	
Compared with Referred by GP										
Self referred			1.15 (1.00 to 1.32)	*	1.19 (1.02 to 1.39)	**	1.08 (0.93 to 1.25)		1.11 (0.93 to 1.32)	
Referred by other source					1.00 (0.85 to 1.17)		0.88 (0.75 to 1.03)		0.85 (0.71 to 1.02)	*

Predicting:	MTR1	MTR2	MTRDEP	MTRAnx	MTRPhob
Primary Diagnosis comparison with Depressive episode					
Alcohol problems		0.36 (0.15 to 0.88) **	0.43 (0.21 to 0.90) **	0.60 (0.30 to 1.18)	0.65 (0.26 to 1.66)
Bipolar disorder			0.99 (0.43 to 2.25)	1.12 (0.48 to 2.64)	0.89 (0.27 to 2.94)
Recurrent depressive disorder	0.82 (0.71 to 0.95) **	0.88 (0.75 to 1.02) *	0.75 (0.64 to 0.88) ***	0.89 (0.76 to 1.05)	0.89 (0.73 to 1.08)
Generalised anxiety disorder	1.13 (1.03 to 1.26) **	1.12 (1.00 to 1.24) **	1.26 (1.11 to 1.44) ***	1.09 (0.96 to 1.23)	0.98 (0.84 to 1.15)
Mixed anxiety and depressive disorder			0.92 (0.83 to 1.02)	0.93 (0.83 to 1.03)	0.93 (0.81 to 1.06)
Agoraphobia			1.20 (0.88 to 1.63)	1.12 (0.84 to 1.49)	0.80 (0.60 to 1.08)
Social phobia			1.10 (0.80 to 1.51)	0.98 (0.74 to 1.32)	0.76 (0.57 to 1.01) *
Specific phobias		0.68 (0.50 to 0.93) **	1.44 (0.90 to 2.32)	1.18 (0.82 to 1.69)	0.51 (0.36 to 0.71) ***
OCD			1.07 (0.81 to 1.41)	0.92 (0.71 to 1.18)	0.81 (0.59 to 1.10)
PTSD			1.17 (0.86 to 1.59)	0.98 (0.73 to 1.32)	0.91 (0.66 to 1.26)
Somatoform disorder			0.90 (0.47 to 1.72)	0.79 (0.45 to 1.39)	0.71 (0.35 to 1.44)
Eating disorders			0.69 (0.39 to 1.21)	0.79 (0.44 to 1.40)	0.88 (0.45 to 1.73)
Family loss	0.68 (0.47 to 1.00) **		0.68 (0.46 to 0.99) **	0.75 (0.51 to 1.09)	0.62 (0.37 to 1.02) *
Sites adding significantly to the model	9	9	25	1	19

Charts 25 to 28. Histograms to show the distributions of change scores for each of the symptom scales and the Work and social Adjustment Scale (Last score minus first score)



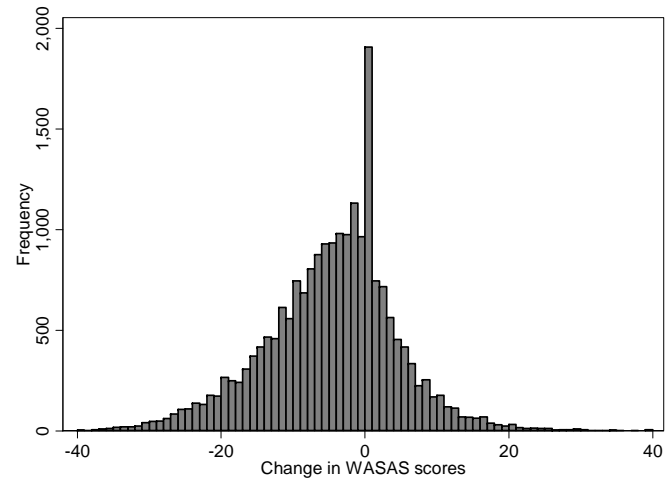
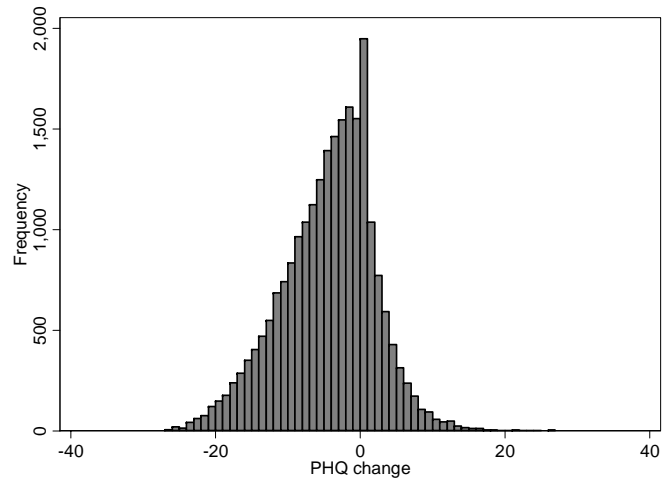


Table 39 Symptom score changes by major categorical variables, study group patients with finished episodes, and two or more contacts.

Gender	PHQ-9	GAD-7	Phobia total	W&SAS	n
Male	-4.67 (+6.45)	-4.23 (+5.73)	-1.86 (+5.24)	-4.99 (+9.12)	7830
Female	-4.98 (+6.53)	-4.42 (+5.87)	-2.10 (+5.41)	-5.20 (+9.36)	14883
Total	-4.87 (+6.51) F=12.31, p=0.0005	-4.35 (+5.82) F=5.61, p=0.0179	-2.02 (+5.35) F=9.17, p=0.0025	-5.13 (+9.28) F=2.38, n.s.	22713
Broad age group	PHQ-9	GAD-7	Phobia total	W&SAS	n
Under18	-5.01 (+7.01)	-3.95 (+5.73)	-1.64 (+5.64)	-4.16 (+8.68)	204
18to34	-4.73 (+6.32)	-4.38 (+5.75)	-2.10 (+5.24)	-5.16 (+9.09)	8416
35to64	-4.99 (+6.63)	-4.35 (+5.90)	-2.02 (+5.49)	-5.23 (+9.49)	11990
65Plus	-5.00 (+6.50)	-4.61 (+5.66)	-2.15 (+5.20)	-4.36 (+8.62)	960
Total	-4.89 (+6.51) F=2.79, p=0.0389	-4.37 (+5.83) F=0.94, n.s.	-2.06 (+5.38) F=0.79, n.s.	-5.15 (+9.29) F=2.9, p=0.0335	21570
Broad ethnic group	PHQ-9	GAD-7	Phobia total	W&SAS	n
White British	-4.99 (+6.55)	-4.47 (+5.84)	-2.04 (+5.26)	-5.22 (+9.15)	14344
Minority White	-4.41 (+6.21)	-3.80 (+5.65)	-1.74 (+5.82)	-5.07 (+9.58)	1049
Mixed	-3.84 (+6.92)	-3.63 (+5.77)	-1.38 (+5.38)	-4.35 (+9.83)	258
Asian	-3.85 (+6.74)	-3.68 (+5.96)	-1.71 (+5.86)	-4.03 (+10.37)	363
Black	-4.97 (+6.98)	-4.11 (+6.26)	-2.30 (+6.55)	-5.59 (+11.53)	361
Other	-4.28 (+6.71)	-3.57 (+5.50)	-2.02 (+6.53)	-4.62 (+10.52)	189
Total	-4.90 (+6.55) F=5.28, p=0.0001	-4.38 (+5.84) F=5.49, p=0	-2.01 (+5.37) F=1.65, n.s.	-5.17 (+9.30) F=1.72, n.s.	16564
Broad referral source	PHQ-9	GAD-7	Phobia total	W&SAS	n
GP	-4.91 (+6.52)	-4.39 (+5.83)	-1.99 (+5.32)	-5.12 (+9.26)	19699
Self	-4.69 (+6.34)	-4.19 (+5.70)	-2.12 (+5.28)	-5.71 (+9.35)	1747
Other	-4.64 (+6.55)	-4.07 (+5.83)	-2.27 (+5.74)	-4.45 (+9.27)	1481
Total	-4.87 (+6.50) F=1.87, n.s.	-4.35 (+5.82) F=2.76, n.s.	-2.02 (+5.35) F=1.98, n.s.	-5.12 (+9.27) F=6.86, p=0.0011	22927

Tables show mean (and standard deviation) of change in symptom score (last rating minus first rating). Records with missing data omitted.

Table 40 Symptom score changes by primary diagnosis and type of high intensity treatment, study group patients with finished episodes, and two or more contacts.

Primary diagnosis	PHQ-9	GAD-7	Phobia total	W&SAS	n
Mixed anxiety depressive disorder	-5.12 (+6.69)	-4.43 (+5.97)	-1.94 (+5.52)	-5.04 (+9.43)	4230
Depressive episode	-5.35 (+6.77)	-4.06 (+5.75)	-1.47 (+5.06)	-5.54 (+9.72)	4167
Generalised anxiety disorder	-4.45 (+5.90)	-4.99 (+5.63)	-2.47 (+5.17)	-4.99 (+8.29)	2966
Recurrent depressive disorder	-5.01 (+6.84)	-3.80 (+5.73)	-1.88 (+5.25)	-5.37 (+9.94)	1053
OCD	-3.96 (+5.81)	-4.95 (+5.58)	-2.08 (+5.08)	-4.87 (+8.43)	327
Agoraphobia	-4.64 (+6.11)	-5.15 (+5.66)	-4.59 (+6.89)	-5.78 (+9.35)	275
PTSD	-5.08 (+7.33)	-5.04 (+6.17)	-3.18 (+6.98)	-6.16 (+10.36)	257
Social Phobia	-3.94 (+5.62)	-4.10 (+5.46)	-3.48 (+5.67)	-4.81 (+7.78)	251
Specific Phobias	-3.16 (+5.25)	-3.83 (+5.45)	-3.95 (+6.45)	-4.82 (+7.89)	222
Family Loss	-4.68 (+5.74)	-3.53 (+5.27)	-1.02 (+4.83)	-4.33 (+8.41)	194
Somatoform disorder	-3.51 (+5.47)	-3.99 (+5.94)	-1.34 (+5.94)	-4.32 (+7.18)	75
Eating Disorders	-3.52 (+5.69)	-2.44 (+4.50)	-1.17 (+5.56)	-2.49 (+7.90)	73
Mental / behavioural disorders due to use of alcohol	-3.57 (+6.37)	-2.96 (+6.89)	-1.31 (+5.88)	-0.26 (+12.46)	49
Bipolar disorder	-3.86 (+6.86)	-4.04 (+4.44)	-2.05 (+5.14)	-5.68 (+10.02)	29
Other specified mental disorders	-3.89 (+6.46)	-4.62 (+6.04)	-3.48 (+5.66)	-5.30 (+8.65)	398
Other diagnoses	-6.37 (+7.49)	-6.41 (+6.31)	-0.89 (+6.12)	-7.45 (+8.34)	30
Mental disorder not otherwise specified	-4.76 (+6.41)	-4.31 (+5.86)	-1.86 (+5.17)	-5.08 (+9.08)	3359
Total	-4.87 (+6.49) F=5.68, p=0	-4.38 (+5.81) F=5.64, p=0	-2.03 (+5.37) F=13.32, p=0	-5.16 (+9.21) F=2.11, p=0.0059	17955

High intensity treatment type	PHQ-9	GAD-7	Phobia total	W&SAS	n
No high intensity treatment	-4.61 (+6.20)	-4.17 (+5.65)	-1.95 (+5.17)	-4.90 (+9.07)	11097
CBT	-5.08 (+6.69)	-4.72 (+6.00)	-2.45 (+5.59)	-5.60 (+9.38)	6970
Counselling	-5.13 (+6.83)	-4.18 (+5.87)	-1.49 (+5.22)	-4.81 (+9.33)	3935
CBT and Counselling	-5.32 (+6.91)	-4.53 (+5.98)	-1.78 (+5.65)	-5.44 (+9.96)	1161
Total	-4.88 (+6.50) F=12.46, p=0	-4.36 (+5.82) F=14.16, p=0	-2.02 (+5.35) F=26.65, p=0	-5.13 (+9.26) F=9.35, p=0	23163

Table 40a Multivariate analysis of factors affecting changes in symptom scores

Predicting:	PHQ-9 Change		Gad 7 Change		W&SAS Change	
Observations	14992		14927		13538	
Proportion of variance explained	0.216		0.237		0.243	
Predictor	Coefficient	p	Coefficient	p	Coefficient	p
Constant	-0.21 (-0.25 to -0.17)	***	-0.20 (-0.24 to -0.17)	***	-0.05 (-0.09 to -0.01)	**
PHQ-9 first score	-0.60 (-0.63 to -0.58)	***	0.16 (0.14 to 0.19)	***	0.18 (0.15 to 0.20)	***
GAD-7 first score	0.07 (0.05 to 0.10)	***	-0.63 (-0.65 to -0.60)	***	0.02 (0.00 to 0.04)	*
Phobia questions total	0.09 (0.08 to 0.11)	***	0.10 (0.08 to 0.12)	***	0.13 (0.11 to 0.15)	***
W&SAS first score	0.12 (0.10 to 0.14)	***	0.08 (0.06 to 0.10)	***	-0.64 (-0.66 to -0.61)	***
Start Month	0.03 (0.03 to 0.04)	***	0.03 (0.03 to 0.04)	***	0.02 (0.02 to 0.03)	***
Whether using psychotropics at initial assessment	0.05 (0.02 to 0.08)	***				
Compared with Male						
Female	0.03 (0.00 to 0.06)	*			0.04 (0.01 to 0.07)	**
Compared with Age 35 to 64						
Under 18						
Age 18to34						
Age 65Plus	-0.06 (-0.13 to 0.00)	*	-0.08 (-0.15 to -0.01)	**	-0.14 (-0.20 to -0.07)	***
Compared with White British						
Minority White			0.07 (0.01 to 0.12)	**		
Mixed	0.13 (0.01 to 0.25)	**	0.12 (0.00 to 0.24)	**		
Asian	0.20 (0.10 to 0.29)	***	0.18 (0.09 to 0.28)	***	0.12 (0.01 to 0.22)	**
Black						
Other			0.20 (0.08 to 0.32)	***		
Compared with Referred by GP						
Self referred	-0.05 (-0.10 to 0.00)	*	-0.06 (-0.11 to -0.01)	**	-0.06 (-0.11 to 0.00)	**
Referred by other source	0.06 (0.01 to 0.12)	**	0.05 (0.00 to 0.11)	*	0.11 (0.05 to 0.17)	***

Table 40a (cont) Multivariate analysis of factors affecting changes in symptom scores

Predicting:	PHQ-9 Change		Gad 7 Change		W&SAS Change	
Primary diagnosis - comparison with depressive episode						
Alcohol problems	0.31 (0.08 to 0.54)	**	0.23 (-0.03 to 0.48)	*		
Bipolar disorder						
Recurrent depressive disorder	0.08 (0.03 to 0.14)	**	0.06 (0.00 to 0.11)	**	0.08 (0.02 to 0.13)	**
Generalised anxiety disorder	-0.12 (-0.16 to -0.09)	***			-0.11 (-0.15 to -0.08)	***
Mixed anxiety depressive disorder			0.07 (0.03 to 0.10)	***		
Agoraphobia	-0.18 (-0.28 to -0.08)	***	-0.10 (-0.20 to 0.00)	**	-0.15 (-0.26 to -0.04)	**
Social phobias	-0.14 (-0.22 to -0.06)	***				
Specific phobias	-0.16 (-0.25 to -0.07)	***			-0.22 (-0.31 to -0.13)	***
OCD			0.13 (0.04 to 0.21)	**		
PTSD						
Somatoform disorder			0.15 (-0.01 to 0.32)	*		
Eating disorder	0.17 (0.00 to 0.34)	*				
Family loss			0.09 (-0.02 to 0.20)	*		
Sites adding significantly to the model		10		8		10

Table 41. Overall changes in symptom scores: means, standard deviations and effect sizes. Remission and recovery rates (with 95% confidence intervals) for comparison with Pilot studies (see text). All patients with finished episodes comprising at least two contacts; each section omits patients missing either first or last rating on the scale concerned.

Method of ending treatment	Completed	Declined or Dropped out	Not suitable	Other	Sites Combined
PHQ-9 ratings changes					
Patients	12,396	6,001	1,186	3,580	23,163
Pre-treatment					
Mean (s.d.)	12.97 (6.57)	14.99 (6.39)	16.14 (6.42)	14.60 (6.62)	13.91 (6.61)
Median (IQR)	13 (8 - 18)	16 (10 - 20)	17 (11 - 21)	15 (10 - 20)	14 (9 - 19)
Post-treatment					
Mean (s.d.)	6.77 (6.38)	11.81 (6.81)	14.40 (7.30)	10.43 (7.00)	9.03 (7.11)
Median (IQR)	12 (7 - 16)	14 (9 - 18)	14 (10 - 18)	13 (8 - 17)	13 (8 - 17)
Effect size	0.97	0.47	0.24	0.60	0.69
Case rate					
Pre-treatment	67.4% (66.6% to 68.3%)	78.6% (77.5% to 79.6%)	82.7% (80.5% to 84.8%)	75.8% (74.3% to 77.1%)	72.4% (71.8% to 73.0%)
Post-treatment	26.9% (26.1% to 27.7%)	60.1% (58.9% to 61.4%)	73.0% (70.4% to 75.5%)	49.7% (48.1% to 51.4%)	41.4% (40.8% to 42.0%)
'Remission rate' (Post / Pre)	0.40 (0.39 to 0.41)	0.77 (0.75 to 0.78)	0.88 (0.85 to 0.92)	0.66 (0.63 to 0.68)	0.57 (0.56 to 0.58)
GAD-7 ratings changes					
Patients	12,384	5,944	1,169	3,540	23,037
Pre-treatment					
Mean (s.d.)	11.75 (5.52)	13.22 (5.34)	13.67 (5.28)	12.67 (5.51)	12.37 (5.50)
Median (IQR)	12 (7 - 16)	14 (9 - 18)	14 (10 - 18)	13 (8 - 17)	13 (8 - 17)
Post-treatment					
Mean (s.d.)	6.09 (5.47)	10.51 (5.85)	12.10 (5.93)	9.22 (5.99)	8.01 (6.08)
Median (IQR)	4 (2 - 9)	10 (6 - 15)	12 (7 - 17)	8 (4 - 14)	7 (3 - 13)
Effect size	1.04	0.46	0.27	0.58	0.72
Case rate					
Pre-treatment	74.1% (73.3% to 74.9%)	82.2% (81.2% to 83.1%)	83.8% (81.6% to 85.8%)	79.1% (77.7% to 80.4%)	77.5% (76.9% to 78.0%)
Post-treatment	30.6% (29.8% to 31.4%)	64.6% (63.4% to 65.8%)	74.9% (72.4% to 77.3%)	54.5% (52.8% to 56.1%)	45.3% (44.6% to 45.9%)
'Remission rate' (Post / Pre)	0.41 (0.40 to 0.42)	0.79 (0.77 to 0.80)	0.89 (0.86 to 0.93)	0.69 (0.67 to 0.71)	0.58 (0.58 to 0.59)

Table 42 Spearman's rank correlations between sites scores on recovery rates, and simple service characteristics.

	MTR1 recovery rate	PHQ-9 recovery rate	GAD-7 recovery rate	Phobia questions recovery rate	Study group size	Proportion still in system	Proportion finishing by completing	Proportion of therapists high intensity I	Proportion of high intensity therapy which is CBT	Proportion of patients having low intensity only	Step fraction
MTR1 recovery rate	1										
PHQ-9 recovery rate	0.945 ***	1									
GAD-7 recovery rate	0.974 ***	0.943 ***	1								
Phobia questions recovery rate	0.795 ***	0.778 ***	0.771 ***	1							
Study group size	0.104	0.150	0.061	0.228	1						
Proportion of patients still in system	0.124	0.193	0.182	0.056	-0.352	1					
Proportion of patients finishing by completing	0.305	0.310	0.305	0.403 *	-0.087	0.071	1				
Proportion of therapists high intensity	0.477 **	0.432 *	0.465 *	0.372 *	-0.338	0.422 *	0.261	1			
Proportion of high intensity therapy which is CBT	0.178	0.080	0.199	0.158	-0.467 **	0.170	0.072	0.308	1		
Proportion of patients having low intensity only	-0.147	-0.062	-0.167	-0.065	-0.048	-0.214	0.137	-0.430 *	0.035	1	

Step fraction	-0.067	-0.063	-0.132	-0.010	0.288	-0.471 **	-0.181	-0.433 *	-0.123	0.394 *	1
---------------	--------	--------	--------	--------	-------	-----------	--------	----------	--------	---------	---

Table 43. Treatment response in corresponding phobia question for patients with primary or secondary diagnosis of a phobic state.

	Initial cases	Total patients	Complete data only		
			Recovery rate	Incidence rate	Net change in prevalence
Diag. Social phobia				Using phobia question 1	
Low intensity treatment	55	82	58.2% (45.0% to 70.3%)	14.8% (5.9% to 32.5%)	-34.1% (-48.5% to -19.8%)
CBT	119	181	47.1% (38.3% to 56.0%)	8.1% (3.5% to 17.5%)	-28.2% (-38.1% to -18.3%)
Diag. Agoraphobia				Using phobia question 2	
Low intensity treatment	66	99	53.0% (41.2% to 64.6%)	21.2% (10.7% to 37.8%)	-28.3% (-41.6% to -14.9%)
CBT	134	171	44.8% (36.6% to 53.2%)	21.6% (11.4% to 37.2%)	-30.4% (-40.1% to -20.7%)
Diag. Specific phobia				Using phobia question 3	
Low intensity treatment	41	67	65.9% (50.5% to 78.4%)	19.2% (8.5% to 37.9%)	-32.8% (-48.7% to -16.9%)
CBT	89	139	52.8% (42.5% to 62.8%)	30.0% (19.1% to 43.8%)	-23.0% (-34.4% to -11.6%)
				Second unknown - no change	
Diag. Social phobia				Using phobia question 1	
Low intensity treatment	61	95	52.5% (40.2% to 64.5%)	11.8% (4.7% to 26.6%)	-29.5% (-43.1% to -15.9%)
CBT	129	197	43.4% (35.2% to 52.0%)	7.4% (3.2% to 16.1%)	-25.9% (-35.4% to -16.4%)
Diag. Agoraphobia				Using phobia question 2	
Low intensity treatment	83	121	42.2% (32.1% to 52.9%)	18.4% (9.2% to 33.4%)	-23.1% (-35.3% to -11.0%)
CBT	151	193	39.7% (32.3% to 47.7%)	19.0% (10.0% to 33.3%)	-26.9% (-36.1% to -17.8%)
Diag. Specific phobia				Using phobia question 3	
Low intensity treatment	50	83	54.0% (40.4% to 67.0%)	15.2% (6.7% to 30.9%)	-26.5% (-41.1% to -11.9%)
CBT	95	151	49.5% (39.6% to 59.4%)	26.8% (17.0% to 39.6%)	-21.2% (-32.2% to -10.2%)

Annex 1. Dataset finally requested from providers.

Dataset for IAPT rollout audit

This document sets out the specifications for a data extract required from IAPT services for the 2008/9 and 2009/10 audits of progress in IAPT implementation. It comprises a subset of the items specified in the document 'Improving Access to Psychological Therapies (IAPT) Outcomes Toolkit 2008/9'.

IAPT service providers will be required to submit files of data relating to periods of time as set out in table 1.

Table 1: Periods covered by audit submissions and timescale for data returns

Data period label	Start of period	End of period	Data file required by

Format of data files

Data files should be submitted as standard Excel format spreadsheets.

One datafile should be submitted by each IAPT service provider, for each data period.

Data files should comprise one row of data for each case assessed and/or treated during the period. Where a person's treatment spans two periods, an appropriate record should be included in both, using the most up-to-date information available at the point of data submission. Referrals not leading to an assessment should not be included.

The fields required for each record are set out in table 2. All are items detailed in the IAPT dataset described in detail in Appendix A of the document *Improving Access to Psychological Therapies (IAPT) Outcomes Toolkit 2008/9*. The first column in table 2 indicates the reference number of the data item in the *Outcomes Toolkit*. In some cases these are simple: the return required should be stored in one place only and in exactly the form required. There are three areas of deviation from this.

1. In the case of rating scales (GAD-7, PHQ-9 and Social and Work Adjustment Scale), in addition to the total scores (identified in the Outcomes Toolkit) for audit purposes the full rating detail is required. Precise formats for these are set out after table 2. This is to examine the extent to which data items are missing from the summary scores and the extent to which the profile of positively rated items alters in addition to changes in the overall total scores.

2. In the case of all 'last' state ratings, dates have been requested. These items are rated at every appointment the patient has. Hence it is not clear whether the final ratings will

all be contemporary, and likely that in some cases they will not be. Dates are requested here to check this aspect of data quality.

3. Two items, intervention given and therapist 'Agenda for Change' band, are rated at every appointment and it is not clear that these will remain the same. In cases where patients are 'stepped' from one therapeutic approach to another, they will change. Service providers are asked to produce summary figures for these with one field for each possible answer, and entries indicating the number of appointments each patient had with the corresponding rating.

Table 2. Detailed specification of data required.

Data item reference number	Data Item	Column heading	Format	Notes
Administrative details				
P3	Organisation code (code of provider)	orgcodeprov	an5	To identify the organisation providing the period of care
Details of patient				
P5	Patients gender	ptgender	n1	
From P6	Patients age at start of episode	Derived from ptdob	an10	
P7	Ethnic category	ptethcat	an2	The patients self-identified ethnic category
P9(i)	Visual disability	visdis	n1	
P9(ii)	Speech disability	speechdis	n1	
P9(iii)	Hearing disability	heardis	n1	
P9(iv)	Mobility disability	mobdis	n1	
P10	Able to communicate in spoken English	spokenenglish	n1	
Care pathway data				
C5	Date of initial assessment	Dtassess	an10	
C6	Date of first therapeutic session	dtfirstther	an10	This may be blank if treatment has not started
C7	Date of end of care pathway	dtendcare	an10	This may be blank if pathway has not ended
C8	Reason for end of IAPT care pathway	reasonend	n1	
Details of problem				

at presentation				
C4	Primary diagnosis	primdiag	ann	List of permissible values set out in <i>Outcomes Toolkit</i> .
C9	PHQ-9 first score total	phqfirst	n2	
Source for C9	PHQ-9 first score detail	phqfirstdetail	an9	These scores are required to produce C9, but are requested in full here to establish the extent of missing item data and the degree of similarity of positive items at assessment and termination.
C11	GAD-7 first score total	GAD-7first	n2	
Source for C11	GAD-7 first score detail	GAD-7firstdetail	an7	These scores are required to produce C11, but are requested in full here to establish the extent of missing item data and the degree of similarity of positive items at assessment and termination.
C13	Work and Social Adjustment Scale first score	wasasfirst	n2	
Source for C13	Work and Social Adjustment Scale first score detail	wasasfirstdetail	an5	These scores are required to produce C13, but are requested in full here to establish the extent of missing item data and the degree of similarity of positive items at assessment and termination.
C32	Presenting problem	presprob	an255	Reason for referral as described by patient
C33	Secondary diagnosis	secdiag	annn	ICD10 codes from list provided or other legal ICD10 codes.
C15	Employment status first	empstatfirst	n2	Employment status at first

				assessment
C17	Sick pay status first	sickpayfirst	n2	Sick pay status at first assessment during care spell
C19	Benefits status first	benstatfirst	n1	
C21	Phobia question 1 first score	phobq1first	n1	
C23	Phobia question 2 first score	phobq2first	n1	
C25	Phobia question 3 first score	phobq3first	n1	
C28	Use of psychotropic medication first	psytropfirst	n1	
Patients condition at latest session for each rating				
C10	PHQ-9 last score total	phqlast	n2	
Metadata for C10	Date of last PHQ-9 score	datephqlast	an10	Date last PHQ-9 recorded
Source for C10	PHQ-9 last score detail	phqlastdetail	an9	These scores are required to produce C10, but are requested in full here to establish the extent of missing item data and the degree of similarity of positive items at assessment and termination.
C12	GAD-7 last score total	GAD-7last	N2	
Metadata for C12	Date of last GAD rating	dateGAD-7last	an10	
Source for C12	GAD-7 last score detail	GAD-7lastdetail	an7	These scores are required to produce C12, but are requested in full here to establish the extent of missing item data and the degree of similarity of positive items at assessment and termination.
C14	Work and Social Adjustment Scale last score	wasaslast	n2	

Metadata for C	Date of Work and Social Adjustment Scale last score	datewasaslast		
Source for C14	Work and Social Adjustment Scale last score detail	wasaslastdetail	an5	These scores are required to produce C14, but are requested in full here to establish the extent of missing item data and the degree of similarity of positive items at assessment and termination.
C16	Employment status last	empstatlast	n2	Employment status at last assessment during care spell
Metadata for C	Date of Employment status last	dateempstatlast		
C18	Sick pay status last	sickpaylast	n2	Sick pay status at last assessment during care spell
Metadata for C	Date of Sick pay status last	datesickpaylast		
C20	Benefits status first	benstatlast	n1	
Metadata for C	Date of Benefits status first	datebenstatlast		
C22	Phobia question 1 last score	phobq1last	n1	
Metadata for C	Date of Phobia question 1 last score	datephobq1last		
C24	Phobia question 2 last score	phobq2last	n1	
Metadata for C	Date of Phobia question 2 last score	datephobq2last		
C26	Phobia question 3 last score	phobq3last	n1	
Metadata for C	Date of Phobia question 3 last score	datephobq3last		
C28	Use of psychotropic medication last	psytroplast	n1	
Metadata for C	Date of Use of psychotropic medication last	datepsytroplast		
Interventions given				
From A4	Intervention given: cCBT (code 01)	interv01num	n3	Indicate number of appointments for which the A4 field in the record for the patients care indicates the

				intervention given as 01: cCBT. If none, enter 0.
From A4	Intervention given: Pure self help (code 02)	interv02num	n3	Indicate number of appointments for which the A4 field in the record for the patients care indicates the intervention given as 02 Pure self help. If none, enter 0.
From A4	Intervention given: Guided self-help (code 03)	interv03num	n3	Indicate number of appointments for which the A4 field in the record for the patients care indicates the intervention given as 03: Guided self-help. If none, enter 0.
From A4	Intervention given: Behavioural activation (code 04)	interv04num	n3	Indicate number of appointments for which the A4 field in the record for the patients care indicates the intervention given as 04: Behavioural activation. If none, enter 0.
From A4	Intervention given: Structured exercise (code 05)	interv05num	n3	Indicate number of appointments for which the A4 field in the record for the patients care indicates the intervention given as 05: Structured exercise. If none, enter 0.
From A4	Intervention given: Psychoeducational groups (code 06)	interv06num	n3	Indicate number of appointments for which the A4 field in the record for the patients care indicates the intervention given as 06: Psychoeducational groups. If none, enter 0.
From A4	Intervention given: CBT (code 07)	interv07num	n3	Indicate number of appointments

				for which the A4 field in the record for the patients care indicates the intervention given as 07: CBT. If none, enter 0.
From A4	Intervention given: Interpersonal therapy (code 08)	interv08num	n3	Indicate number of appointments for which the A4 field in the record for the patients care indicates the intervention given as 08: Interpersonal therapy. If none, enter 0.
From A4	Intervention given: Counselling (code 09)	interv09num	n3	Indicate number of appointments for which the A4 field in the record for the patients care indicates the intervention given as 09: Counselling. If none, enter 0.
From A4	Intervention given: Couples therapy (code 10)	interv10num	n3	Indicate number of appointments for which the A4 field in the record for the patients care indicates the intervention given as 10: Couples therapy. If none, enter 0.
From A4	Intervention given: Other (code 11)	interv11num	n3	Indicate number of appointments for which the A4 field in the record for the patients care indicates the intervention given as 11: Other . If none, enter 0.
From A1	Therapist level: Number of appointments with therapist at Agenda For Change band 1	thafc01	n3	Indicate number of appointments for which the A1 field in the record for the patients care indicates the therapist's agenda for change band as Band 1. If none, enter 0.

From A1	Therapist level: Number of appointments with therapist at Agenda For Change band 2	thafc02	n3	Indicate number of appointments for which the A1 field in the record for the patients care indicates the therapist's agenda for change band as Band 2. If none, enter 0.
From A1	Therapist level: Number of appointments with therapist at Agenda For Change band 3	thafc03	n3	Indicate number of appointments for which the A1 field in the record for the patients care indicates the therapist's agenda for change band as Band 3. If none, enter 0.
From A1	Therapist level: Number of appointments with therapist at Agenda For Change band 4	thafc04	n3	Indicate number of appointments for which the A1 field in the record for the patients care indicates the therapist's agenda for change band as Band 4. If none, enter 0.
From A1	Therapist level: Number of appointments with therapist at Agenda For Change band 5	thafc05	n3	Indicate number of appointments for which the A1 field in the record for the patients care indicates the therapist's agenda for change band as Band 5. If none, enter 0.
From A1	Therapist level: Number of appointments with therapist at Agenda For Change band 6	thafc06	n3	Indicate number of appointments for which the A1 field in the record for the patients care indicates the therapist's agenda for change band as Band 6. If none, enter 0.
From A1	Therapist level: Number of appointments with therapist at Agenda For Change band 7	thafc07	n3	Indicate number of appointments for which the A1 field in the record for the patients care indicates the therapist's agenda for change band as Band 7. If none, enter 0.

From A1	Therapist level: Number of appointments with therapist at Agenda For Change band 8a (code 08)	thafc08	n3	Indicate number of appointments for which the A1 field in the record for the patients care indicates the therapist's agenda for change band as Band 08a (code 08). If none, enter 0.
From A1	Therapist level: Number of appointments with therapist at Agenda For Change band 8b (code 10)	thafc10	n3	Indicate number of appointments for which the A1 field in the record for the patients care indicates the therapist's agenda for change band as Band 8b (code 10). If none, enter 0.
From A1	Therapist level: Number of appointments with therapist at Agenda For Change band 8c (code 11)	thafc11	n3	Indicate number of appointments for which the A1 field in the record for the patients care indicates the therapist's agenda for change band as Band 8c (code 11). If none, enter 0.
From A1	Therapist level: Number of appointments with therapist at Agenda For Change band 8d (code 12)	thafc12	n3	Indicate number of appointments for which the A1 field in the record for the patients care indicates the therapist's agenda for change band as Band 8d (code 12). If none, enter 0.

PHQ-9 score detail (required for first and last PHQ-9 scores).

Alphanumeric, 9 characters, 0,1,2,3 or X. This score comprises one rating for each of nine questions as follows:

0	Not at all
1	Several days
2	More than half the days
3	Nearly every day

The answers should be concatenated to form a nine character alphanumeric field. The answer to question 1 should be at position 1, question 2 at position 2, and so on. Missing answers should be reported with X characters.

GAD-7 score detail

Alphanumeric, 7 characters, 0,1,2,3 or X. This score comprises one rating for each of seven questions as follows:

0	Not at all
1	Several days
2	More than half the days
3	Nearly every day

The answers should be concatenated to form a seven character alphanumeric field. The answer to question 1 should be at position 1, question 2 at position 2, and so on. Missing answers should be reported with X characters.

Work and Social Adjustment Scale score detail

Alphanumeric, 5 characters, 0, 1, 2, 3, 4, 5, 6, 7 or 8; X for missing ratings. This score comprises a self assign rating of 0 (no impairment) to 8 (very severe impairment) for each of five questions. The answers should be concatenated to form a five character alphanumeric field. The answer to question 1 should be at position 1, question 2 at position 2, and so on. Missing answers should be reported with X characters.