COMPUTERISED CBT FOR COMMON MENTAL DISORDERS: RCT OF A WORKPLACE INTERVENTION

REPORT TO BRITISH OCCUPATIONAL HEALTH RESEARCH FOUNDATION

Corresponding author:

Justine Schneider, Professor of mental health and social care, University of Nottingham

Department of Sociology & Social Policy
Law & Social Sciences Building
University Park
Nottingham
NG7 2RD

Justine.Schneider@nottingham.ac.uk

Tel: 0115 846 7307

September, 2012

AUTHORS' TITLES AND AFFILIATIONS (ALPHABETICAL)

See Appendix 6 for authors' contributions to the study

Kylie **Bennett**, Australia National University

Anthony **Bennett**, Australia National University

Dr. Paul **Grime**, University of Cambridge and Cambridge University Hospitals

Dr. Bob **Grove**, Centre for Mental Health, London

Dr. Pooria **Sarrami Foroushani**, formerly University of Nottingham, now University of New South Wales

Dr. Morven **Leese**, Institute of Psychiatry, Kings College London

Professor Paul **McCrone**, Institute of Psychiatry, Kings College London

Iris **Molosankwe**, Institute of Psychiatry, Kings College London

Professor Richard **Morris**s, Institute of Mental Health, University of Nottingham

Lesia **Nel**, Institute of Mental Health, Nottingham

Rachel **Phillips**, Institute of Psychiatry, Kings College London

Professor Justine **Schneider**, Institute of Mental Health, University of Nottingham

Professor Graham **Thornicroft**, Institute of Psychiatry, Kings College London

Dr. Paul **Walker**, Institute of Psychiatry, Kings College London

ACKNOWLEDGEMENTS

The study was funded by the British Occupational Health Research Foundation.

This research was made possible through the commitment and support of Catherine Kilfedder Group Health Adviser at BT and Alison Dunn, Head of Treatment Services in the Occupational Health Department at Transport for London. In the final year of the study, with the support of the NHS Confederation mental health lead, Steve Shrubb, a link to the trial portal was added to the website of NHS Employers, enabling individuals with access to the website to enter the study or to link it to their own workforce by any means. The research was met with enthusiasm and support by these employer organisations, their health and wellbeing and occupational health staff, too numerous to mention here, and above all from their employees who voluntarily joined this study and participated by contributing their data, their observations about the process and their time.

The study was executed by an international team of researchers as shown in the table of authors. The clinical studies officers of the UK Mental Health Research Network (East Midlands and South Yorkshire hub) conducted the telephone interviews. The Clinical Trials Unit at the University of Nottingham advised on randomization and generated the algorithm used for this purpose. Kylie Bennett (e-hub development manager) and Anthony Bennett (e-hub IT manager) of The Australian National University developed the portal and managed the data which it collected. Lesia Nel and Pooria Sarrami Foroushani managed the trial; Paul Grime, Bob Grove, Brian Kazer, Morven Leese, Paul McCrone, Richard Morriss, Graham Thornicroft and Paul Walters advised on the study's design. Rachel Phillips and Morven Leese conducted the statistical analyses and wrote the statistical appendices. Iris Molosankwe and Paul McCrone conducted the economic analysis and wrote the relevant sections. Justine Schneider was chief investigator and compiled the report. A positive ethical opinion was granted by Derby Research Ethics Committee and by the Australia National University.

CONTENTS

MAIN FIGURES

ABSTRACT

BACKGROUND

Depression and anxiety are major causes of absence from work and underperformance in the workplace. Cognitive behaviour therapy (CBT) can be effective in treating such problems but the need exceeds the supply of face to face therapists. Computerised CBT has many advantages and has been approved by NICE, but its utility in a workplace setting has not been fully investigated.

DESIGN

A two-arm, parallel, randomised controlled trial with a six week intervention and follow-ups at six and twelve weeks was undertaken to measure, in employees with diagnosable depression, the impact of one computerised CBT intervention, MoodGYM, compared to an 'attentional' control; five websites with general information about mental health. The main outcome was scored on the Work and Social Adjustment Scale. Depression, anxiety, costs and acceptability of the interventions were also assessed.

METHOD

A research portal was designed and implemented to provide information about the study, screen potential participants, record consent and randomise eligible participants. Consent also entailed agreeing to be telephoned once a week for six weeks. Baseline and follow-up measures were administered through the portal, while service use audit and risk assessment were completed by telephone interviewers. Access to the portal was extended to employees of three UK-based employers, over 2 years. Data presented here were analysed from the perspective of outcomes, acceptability and costs. Analyses of presenteeism and job control will follow.

RESULTS

There were 9,305 visits to the website, 1,111 completed the screening and 637 were randomized, with 359 completing six-week and 231 completing twelve-week follow-ups. Longitudinal analysis across both time points showed no evidence for a difference in the average treatment effect on the Work and Social Adjustment Scale score (effect -0.47, 95% CI (-1.84, 0.90), P=0.5) and no evidence for a difference in the average treatment effect for any of the secondary outcomes. Yet most participants saw advantages to cCBT at the outset and retained favourable attitudes when they had used it. Participants in the MoodGYM arm had fewer days off work during the intervention period but this was not statistically significant.

CONCLUSIONS

This study found no evidence that MoodGYM was superior to informational websites in terms of psychological outcomes or service use. However, users of MoodGYM took less time off work during the study period, leading to lower costs for employers. The qualitative comments from participants offer information for further development of Internet interventions to meet the continuing demand for cognitive behaviour therapy.

BACKGROUND

Depression and anxiety are recognised as major causes of underperformance at work[1] and long-term absences of employees. Up to 9% of the UK population is likely to be affected[2] by these treatable mental health issues, and the economic loss is estimated to be of the order of £9 billion annually, mostly due to lost productivity at work. A survey for the UK Department of Work and Pensions found that more than 85% of employees consider that employers should take steps to help depressed employees to carry on working.

Considerable investment has been directed at effective, brief treatments, based largely on cognitive behaviour therapy (CBT)[3]. Interventions of this kind offer a structured approach to recognising and changing negative thinking patterns, which requires active participation by the patient. Such is the demand for CBT that despite the development of brief, stepped interventions through the Increasing Access to Psychological Therapies (IAPT) programme, waiting lists are long, typically up to 6 months for an NHS referral.

Computerised interventions have potential to meet rising demand for CBT, and also appear to have many other advantages. They are almost certain to be less costly than face to face therapy, and are convenient and easy to access thanks to growing use of the Internet. They appear relatively free of the stigma which often attaches to formal mental health services, and computerised interventions can be provided consistently to large numbers of people maintaining rigorous quality standards.

Two computerised CBT (cCBT) packages received approval from the National Institute for Clinical Excellence (NICE) in 2006[4], one for panic and phobia (Fearfighter) and one for mild and moderate depression (Beating the Blues). Primary care providers in England and Wales were then instructed by NICE to make cCBT available to all patients[5]. Evidence about these and other cCBT packages remains promising but inconclusive, making it desirable to investigate the costs and benefits of each package in relation to specific diagnostic groups (Sarrami et al., 2011)[6].

Moreover, the utility of cCBT in an employment setting has not been investigated. With the exception of one small trial in the NHS of Beating the Blues[7], at the outset of this trial we could find no published studies of cCBT in workplace settings.

RATIONALE FOR THIS STUDY

We were commissioned by the British Occupational Health Research Foundation to investigate an intervention that could be administered in the workplace in order to explore the potential benefits to employers of making it available, if effective. MoodGYM was chosen for this study because it is freely available on the Internet[8], has some evidence of effectiveness from one community-based trial, in which users were supported by weekly phone calls. [9],[10] In that study, 79% completed the intervention. Although the dropout was greater from MoodGYM than from the control group, the effect of MoodGYM endured longer than the other intervention.

Beating the Blues (BtB) has been shown to be effective[11],[12],[13] but it is designed to be completed in a routine care setting with close supervision from a professional, so the costs of implementing BtB are much greater. BtB has been trialled in one workplace setting[14]. In that study acceptability was low and drop out was high, leading the researcher to infer that greater flexibility and accessibility might improve uptake. MoodGYM offers this flexibility but had not been tested in an employment context, apart from an unpublished pilot study in an Australian bank.

The hypothesis was that users of MoodGYM would experience less functional impairment at work than users of an alternative; informational websites about mental health.

AIM

To measure the impact of MoodGYM, an interactive cCBT programme, on employees' work-related performance and psychological well-being, compared to that of an 'attentional' control (five websites with general information about mental health).

OUTCOMES

The Work and Social Adjustment Scale (WSAS)[15] was used to measure the primary outcome, subjective, work-related performance. Secondary outcomes included depression rating on the PHQ-9[16], performance at work measured by the Work Limitations Questionnaire (WLQ)[17], self-assessed absence from work, and acceptability of the interventions. Service use (CSRI) and quality of life (EQ-5D) were measured to permit cost-effectiveness analysis of the trial.

METHOD

Design

A two-arm, parallel randomised controlled trial with a six week intervention and follow-ups at six and twelve weeks.

Participants

Employees were given access to the trial website with the co-operation of their organisation's occupational health department. They were free to access the website to find out if they were eligible for the trial or not. On-line screening of potential participants offered the option of joining the trial if they were aged over 18 and met the following criterion: On Patient Health Questionnaire-9 (PHQ-9) employees scored 2 or more on 5 of the 9 items, including 2 or more on item 1 (little interest in doing things) or item 2 (feeling hopeless). To be eligible the employee also had to confirm that at least one of the items identified as a problem for them made it difficult to work, take care of things at home, or get along with other people.

At the outset the UK protocol stated that employees would be excluded (and given advice on other sources of support) if the screening determined that they had a primary alcohol or drug use problem, depression secondary to organic brain disease (e.g. stroke) bipolar disorder or other serious mental illness apart from uni-polar depression, were already receiving cognitive behaviour therapy, had current suicide plans requiring emergency psychiatric treatment or did not give written informed consent to the study.

During the implementation of the trial a pragmatic decision was made to allow the people who screened positive for depression but who fitted the following exclusion criteria to enter the trial: a primary alcohol or drug use problem; depression secondary to organic brain disease (e.g. stroke) bipolar disorder or other serious mental illness apart from uni-polar depression or were already receiving cognitive behaviour therapy. This was explicit in the ethical application to Australia National University1 but not in the UK ethics application. The deviation from the protocol only

[1] "Excerpt from ANU ethics application:

came to light after the study had ended. At that point, the research team decided to retain the group who had initially been deemed ineligible, on the grounds that they had consented to the trial, despite being told that their data 'may not' be used, and that the results had greater real-world relevance (ecological validity) if they were included. Nobody was told that they were not in the study and then included. These people were (inadvertently) not screened out as we had originally intended so they were unaware that they might not have been eligible.

The impact of the research team's considered decision to retain the 41 individuals affected is fully explored in Appendix 5. Their inclusion made no difference to the results of the analyses presented here.

Settings

With support from BOHRF and its members, we were granted access to three UK workforces. The research portal was first piloted with BT, and then implemented incrementally with most of the BT business units. It was promoted through the Health, Safety and Wellbeing section of the Human Resources department using internal communications and regular telephone conferences with key personnel over the duration of the study (2 years). Transport for London joined the trial a few months later and promoted it through their intranet as well as referring individuals through the in-house counseling service, so offering it to employees who were deemed likely to benefit as well as to the wider workforce. After one year in the field, trial recruitment was lagging, so we recruited NHS Employers to the study in October, 2010. This entailed making access available on their website, with minimal marketing. The recruitment numbers reflect the size of the employers and the nature of the workforces as well as the extent to which the study was actively promoted. At baseline assessment, BT recruited 396 (62%), Transport for London 100 (16%), and NHS Employers 141 participants (22%).

Interventions

The MoodGYM intervention is a modularised course developed at Australia National University. It is designed to last five weeks with assessments in the sixth week, although participants proceed at their own pace. It is freely available for use by the public and not restricted to any particular delivery setting.[7] It's website describes it as 'A free self-help program to teach cognitive behaviour therapy skills to people vulnerable to depression and anxiety'. The websites selected for the control group were judged to be reliable sources of information about mental health problems, and are listed in Table 1. They were known to the chief investigator from a previous review of self-help in mental health, and had been identified for teaching purposes as suitable materials to inform health and social care professionals.

Table 1: Website links sent weekly to participants in control arm.

The first time the participant logs into the portal they will be automatically directed to a Demographics Survey, followed by the Possible Exclusion Survey. If a participant answers yes to any of the possible exclusion survey questions, the user is notified that they will be able to access the online self-help introduced in the study, but that their data might not be used."

url	Title on trial portal
http://www.hse.gov.uk/stress	Stress at work
http://www.nhs.uk/Conditions/Mental-health/Pages/Introduction.aspx?url=Pages/What-is-it.aspx	NHS Choices: mental health
http://www.mentalhealth.org.uk/information/mental-health-a-z/	A to Z of mental health problems
http://www.bbc.co.uk/health/conditions/mental_health/emotion_stress.shtml	Tackling stress
http://www.rethink.org/living_with_mental_illness/rights_and_laws/index.html	Your rights to care and treatment

All participants were required to give a telephone number as a condition of joining the study. For the purpose of the trial, weekly telephone calls were made, lasting about ten minutes on average. Their aim was, firstly, to maintain engagement with the study, and secondly to collect service use data for costing purposes. The telephone input was provided to both arms of the trial by the Mental Health Research Network's clinical studies officers. The telephone interviewers were not 'blind' to the status of the participants, but they only recorded service use measures.

Risk of adverse events

The telephone interviewers also screened for risk of self-harm or suicide, and followed a protocol which permitted them to breach confidentiality if participants seem so unwell that immediate professional care was indicated but they were unwilling or unable to access this without assistance. This was invoked twice. On six occasions telephone interviewers contacted the relevant occupational health department with the client's permission. In addition, participants were told at the outset and reminded periodically in the telephone interviews that they should consult their GP or Occupational Health Department (as appropriate for each organisation) if they had intentions to harm themselves.

Development and implementation of the trial website

The trial was designed to be administered on-line; while a novel departure for most of the UK researchers, this kept the costs of the trial exceptionally low. The MoodGYM developers were commissioned by the study investigators to construct a research portal which could allocate login identities, screen, take consent, randomise and direct participants to the appropriate arm of the study. This issued emails which prompted participants to complete assessments at six and twelve weeks. The trial manager could be contacted by email in case of problems, such as forgotten passwords. The portal was piloted in June-September, 2009, following adjustments to the text displayed on screen the trial was launched in November of that year and the trial portal was closed to new recruits on 31 May, 2011.

Ethical oversight

A favourable ethical opinion was granted by Derbyshire Local Research Ethics Committee on January 6, 2009. The Mental Health Research Network adopted the study in February, 2009.

Australia National University's ethics committee also approved the study before the portal was piloted.

Sample size

The power calculation was designed to detect a mean difference of 3 points in the Work and Social Adjustment Scale, which was judged by the study team to be a clinically significant difference. With a standard deviation of 9 and 80% power at a 5% significance level, we required 142 participants per arm to complete the study.

Randomisation

Once potential participants had completed the screening questions, if eligible for inclusion in the trial, they were given a study ID, allocated through the website, and they were then invited to join the trial. A list was produced by the Nottingham Clinical Trials Unit to allow simple (unrestricted) randomisation.[18] If participants consented, they were randomised and sent to the portal designers to be incorporated in its pathway. In this way the randomisation status of participants was concealed from their employers and from the research team until the study was completed.

Blinding

The study design was double-blind. Participants were told that they were participating in a trial of 'on-line self-help', comparing two approaches, and researcher bias was avoided completely, at least with respect to the primary and secondary outcome measures administered through the research portal.

MEASURES

Demographic

All those people who consented to participate were asked to provide basic demographic data on-line: gender, age, marital status, educational attainment, hours worked per week, type of job (manual/non-manual etc.), length of time in the job and the job content questionnaire developed by Robert Karasek.[19] The latter will be the focus of a separate analysis led by Paul Grime, to follow this report.

Service use

Telephone interviewers recorded the use of health and social care services by study participants with an adapted version of the Client Service Receipt Inventory (CSRI)[20]. At baseline, participants were asked for details on the use of services during the previous six weeks. At each of the five subsequent weekly telephone calls they were asked for details of services used since the last call.

These interviews also recorded sick leave, asking 'In the past six months/week have you had any days off due to ill-health?' and 'If so, how many of these days would you say were due to your mental health?'.

Other measures

All other measures were collected on-line, as shown in Table 2. Permissions were obtained to use standardised instruments, where applicable. The study team rated the importance of five pre-coded statements about the acceptability of cCBT (1-3, where 3 is 'very important). These

acceptability items are set out below, together with the open-ended statement which invited further comments.

1. I can use the computer at my own pace.
2. Using a computer is anonymous, I don't need to tell people about my problems.
3. It is convenient for me to access help via the internet and not to have to go to a health centre or clinic.
4. I can access help at any time that suits me.
5. The computer will not criticise me.
6. Any other reasons to like or dislike help via the internet (please give brief details).

Table 1 summarises the data collection instruments and number of questions posed at each time point.

Table 2: Data collection schema

Measure	**Week completed on-line**		
Self-completion instruments (no. questions)	**B'line**	**Wk 6**	**Wk 12**
Demographics and job-related questions (20)	X		
Work and social Adjustment scale (8)	X	X	X
Patient Health Questionnaire (9)	X	X	X
Work Limitations Questionnaire (25)	X	X	X
EQ-5D (5)	X	X	X
Acceptability questions (6)	X	X	X
Total est. number of questions	73	53	53

Acceptability analysis

Agreement with statements about acceptability and comparisons of on-line self-help with professional input were analysed using descriptive statistics. The percentage agreeing that each aspect was important or very important was compared at baseline and after six weeks. Change over time in individual ratings of the acceptability of cCBT was explored using paired t-tests, and differences between groups (intervention v control and positive v negative expectations) were compared using t-tests of independent means. Responses to open-ended questions were sorted and coded using a grounded approach; categories were not pre-set but evolved from the data, which were sorted and coded to summarise findings as concisely and clearly as possible. Where categories emerged, comments were allocated and counted. Comments were also coded as broadly negative, positive or neither, and the differences between study arms in numbers of comments so coded at baseline and follow-up were tested for significance using Chi-square tests.

Cost effectiveness analysis

Service use and lost employment were recorded for the six-month period prior to randomisation and then for the six weeks following randomisation. Services included contacts with primary and secondary healthcare professionals and inpatient stays. Lost employment was also recorded. Costs were calculated by combining the service use data with information on unit costs[21]. Lost employment was valued by combining lost work days with average earnings data.

Costs at follow-up with and without lost employment were compared between the two groups. Due to expected skewness in the cost distribution, a bootstrapped regression model was used, with baseline costs controlled-for.

Quality adjusted life years (QALYs) were computed from the EQ-5D using area under the curve methods. If MoodGYM results in lower costs and better outcomes then it is deemed 'dominant'. If it, or the control, had both higher costs and better outcomes then an incremental cost-effectiveness ratio was calculated, defined as the difference in costs divided by the difference in QALYs. This shows the extra cost incurred to produce an extra QALY.

STATISTICAL ANALYSIS

Sample profile

The demographics of the trial participants were compared where possible with the whole workforce for each of the organisations involved in the trial to assess the representativeness of the study sample from employers' perspectives. The demographics of participants who completed screening and chose to proceed were compared to those who completed screening and did not proceed. The sensitivity of our results to possible bias deriving from gender imbalance and drop out was explored.

Loss to follow-up and other missing data

Missing baseline items were imputed to create complete scores where no more than 20% of items were missing. Baseline demographics and outcome values were compared for those with complete sets of scores and those with missing scores and six and twelve weeks and are reported in Appendix 2.

Main analysis

The probability of falsely claiming a statistically significant result (type 1 error) increases where multiple significance tests are interpreted simultaneously. For this reason, the main research question was addressed by a single statistical model; no adjustments were made to the 5% significance level.

Data were analysed using Stata 11 according to the intention-to-treat (ITT) principle: i.e. patients are analysed in the treatment groups to which they were randomised irrespective of treatment received as long as outcome data are available. The significance level was 5% (2 sided) for specified analyses.

The main statistical analysis was to estimate the difference in mean outcome (WSAS) of scores between participants randomised to MoodGYM and control across the two follow-up points (six and twelve weeks).

A linear mixed effect model for longitudinal data (random intercept model) was used to estimate, using maximum likelihood, the difference between treatment arms in WSAS score at six and twelve weeks overall (taking account of any time trends). This approach allows the simultaneous modeling of the two time points at which we measured outcome, reflecting the estimated difference in randomised groups across the entire follow-up period. The WSAS total score at both follow-up points constitutes the dependent variable.

The assumption of normality was checked as follows. A Q-Q plot of level one standardised residuals against their normal scores and Q-Q plots of level two residuals (subject level, for intercepts) against their normal scores were produced. These plots verified that the residuals were approximately normal.

The pre-specified covariates that were included in the model consisted of the baseline outcome score, the randomisation group and the organisation. Time was included to estimate the effect over time (this allows for non-constant covariance between observations on the same person, because the covariance depends on the time gap between occasions). An interaction between time and intervention was included in a subsequent model to test for evidence of a differential effect over time

Variables that were associated with missing outcomes were investigated so that they could be included in models so as to minimize bias due to non-response. Analysis showed the following variables were associated with not completing follow-ups: age, organisation and baseline psychiatric scores (Appendix 2). However, organisation was already controlled for, as were each of the baseline psychiatric measures, so we only added in age.

RESULTS

RECRUITMENT AND REPRESENTATIVENESS OF RECRUITED PARTICIPANTS

Participant flow through the trial is summarised in a consort diagram in Appendix 1. There were 9,305 visits to the website, 1,111 completed the screening and 637 were randomised with 359 completing six-week and 231 participants completing twelve-week on-line assessments. The consort diagram shows how the participants with more severe mental health issues who were included in the analysis were distributed through the study stages.

Representativeness of sample by setting

We compared the basic demographics of the three participating workforces with those participants who entered the study from two of these organisations. No comparable information on the NHS workforce was gathered because a number of different workplaces were involved and we could not link the respondents to their employer in the NHS.

Table 3 shows that 78% of the TFL workforce is male, however the TFL sample that entered this study was largely female (54%) and this difference was statistically different. The TFL study sample mean age is younger than the TFL workforce as a whole; this difference was also statistically different.

Table 4 shows that the BT workforce is primarily male (78%) (Table 2), the study sample was also primarily male (if not as dominant) (58%), as in TFL, this relative bias towards female participants in the trial was statistically significant.

DEMOGRAPHIC COMPARISONS

Table 5 shows the baseline demographics according to intervention arm. There were no major differences; subjects had similar mean age, years in school and levels of alcohol consumption compared across both arms. Randomisation seemed relatively evenly split across each of the occupations. It is worth noting that more males were randomised to control than MoodGYM (50% versus 43%) and more females were randomised to MoodGYM than control (55% versus 48%). More single people were randomised to control than MoodGYM (26% versus 21%) and more of the married/cohabiting group were randomised to MoodGYM than control (69% to 61%).

Table 5b breaks down the occupational categories by the three employers. This reflects the different workforces, with more professionals in the NHS and more customer service personnel in BT. Within each employer the participating sample is not likely to reflect the whole workforce, because the study was rolled out by BT and TFL to the groups for whom it was most relevant and feasible (i.e. not actively promoted to staff groups who were not office-based). NHS Employers' membership is made up largely of professionals and managers, not a cross-section of NHS employees.

Clinical outcomes and missing data

The WSAS scale ranges from 0-40 with higher scores indicating more disability. The secondary outcomes were the PHQ-9 (n=636), the CORE10 (n=634) and the GAD (n=608). Scores on the PHQ-9 range from 0-27 with higher scores indicating more depressive symptoms. The CORE10 score can range from 0-40 with higher scores indicating individuals are reporting more problems and experiencing more distress. The GAD score can range from 0-21 with increasing scores indicating increasing anxiety.

At baseline there were varying numbers for which we were able to calculate the psychiatric summaries due to incomplete questionnaires. At baseline one participant did not complete any of the measures, hence we report 637 people recruited to the study, but only 636 baseline assessments. We had data at one or more assessment points over the twelve weeks for 401 subjects (63%). Our primary outcome was the WSAS scale which was completed by 636/637 subjects at baseline, 359/637 (56%) at six weeks and 231/637 (36%) at twelve weeks.

Table 6 shows WSAS score by intervention arm and organisation. Baseline WSAS scores were similar across organisations. The biggest difference at six weeks was at the NHS where the mean score varied by one point between intervention arms. By twelve weeks the scores on control were more varied (15.2 at the NHS and 17.2 at TFL). On the MoodGYM arm there was no difference between the six and twelve week mean score at BT, but scores at both the NHS and TFL continued to fall. The biggest difference at twelve weeks was at TFL where the mean score on the control arm was 3.2 points higher than the mean score on MoodGYM, indicating greater impairment in the control participants.

Retrospective examination of the data showed that those participants who were strictly ineligible according to the secondary exclusion criteria had significantly higher scores on the main outcome (WSAS) at all assessment points; baseline, six and twelve weeks. Twenty two of these 'difficult' cases were allocated to the intervention, and 19 to the control arm. We investigated the sensitivity of our results to their inclusion and found that the statistical significance of our findings did not differ if they were excluded. Details are given in Appendix 5.

Primary Analysis

Table 7 shows the summary psychiatric outcome measures by treatment arm. Mean baseline scores were similar across treatment arms. Figure 1 displays the profile of the unadjusted means for the WSAS score (observed scores) and Figure 2 compares the observed mean profiles with the mean profiles adjusted for covariates from the model in Table 6.

Across follow-ups each of the psychiatric measure means in both treatment arms decreased. This is illustrated for our main outcome measure WSAS score by Figures 1 and 2.

In Model 1, the random effects model which combines data from six and twelve weeks to give a combined effect size (Table 7), there was no statistically significant difference in the average treatment effect on the WSAS score (primary outcome) over the two time points; and there was no evidence for an interaction between time point and intervention so no evidence of a differential effect over time. This justifies combining the effect over the two time points. This model adjusted for baseline WSAS score and the organisation that the subject worked for. For the WSAS score the estimated treatment effect across follow-ups was -0.47 (-1.84 to 0.90, p=0.5). Figure 2 shows that this fitted model (Model 1, Table 7) was quite close to the observed profile.

We also considered an interaction between intervention and organisation to investigate whether there was a difference in the intervention effect according to the organisation the subject worked for. However, there was no evidence to support such an effect.

There was no evidence of a statistically significant treatment effect for any of the secondary outcomes.

We went on to compare Model 1 from our primary analysis with two additional models. Model 2 accounts for a possible gender imbalance at baseline; and Model 3 controls for any variables that were associated with subjects missing follow-ups (Appendix 2). These analyses did not give us any reason to alter our main finding that there was no evidence of a greater treatment effect from MoodGYM.

Table 3 Comparisons for the Transport for London (TFL) workforce (n=100)

TFL		**Whole workforce**		**Study sample**		P-value
		N	**%**	**n**	**%**	
Gender*	Male	18149	78%	43	44%	<0.001**
	Female	5221	22%	54	56%	
	Total	23370		97		
		Mean	**SD**	**Mean**	**SD**	
Age		42.9	10.4	40.1	9.8	0.0063***

* Gender status for 3 participants was missing in the study sample.
** Testing whether the proportion of females in the sample differs from the proportion of females in the whole workforce.
*** Testing whether the sample mean differs from the whole workforce mean.

Table 4: Comparisons for the BT workforce (n=393)

BT		**Whole workforce**		**Study sample**		P-value
		N	**%**	**N**	**%**	
Gender*	Male	73337	78%	223	58%	<0.001**
	Female	20309	22%	162	42%	
	Total	93646		385		
		Mean	**SD**	**Mean**	**SD**	
Age**		43.6	-	43.4	8.8	0.6199***

* Gender status for 8 participants was missing in the study sample.
** Testing whether the proportion of females in the sample differs from the proportion of females in the whole workforce.
*** Testing whether the sample mean differs from the whole workforce mean.

Table 5: Characteristics of subjects according to study arm

		Control (n=319)		MoodGYM (n=318)	
Gender	Male	160	50%	136	43%
	Female	152	48%	176	55%
	Missing	7	2%	6	2%
Age (years) (Mean (SD))		42.7	9.6	42.2	9.6
Marital Status	Single	82	26%	67	21%
	Married/cohabiting	196	61%	218	69%
	Widowed/separated	34	11%	27	8%
	Missing	7	2%	6	2%
Organisation	BT	195	61%	198	62%
	TFL	51	16%	49	15%
	NHS	73	23%	71	22%
Occupation	Manager or senior official	92	29%	91	29%
	Professional	66	21%	63	20%
	Associate professional or technical	33	10%	32	10%
	Administrative and secretarial	34	11%	52	16%
	Skilled trade	14	4%	5	2%
	Personal services	0	0%	2	1%
	Sales and customer service	60	19%	61	19%
	Process, plant and machine operative	1	0.3%	1	0.3%
	Other	19	6%	10	3%
	Missing	0	0%	1	0.3%
Education post-16(yrs) (Median, range (IQR))		3	0-39 (5)	3	0-39 (5)
Alcohol consumption (units) (Median, range (IQR))		5	0-140 (14)	4	0-70 (11)

Note: *Figures are numbers (%) of subjects unless stated otherwise.*

Table 5a Occupation of subjects according to study arm and employer

Control (n=319)	**BT**		**NHS**		**TFL**	
	195		**73**		**51**	
Manager or senior official	60	31%	15	21%	17	33%
Professional	25	13%	28	38%	13	25%
Associate professional or technical	22	11%	7	10%	4	8%
Administrative and secretarial	9	5%	16	22%	9	18%
Skilled trade	10	5%	1	1%	3	6%
Sales and service	57	29%		0%	3	6%
Process, plant and machine operative	1	1%		0%		0%
Other	11	6%	6	8%	2	4%
MoodGYM (n=318)	**BT**		**NHS**		**TFL**	
	198		**71**		**49**	
Manager or senior official	60	30%	13	18%	18	37%
Professional	26	13%	24	34%	13	27%
Associate professional or technical	20	10%	8	11%	4	8%
Administrative and secretarial	21	11%	25	35%	6	12%
Skilled trade	5	3%	0	0%	0	0%
Sales and service	59	30%	1	1%	3	6%
Process, plant and machine operative	0	0%	0	0%	1	2%
Other	6	3%	0	0%	4	8%
Missing	1	1%	0	0%	0	0%

Table 6: Comparison of intervention group for WSAS score by organisation[1]

		Control			**MoodGYM**		
	Outcome	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**
BT	**WSAS**						
	Base	195	20.1	7.9	197	19.8	7.9
	6 weeks	108	16.2	8.5	97	15.5	9.0
	12 weeks	78	15.7	8.2	70	15.5	10.1
NHS	**WSAS**						
	Base	73	19.3	7.8	71	20.4	7.5
	6 weeks	46	17.0	8.8	40	16.0	8.4
	12 weeks	25	15.2	10.6	16	13.6	9.6
TFL	**WSAS**						
	Base	51	20.7	6.6	49	19.5	9.2
	6 weeks	34	17.0	8.5	34	17.6	10.4
	12 weeks	26	17.2	7.8	16	14.0	10.6

[1] The Work and Social Attitudes Scale rates responses (0 – 'not at all', through to 8 - 'very severely'. It goes as follows: 'People's problems sometimes affect their ability to do certain day-to-day tasks in their lives. To rate your problems look at each section and determine on the scale provided how much your problem impairs your ability to carry out the activity.' The five items are work, home management, social leisure activities, private leisure activities and family & relationships.

Table 7: Comparison of intervention groups for primary and secondary outcome measures

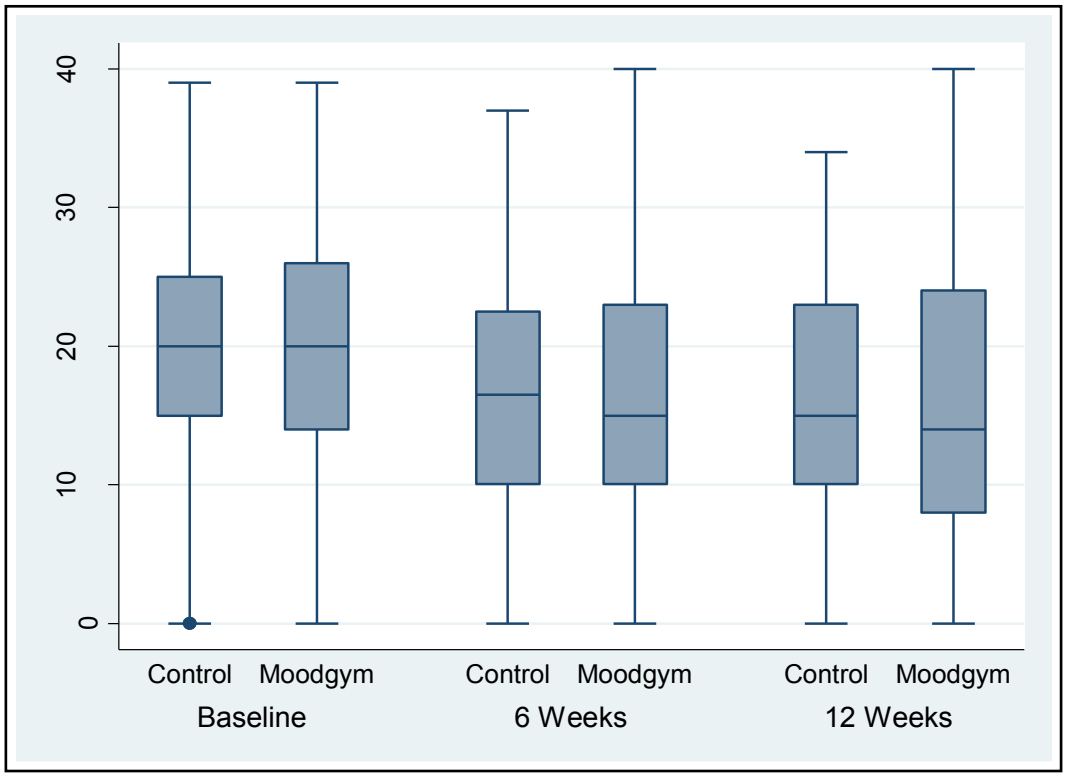
	Control			MoodGYM			Treatment effect
Outcome	**N**	**Mean†**	**SD**	**n**	**Mean†**	**SD**	(95% CI); P-value
Primary							
WSAS							
Base	319	20.0	7.7	317	19.9	8.0	-0.470 (-1.837, 0.897); 0.50
6 weeks	188	16.5	8.6	171	16.0	9.1	
12 weeks	129	15.9	8.6	102	15.0	10.1	
Secondary							
PHQ**							
Base	318	14.6	5.6	311	14.6	5.4	-0.429 (-1.454, 0.595);0.41
6 weeks	176	10.2	6.0	164	9.9	6.1	
12 weeks	122	10.3	6.9	97	9.3	6.9	
CORE10							
Base	318	18.3	5.3	316	18.4	5.9	-0.257 (-1.244, 0.731); 0.61
6 weeks	187	15.3	6.1	171	15.3	6.1	
12 weeks	129	15.0	6.9	101	14.1	7.3	
GAD							
Base	305	13.2	5.0	303	13.0	5.4	-0.341 (-1.334, 0.651); 0.50
6 weeks	181	10.2	5.7	166	9.5	6.0	
12 weeks	123	10.1	6.5	98	8.4	6.4	

* MoodGYM – Control, based on 6 &12 weeks combined; adjusted for time, baseline value of outcome and organisation.

† Refers to the estimated mean difference, across both time points.

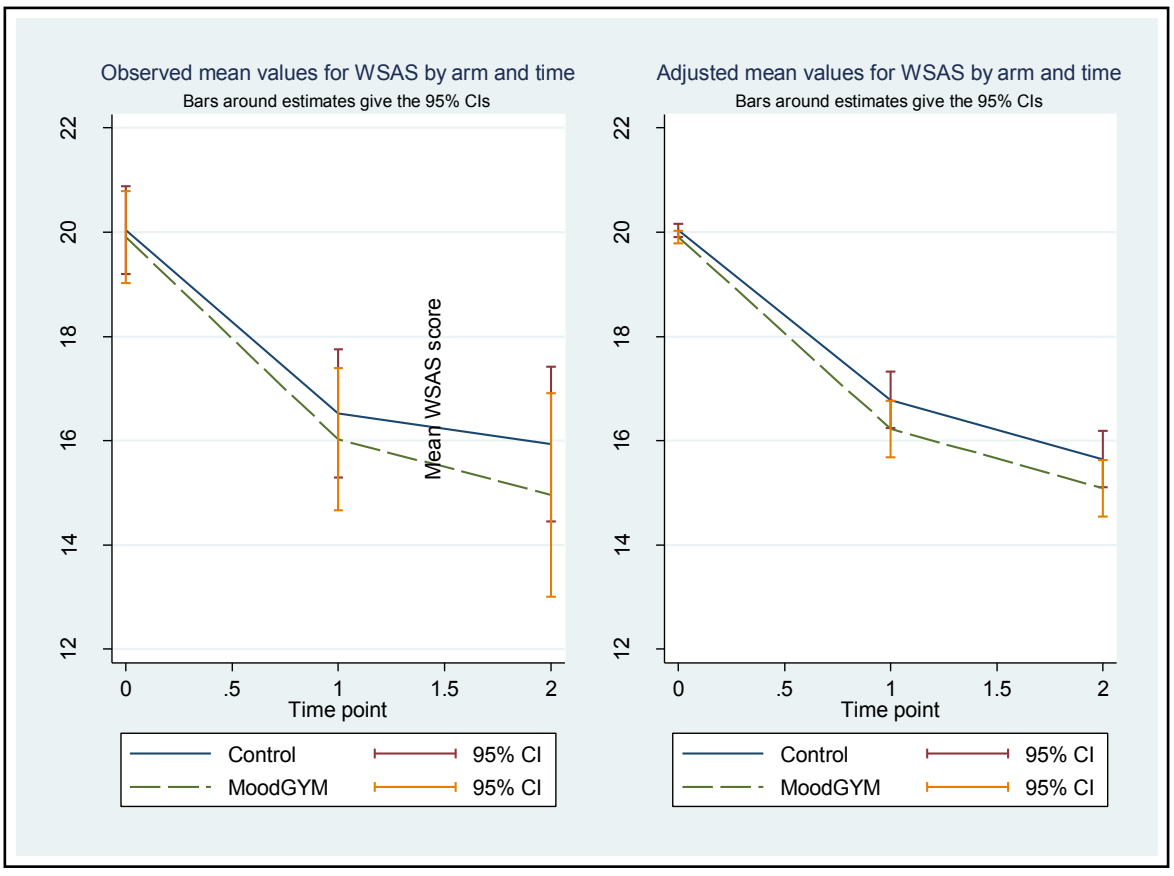
** Threshold for inclusion in the study was PHQ score greater than or equal to 10.

FIGURE 1 BOX PLOTS OF WSAS SCORES AT EACH TIME POINT FOR THE TWO TRIAL ARMS.



n=319 n=318 n=188 n=171 n=129 n=102

FIGURE 2 OBSERVED AND ADJUSTED MEAN PROFILES FOR WSAS SCORES BY TREATMENT ARM (BARS AROUND MEANS ARE 95% CI).



ECONOMIC ANALYSIS

Service use and lost employment were recorded by telephone interviews for the six-month period (26 weeks) prior to randomisation and then over the five weeks during participation in the study. Services included contacts with primary and secondary healthcare professionals and inpatient stays. Participants were asked about their use of counselling services and occupational health as well as other providers such as self-help groups or The Samaritans. Lost employment was also recorded.

Costs were calculated by combining the service use data with information on unit costs (Curtis, 2010). Lost employment was valued by combining lost work days with average earnings data.

Costs at follow-up with and without lost employment were compared between the two groups. Due to expected skewness in the cost distribution a bootstrapped regression model was used, with baseline costs controlled for.

QALYs were computed from the EQ-5D using area under the curve methods. If participants in the MoodGYM arm had lower costs and better outcomes then it is deemed to be 'dominant'. If it, or the control, had (both) higher costs and better outcomes then an incremental cost-effectiveness ratio was calculated, defined as the difference in costs divided by the difference in QALYs. This shows the extra cost incurred to produce an extra QALY.

RESULTS

There were no major differences between the two groups at baseline. In the six months prior to randomisation at least two-thirds of people had GP contacts and one in five attended hospital as an outpatient (Table 8). A similar proportion – about one-fifth of participants, used some kind of community-based provider. Most were in receipt of psychotropic medication; antidepressants and anxiolytics.

Around half the sample had taken time off work due to sickness, on average about 20 days over six months. Respondents stated that most of this time (52%) was taken off due to their mental health, which affected 28% of employees. People who were absent for this reason took longer on average to return to work – about 28 days (Table 8).

During the five-week follow up period, there were no major differences in service use between the groups (Table 9). About 16% of participants took some time off sick during this period – on average seven days. Most of this sick leave (9% of employees, 56% of those people off work) was attributed to mental health difficulties, and again the average length of absence was longer: 11 days (Table 9).

The costs and service use figures given in Tables 8 and 9 are not directly comparable because of different time-frames, but Figure 3 standardises costs per week of sickness absence at baseline and follow-up.

The cost of lost employment was higher for the control group at follow-up (£111 v £96, a difference of two days on average over all participants during the five-week follow-up period). However, this did not attain statistical significance (t-test of independent means p=0.759, 95% confidence interval -126 to 92). A similar pattern is seen in the total costs of absence from work, where control participants had higher costs (£143 v £119) but this was not significant either (p=0.644, 95% CI -137 to 84).

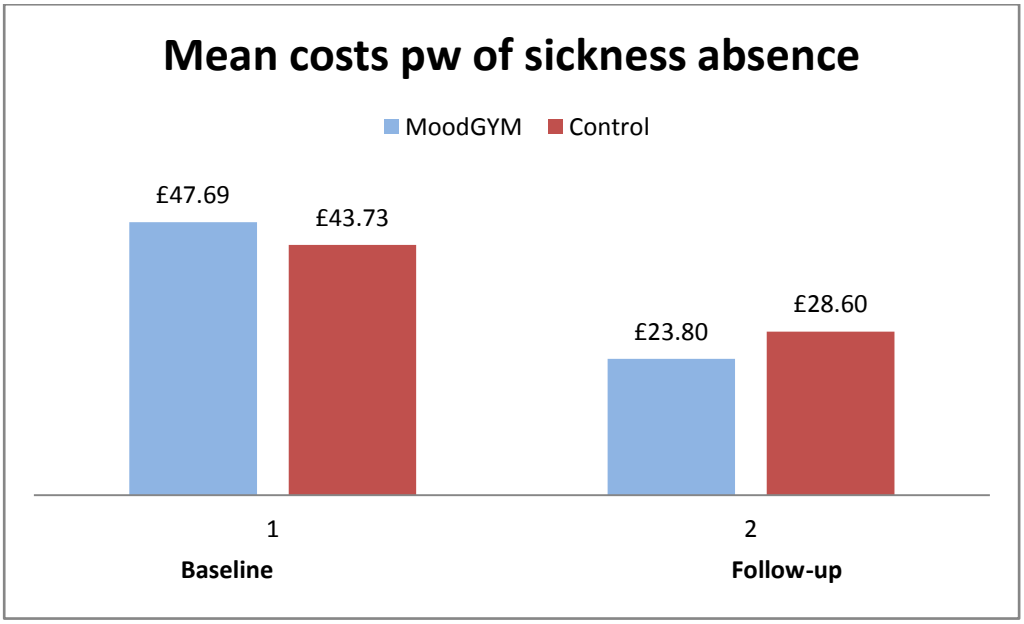
Table 8: Service use and costs (baseline)

	Intervention group 318			Control group 319		
Service	**N participants (%) using**	**Mean (SD) contacts (users only)**	**Mean (SD) cost (all participants)**	**N (%) using**	**Mean (SD) contacts (users only)**	**Mean (SD) cost (all participants)**
Hospital services						
Inpatient	20 (6%)	2.6 (3.3)	101 (854)	27 (8%)	3.4 (6.9)	345 (3733)
Outpatient	61 (19%)	4.5 (16.7)	90 (748)	73 (22.9%)	2.8 (4.7)	77 (315)
Community services						
GP	225 (71%)	3.6 (3.4)	98 (187)	213 (67%)	3.6 (2.8)	88 (130)
Psychiatrist	14 (4%)	3.7 (5.0)	42 (484)	4 (1%)	2.0 (1.4)	5 (52)
District nurse	44 (14%)	2.3 (2.9)	4 (17)	46 (14%)	1.6 (1.1)	3 (10)
Counsellor	47 (15%)	5.9 (6.4)	33 (123)	52 (16%)	5.6 (5.5)	36 (122)
Occ health providers	35 (11%)	1.4 (1)	3 (12)	44 (13.8)	2.1 (1.8)	6 (28)
Other providers	61 (19%)	4.2 (7.3)	39 (254)	73 (23%)	3.9 (4.8)	33 (143)
Medication	307 (97%)	1.4 (1.4)		311 (98%)	1.5 (1.5)	
Lost work						
Total absence	170 (54%)	21.1 (34.4)	1240 (2996)	168 (53%)	19.6 (30.3)	1137 (2643)
Due to illness	91 (29%)	29.3 (41)	924 (2729)	86 (27%)	27.5 (37.6)	818 (2526)
Total service cost	**318**		**411 (1330)**	**319**		**592 (3898)**
Total cost (including lost work)			**1651 (3531)**			**1730 (3531)**

Table 9: Service use and costs (follow-up)

	Intervention group 318			Control group 319		
Service	**N participants (%) using**	**Mean (SD) contacts (users only)**	**Mean (SD) cost (all participants)**	**N (%) participants using**	**Mean (SD) contacts (users only)**	**Mean (SD) cost (all participants)**
Hospital services						
Inpatient	2 (1%)	1.5 (0.7)	3 (46)	2 (1%)	1 (0)	1 (22)
Outpatient	22 (7%)	1.2 (0.4)	11 (73)	30 (9%)	1.7 (1.2)	25 (118)
Community services						
GP	74 (23%)	1.4 (0.7)	6 (17)	82 (26%)	1.6 (1.2)	7 (18)
Psychiatrist	5 (2%)	1.8 (1.3)	4 (38)	4 (1%)	1.3 (0.5)	1 (10)
District nurse	11 (4%)	1.1 (0.3)	0.12 (0.71)	19 (6%)	1.2 (0.6)	1 (3)
Counsellor	20 (6%)	1.6 (0.9)	2 (8)	23 (7%)	1.7 (0.8)	2 (9)
Occ Health provider	7 (2%)	1.1 (0.4)	0.5 (5)	18 (6%)	2.7 (3.3)	3 (15)
Other providers	21 (7%)	2.8 (4.3)	5 (44)	18 (6%)	2.7 (3.3)	3 (15)
Medication	190 (60%)			194 (61%)		
Lost work						
Total absence	55 (17%)	6.3 (7.8)	119 (440)	50 (16%)	8.3 (19.5)	143 (906)
Absence due to illness	34 (11%)	8.1 (8.8)	96 (417)	23 (7%)	14 (27.8)	111 (898)
Total service cost			**29 (110)**			**38 (125)**
Total cost (including lost work)			**125 (451)**			**149 (908)**

FIGURE 3 STANDARDISED COSTS AT BASELINE AND FOLLOW-UP



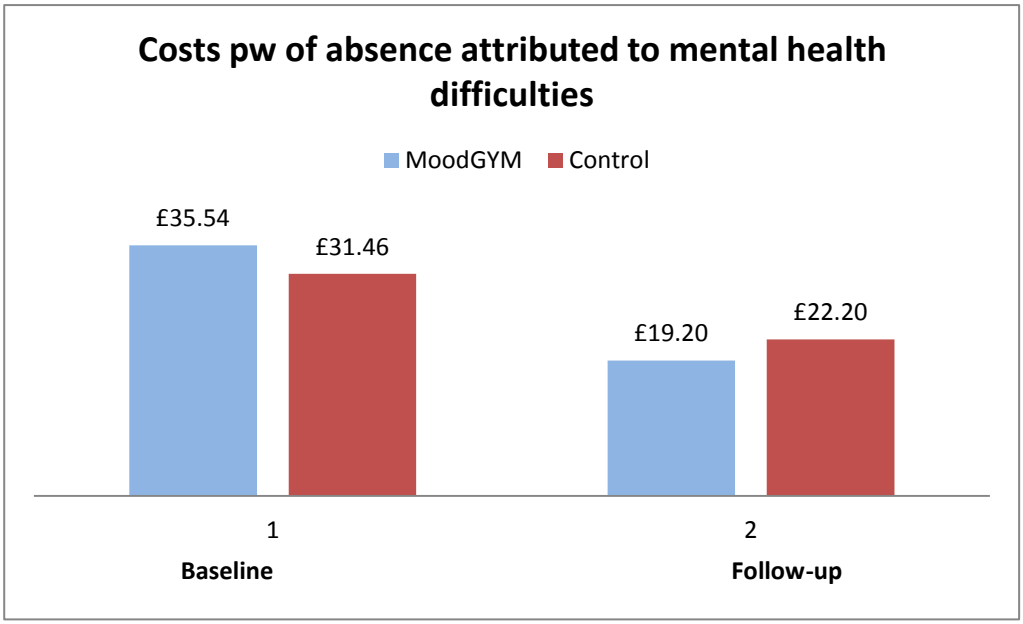


FIGURE 4 UTILITY VALUES BASED ON EQ-5D.

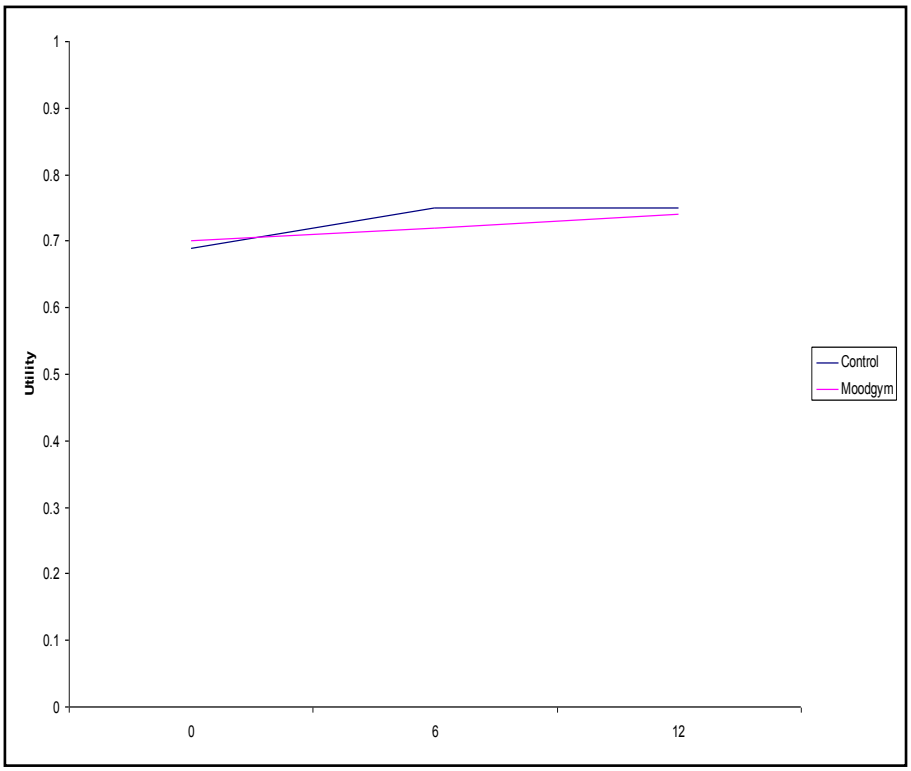
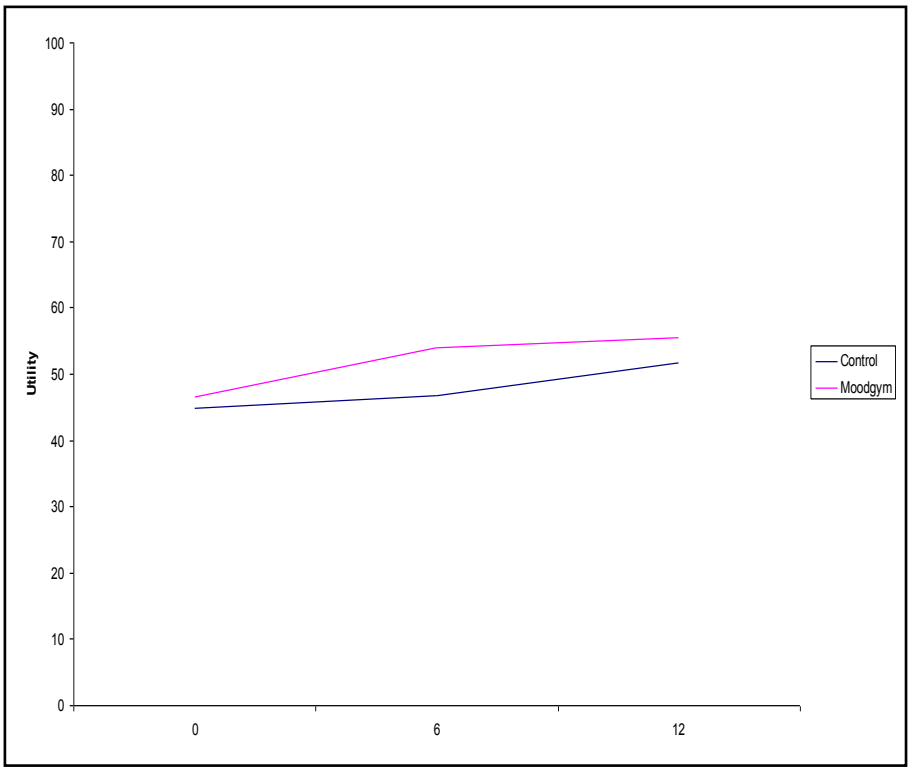


FIGURE 5 UTILITY VALUES BASED ON VAS



QUALITATIVE ANALYSIS

The acceptability items were completed by 554 people at baseline (84%), 276 people at six weeks and 215 at twelve weeks. Figure 5 illustrates the importance assigned by participants to the acceptability statements at baseline. The figure in brackets after each item below is the difference between six week and baseline results in the total percentage stating that this factor was 'very' or 'quite' important (6w-BL).

1. I can use the computer at my own pace. (2%)
2. Using a computer is anonymous, I don't need to tell people about my problems. (-1%)
3. It is convenient for me to access help via the internet and not to have to go to a health centre or clinic. (-3%)
4. I can access help at any time that suits me. (-6%)
5. The computer will not criticise me. (-4%)

There was strong agreement with all of these assertions at six and twelve week follow-ups. At six weeks, the importance assigned to accessing help 'at any time' was significantly lower (paired t test: t 3.396 p 0.001, 359 df). At twelve weeks, compared to baseline, when the intervention had been completed at least one month earlier, there was a statistically significant drop in the importance given to all but one statement: 'I can use the computer at my own pace', nevertheless a majority of respondents still regarded all five features as important.

In comparison to consulting a healthcare professional, the majority of respondents regarded self-help on-line to be equally or more acceptable (Figure 6) at both follow-ups. The reduction in acceptability over time shown in Figure 6 is due to growing dissatisfaction in the control group. The intervention group's ratings of self-help (N=170) increased between baseline and six weeks, although not significantly so (paired t-test), while the control group's ratings (N=190) decreased significantly in relation to three of the four alternatives: GP (t 2.472, p 0.014), counsellor (t 2.206, p 0.029) psychologist (t 1.527, p 0.129) and psychiatrist (t 2.267, p 0.025). The twelve week data showed that the difference between the arms was sustained over time.

FIGURE 6 AGREEMENTWITH REASONS FOR USING ON-LINE SELF-HELP AT BASELINE

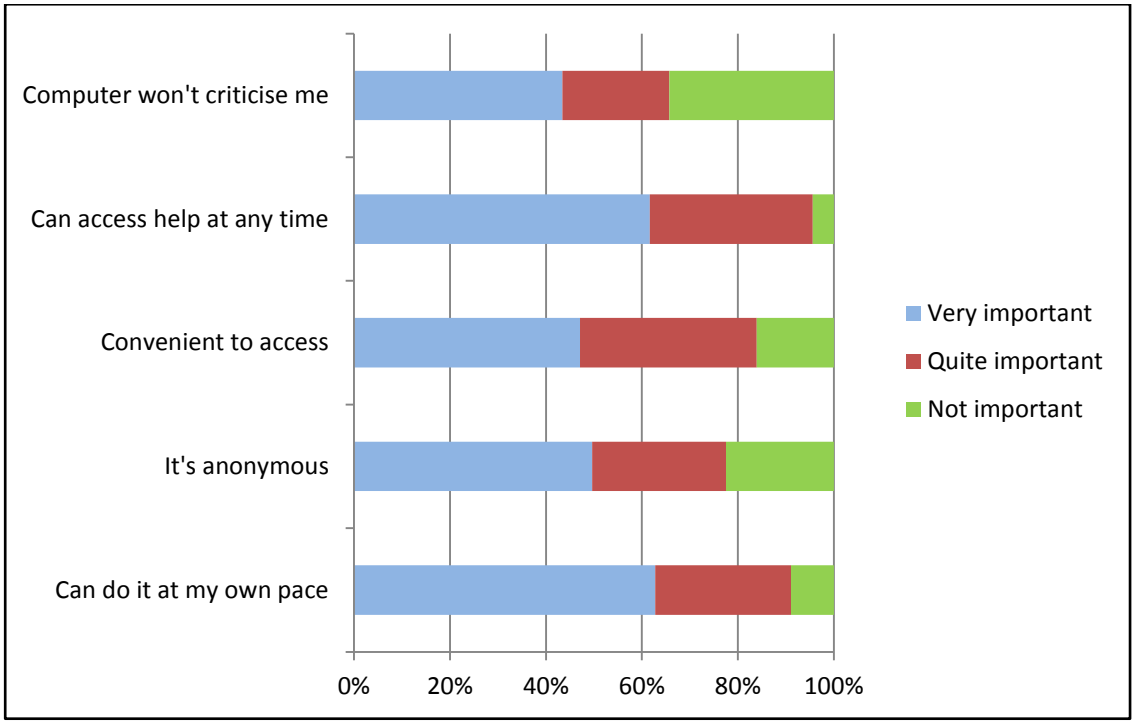
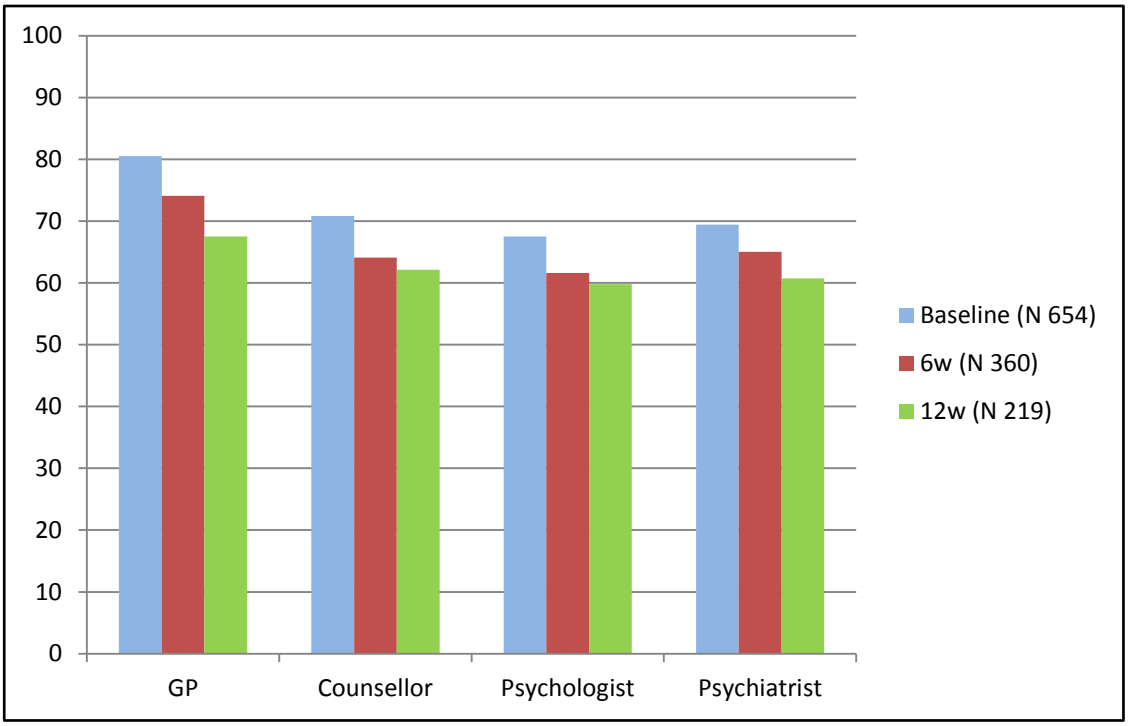


FIGURE 7 PERCENTAGE RATING ON-LINE SELF-HELP 'EQUAL' OR 'MORE ACCEPTABLE' BY COMPARISON WITH PROFESSIONALS



When we categorised the qualitative comments we also found differences between intervention and control arms; the latter judged the process to be less acceptable over time. Barriers to use fell into four categories: intrinsic, intra-personal problems; extrinsic technical problems; generic issues mostly pertaining to perceptions of cCBT; and specific issues about the intervention or control condition.

Participants were fairly evenly divided between those who disliked the impersonality of on-line self-help (N=93) and those who found it advantageous in some way (N=97). While some found the internet non-threatening, convenient and anonymous, a similar number felt that what they needed most was someone to talk to.

Typical advice from the control arm was that:

'It's a little TOO impersonal but it's an excellent complementary way to handle emotional/psychological issues and to provide self-help techniques. I don't think this can replace the more usual means of obtaining help but it certainly has its place'.

A MoodGYM user concluded that:

'help via the internet is far more accessible than having to wait on a GP referral to a counsellor. Some emotional problems can certainly be helped by following a CBT programme on the internet.'

A selection of representative comments from control and intervention groups is given in Box 1. The responses to the open-ended question about users' experiences of on-line CBT have been fully analysed in a separate paper.[22]

Box 1: Representative comments by study arm

CONTROL GROUP

Dislike psycho analysis by computer because it is impersonal and human conditions do not readily fit computer program algorithms so accurate or appropriate diagnosis is a hit and miss process.

I am dyslexic and in my case I find comprehension of sentences in reading difficult to take in quickly, and also find the background is not the best.

I am less likely to "attend" as people won't know if I miss it. It's good as I won't feel judged/as anxious about this. I worry it would be less adaptable to my particular needs than a person would be.

I can print things off and read them at my own pace - as well as ensuring that I understand what is going on. The only issue is still feeling very alone because the computer is not a person.

I feel that if you think you have a slight problem, then this is a good first step. If you feel like there is more treatment needed then other avenues can be looked at. This is a non-committed way of dealing with a problem.

In theory I like help via the internet but to be honest I have found this programme completely useless. It was not what I expected at all. All it did was provide me access to websites which are the sort of thing I can find easily on the internet anyway.

INTERVENTION GROUP

I have struggled to find the time/prioritise with a very high workload. I should have done more modules and prioritised this higher.

I believe the computer can only go so far with alleviating problems/symptoms. After a certain level one to one contact would be much more beneficial.

It's too generic, everyone is different.

Just the fact it's there when you need it, although does not suggest to get help if trend is getting worse. The questions didn't ask about any traumatic/stressful events during the week …

The internet is a useful way of accessing information and help but there will always be occasions when it could not replace talking to a person face to face.

There were a number of the personal logs that were unclear and needed to be discussed with someone before completion. This meant that I didn't necessarily get the best out of the programme.

DISCUSSION

MAIN EFFECTS

There was no evidence that the experimental intervention (MoodGYM) was more effective than the control intervention at reducing symptoms of depression and anxiety six weeks after participants joined the trial. Both groups improved in terms of their mental health. This could be caused by spontaneous remission, since we found no significant association between the amount of time people spent using the on-line intervention and the changes in their test scores (Appendix 4). Another explanation for the change is regression to the mean, which arises due to the combined effect of measurement error and the inclusion criterion: some people meet this because of random fluctuations and/or measurement error. Those people will improve on the rating scales whatever intervention (or none) is given.

A second possible explanation for the results is that participants on average did not receive a sufficiently strong 'dose' of the intervention for it to have the desired effect. We explored this possibility in Appendix 4, and while preliminary findings indicate that high users did not gain greater benefits, a more sophisticated statistical approach, Complier Average Causal Effect (CACE) should be applied to the study data to ascertain this.

Another explanatory factor is the weekly telephone calls to participants which both arms of the trial received; human contact provided through these calls might explain the improvement found across both groups. Yet qualitative evidence suggests that some participants preferred MoodGYM with its structured exercises to the control condition of informational websites about mental health. This is supported by discrete comments and by the increasing preference over time in the

control group for professional input. The control condition made up of five authoritative websites about mental health differed from the intervention in that participants were free to use the former as they wished. By contrast, MoodGYM took the intervention arm participants through a structured programme and asked them to actively engage in exercises and to complete quizzes.

There also remains the possibility that the control intervention (informational websites) was just as effective as the experimental condition (MoodGYM). Had we included a third arm, offering treatment as usual (TAU) it would have been possible to test this hypothesis. Other studies have done this, most recently the REEACT trial which is soon to report its findings. It used a three-way design: Beating the Blues with GP care v MoodGYM with GP care v GP care only (details can be found at: http://www.controlled-trials.com/ISRCTN91947481). In this study, GP care is equated with receipt of antidepressant medication. The comparison between the two CBT arms of REEACT and the present study should be illuminating.

Acceptability

The disproportionate number of women in the study may indicate that the approach is less acceptable to men.

Many participants found an on-line trial to be impersonal and there was a clear message from the qualitative data: one-third of the comments made indicated that people would have preferred to have contact with a sympathetic human being. The positive and negative views expressed by participants need to be interpreted cautiously, because the strongest critics of on-line CBT may well have declined to join the study. We counted roughly twice as many negative comments as positive ones expressed towards on-line self-help at the six week follow-up, with no difference between control and experimental arms. Participants in the control arm grew more dissatisfied by week twelve.

Nevertheless, many people were enthusiastic after engaging with the study, and some expressed high levels of satisfaction, saying that the approach suited them better than face to face counselling for a number of reasons, such as preserving their privacy and fitting in with their lifestyles. The main inference to be drawn from our acceptability data is that on-line self-help suits some people very well indeed. These findings endorse Wade's (2010)[23] recommendation that more knowledge is needed about the characteristics of people suited to Internet therapy. Similarly, more knowledge is needed about how cCBT programmes can be adapted and improved. Users' expectations of computer programmes are likely to be raised by the growing use of social networking and computer gaming. Participants in this study made numerous constructive comments that could guide future development.

Costs

Compared to the control intervention, MoodGYM was not associated with greater improvement in health-related quality of life over the follow-up period. Use and cost of services showed no difference between study arms, but the number of work days lost during the intervention period was significantly higher for the control group.

It is difficult to distinguish the immediate cause of any given period of sickness absence. In this study we simply asked the participants to tell us if their time off was due to their mental health, and indeed this was the attribution of most of the time off which was reported, both in the six months preceding the study and during the five week intervention period.

Findings suggest that MoodGYM was not cost-effective over a short follow-up period for this client group, when judged from a societal perspective. Yet MoodGYM participants who remained engaged with the study until the six-week follow-up had less time off work than the control group. Average sick leave absence of 2 additional days over five weeks for the control group extrapolates to an additional four weeks' absence per annum. Although not statistically significant, which means that this could be a chance finding, from an employer's perspective these costs may be taken as evidence to prefer MoodGYM over informational websites for employees who have common mental disorders.

METHODOLOGICAL CONSIDERATIONS

A notable aspect of this study is its approach to recruitment, screening and participation on-line, whose potential applications in clinical research have yet to be fully realised. This proved challenging and instructive. The experience gained in implementing an on-line trial has helped to build capacity for future studies and advances knowledge around the pros and cons of administering trials through the Internet. The advantages include not only the low cost relative to face to face research, but the ability to involve people from anywhere in the world, its consistent delivery of both study conditions, a certainty that consent is genuine and on-going, and minimal researcher bias, since the research team had no contact with the participants.

A key disadvantage seems to be that there are limited ways to maximise retention of participants in the study, since their engagement relies largely on their own initiative and commitment. Despite phone calls and email prompts to complete assessment tools, the retention rate from this study (56% at six weeks, 36% at twelve) was low by comparison with studies where there is face to face contact between the participant and the research team. While weekly contact with the phone interviewers was expected to encourage engagement, the number of calls they made to complete their data collection was limited to five calls at different times on different days which participants had indicated were suitable for them.

Our intention-to-treat design led us to include everyone who provided follow-up data at six weeks irrespective of how much of the intervention they used. Pittaway et al. (2009)[24], whose participants were referred by GPs but whose intervention took place outside the surgery settings, had a 50% completion rate of the treatment protocol. In a general practice-based study of Beating the Blues, Proudfoot et al. (2004)[11] similarly retained 74% of their sample between baseline and post-intervention assessment, and in a smaller sample (n=48) Grime (2004)[6] retained 90% at follow-up.

A more direct comparison of retention rates may be made with studies conducted entirely on-line. A community-based Internet trial of MoodGYM by Christensen et al. (2006)[25] found that only 30% of those people who complied with the baseline assessment completed even one module of the programme. However a study by the same team[9] (a three-way trial of MoodGYM, a psycho-educational site, BluePages, and control) had a completion rate of 81% of a highly self-selected sample.

We found that more people were lost to follow-up from the intervention arm than from the control, although this was not significant. Christensen, Griffiths and Jorm (2005)[9] also found that MoodGYM subjects were more likely to drop out than people using BluePages (25% v 15%). Such a trend could be due to the more demanding requirements of the interactive CBT aspects of MoodGYM with its exercises and quizzes to complete. There may be scope to make the user interface with MoodGYM more engaging, learning from developments in on-line marketing.

The present study's indirect approach to data collection, complicated further by remote working between members of the research team in Nottingham, London and Canberra, with study participants all over the UK, may have led to the inadvertent inclusion in the analysis of data from 41 people whom we had originally deemed to be ineligible because they had more complex mental health problems such as psychosis or bipolar disorder. After some deliberation we decided to retain them on the grounds that the study results would be more representative of the real world to do so; in fact their exclusion would have made no difference to the results. However, they did differ systematically from other participants, having more severe scores on the psychometric scales. This is important to note in comparing the present study with other studies of cCBT, which typically exclude people with severe mental health problems.

The choice of a linear mixed-effect model for longitudinal data to test our main hypothesis was made after considerable discussion, on the grounds that this permitted us to use the data from follow-up assessments at both time points and to take account of the impact of time on the participants' outcomes. We only have data twelve weeks after completing the intervention and the retention rate at this stage was low (36%). The short follow-up period is the most serious limitation of this study; we do not know whether the relative recovery rates for people in the MoodGYM and control arms continue in parallel over time or not.

CONCLUSION

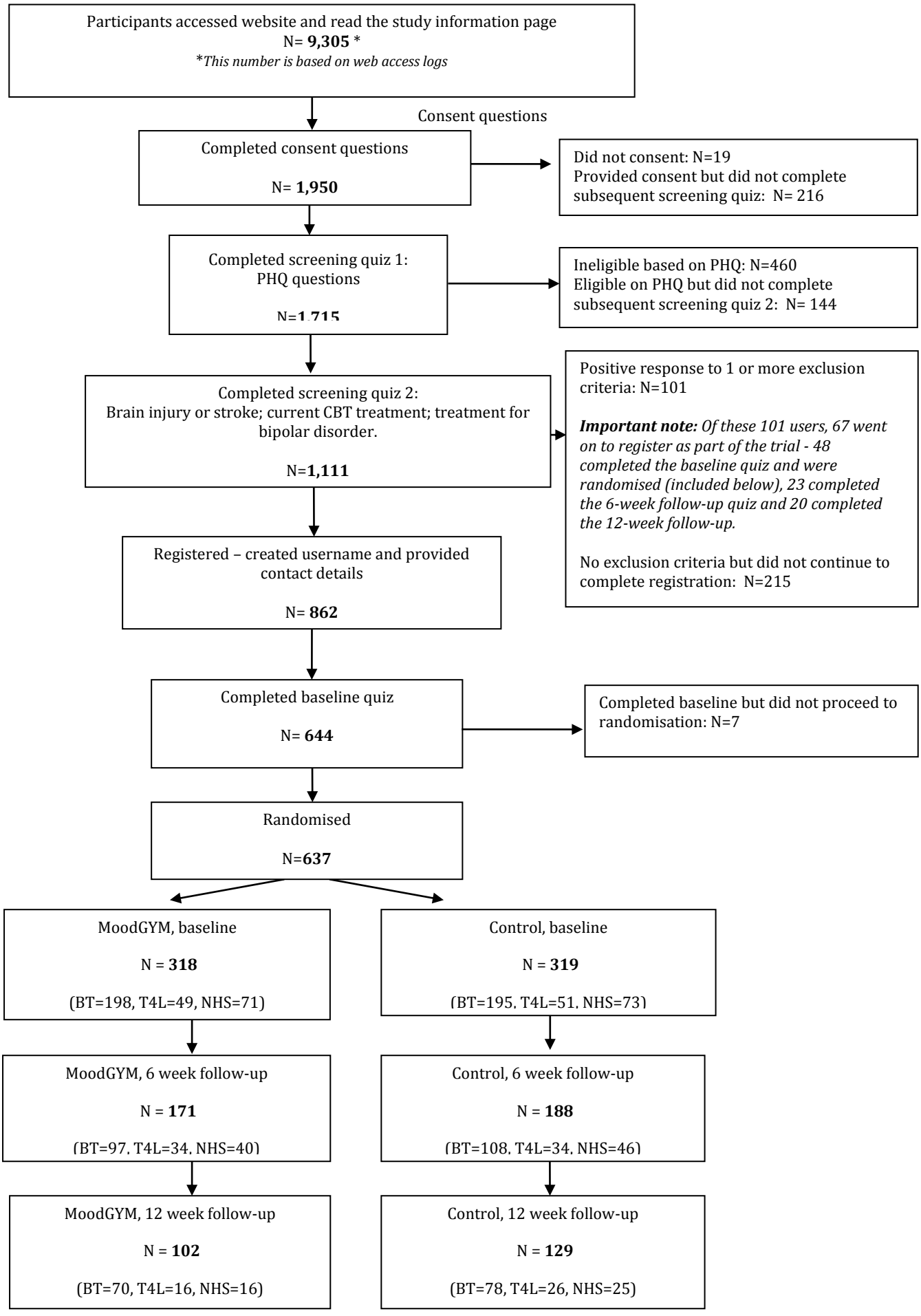
On average, participants in this trial improved their mental health. There were no adverse outcomes from the trial, and it adds to the evidence that on-line self-help is safe and widely acceptable.

There were no differences between the intervention and control participants in any of the psychological outcome measures applied, but the costs of lost employment differed significantly post-intervention, favouring the MoodGYM arm. Two considerations temper this finding: 46% of recruits to the study did not complete the post-intervention interview so it is possible that their outcomes, if known, could negate this result; and the results at six weeks may not be sustained over time.

The on-line approach to self-help for people in employment was popular with a substantial number of users. Although most would have preferred face-to-face contact, there was a minority who greatly preferred the on-line approach. The qualitative data gathered through this study are a rich resource for future development of interactive, on-line resources to improve mental health.

From an employer's perspective, the results of this study suggest that it is advisable to encourage employees who are experiencing mental health problems related to depression or anxiety to use on-line resources. In particular, MoodGYM may bring additional benefits of reducing the time they take off work. One obstacle to wider adoption of on-line self-help is that individuals vary in their attitudes towards it. It may be possible to increase uptake by presenting self-help, not as alternative to conventional counselling, but as a first step, and by promoting and communicating widely the advantages and benefits reported by participants.

APPENDIX 1: CONSORT DIAGRAM



APPENDIX 2: INVESTIGATING THE IMPACT OF GENDER, AGE AND MISSING VARIABLES ON OUTCOMES

Model one is described in the findings, Table 7 and Figure 2. As a result of several subjects missing baseline characteristics Models two and three are fitted on seven fewer subjects.

Model two: Gender itself was not significant in the model. In the random effects model that accounts for a possible gender imbalance in the analysis shown in Table 7, there was no statistically significant difference in the estimated average treatment effect on the WSAS score (primary outcome) over the two time points; and there was no evidence for an interaction between time point and intervention so no evidence of a differential effect over time (Table A2.1, Model 2). For the WSAS score the treatment effect across follow-ups was -0.58 (-1.97 to 0.81, P=0.41), which is a slightly bigger effect than the Model 1, but it is still not significant.

When we accounted for the gender imbalance there was still no evidence of a treatment effect for any of the secondary outcomes.

Model three: This included additional variables that were associated with missing follow-ups, specifically age, organisation and baseline psychiatric scores. Age was not significant in the model. Its addition did not alter our conclusions. In the random effects model that accounts for a possible gender imbalance and age being associated with missing follow-ups (Table A2.1 model 3) there was no statistically significant difference in the average treatment effect on the WSAS score over the two time points; and there was no evidence for an interaction between time point and intervention so no evidence of a differential effect over time. For the WSAS score the treatment effect across follow-ups was -0.57 (-1.96 to 0.82, P=0.42).

When we accounted for the gender imbalance and age being associated with missing follow-ups there was still no evidence of a treatment effect for any of the secondary outcomes.

Table A2.1 Primary and secondary outcome measures – model comparisons

Outcome	**Model One** (n=401) **Treatment effect * †** (95% CI); P-value	**Model Two** (n=394) **Treatment effect ** †** (95% CI); P-value	**Model Three** (n=394) **Treatment effect *** †** (95% CI); P-value
Primary			
WSAS	-0.470 (-1.837, 0.897); 0.50	-0.582 (-1.974, 0.810); 0.41	-0.571 (-1.961, 0.818); 0.42
Secondary			
PHQ	-0.429 (-1.454, 0.595); 0.41	-0.566 (-1.605, 0.472) 0.29	-0.559 (-1.587, 0.478): 0.29
CORE10	-0.257 (-1.244, 0.731): 0.61	-0.353 (-1.348, 0.642); 0.49	-0.355 (-1.350, 0.640); 0.48
GAD	-0.341 (-1.334, 0.651); 0.50	-0.450 (-1.444, 0.543); 0.38	-0.447 (-1.441, 0.547); 0.38

* MoodGYM – Control, based on 6 and 12 weeks combined; adjusted for time, baseline value of outcome measure and organisation.

** MoodGYM – Control, based on 6 and 12 weeks combined; adjusted for time, baseline value of outcome measure, organisation and gender.

*** MoodGYM – Control, based on 6 and 12 weeks combined; adjusted for time, baseline value of outcome measure, organisation, gender and age.

† Refers to the estimated mean difference, combined across the two time points.

APPENDIX 3: CORRELATION MATRICES OF PRIMARY AND SECONDARY OUTCOMES

Across the entire trial period

	WSAS	PHQ9	CORE10	GAD
WSAS total score	1.00			
PHQ9 total score	0.60	1.00		
CORE10 total score	0.58	0.67	1.00	
GAD total score	0.58	0.68	0.73	1.00

Correlation matrix of baseline scores

	WSAS	PHQ9	CORE10	GAD
WSAS	1.00			
PHQ9	0.46	1.00		
CORE10	0.47	0.55	1.00	
GAD	0.46	0.51	0.63	1.00

Correlation matrix of 6 week scores

	WSAS	PHQ9	CORE10	GAD
WSAS	1.00			
PHQ9	0.61	1.00		
CORE10	0.58	0.68	1.00	
GAD	0.57	0.72	0.76	1.00

Correlation matrix of 12 week scores				
	WSAS	PHQ9	CORE10	GAD
WSAS	1.00			
PHQ9	0.71	1.00		
CORE10	0.66	0.76	1.00	
GAD	0.67	0.79	0.77	1.00

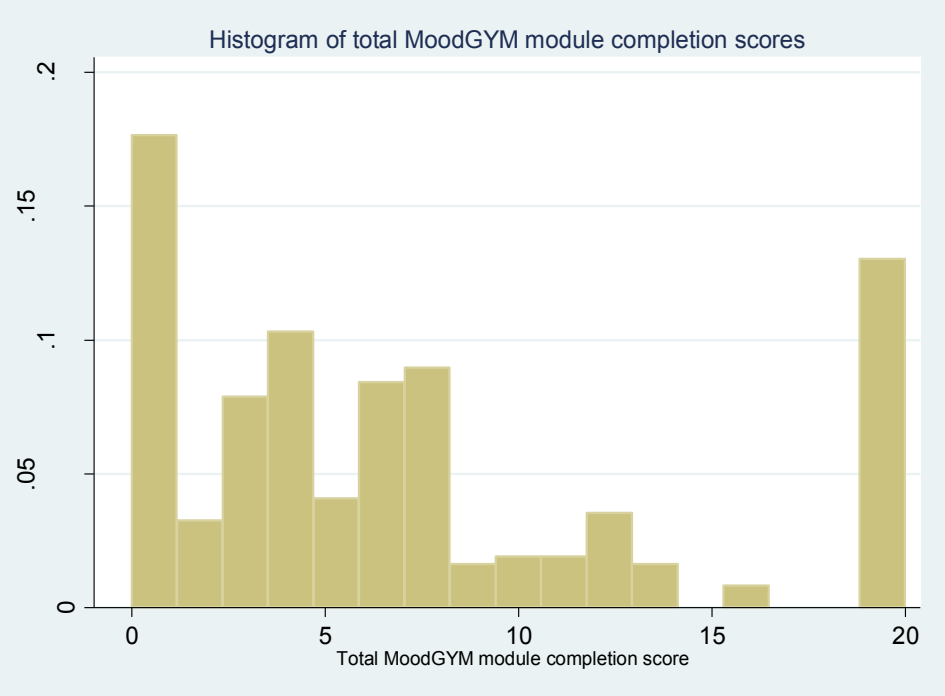
These show that all of our primary and secondary outcomes are correlated as we would expect. The increase in the correlation coefficients at twelve weeks is as a result of a decreasing sample size, which results in a reduction of variance amongst our scores.

APPENDIX 4: ASSESSMENT OF PROGRESSION THROUGH THE MOODGYM MODULES

Each participant on MoodGYM should complete 5 modules. Each participant is assigned a score from 0 to 4 (0= 0%, 1=25%, 2=50%, 3=75% and 4=100%) to indicate the level of completion for each of these modules. To describe the total amount of MoodGYM modules participants experienced, we totalled these scores for each of the five modules. Each participant on MoodGYM should thus have a score from 0-20, where 0 indicates none of the MoodGYM modules was completed to any level and 20 indicates all 5 modules were completed fully. Five participants had missing information on progression through the MoodGYM modules.

Scores to reflect progression through the MoodGYM modules can range from 0 to 20. The mean score for the 313 subjects randomised to MoodGYM who have information on progression is 7.2 (SD=6.6). The histogram below shows the distribution of these total completion scores.

Figure A4.1: Histogram of total module completion scores



In order to explore whether completion was related to outcomes from the study, the data were split into quartiles (Table A4.1), and then into tertiles (Table A4.2). Using tertiles allows us to have a measure that reflects a low level of completion, a moderate level of completion and a high level of completion. However, the quartiles also show that the scores could easily be dichotomised into less than half complete and more than half complete.

Table A4.1: Categorisation into quartiles of scores for level of completion

Categorised level of MoodGYM completion	N	%
0-3	106	34%
4-5	53	17%
6-10	77	25%
11-20	77	25%
	313	

Table A4.2: Categorisation into tertiles of scores for level of completion

Categorised level of MoodGYM completion	n	%
0-3	106	34%
4-8	117	37%
9-20	90	29%
	313	

We went on to look at the demographics of the MoodGYM subjects according to their level of MoodGYM completion, when this is categorised as low, moderate and high level of completion (Table A4.3). It appears that those that complete the highest level of MoodGYM modules are marginally older but this difference is not statistically significant.

There appears to be no significant difference in the median years of education or median units of alcohol consumed across the groups. There appears to be no significant trend in gender proportions across the levels of completion status. Married people have the highest proportion in the high completion category compared to single/widowed/divorced people but this trend across MoodGYM completion levels is not significant.

There appears to be a borderline significant trend in the organisation proportions across the ordered levels of MoodGYM completion. NHS workers have the highest proportion in the high completion category, compared to BT/TFL.

Table A4.3: **Characteristics of MoodGYM subjects according to their level of completion**

		Level of MoodGYM completion						P-value*
		Low (n=106) Score 0-3		Moderate (n=117) Score 4-8		High (n=90) Score 9-20		
Age (years)	Mean (SD)	42	9.1	42	9.7	43	10.2	0.46
Gender (missing 6)	Male	49	36%	46	34%	40	30%	0.727
	Female	55	32%	68	40%	49	28.5%	
Organisation	BT	70	36%	74	38%	50	26%	
	NHS	18	26%	25	36%	27	39%	0.071
	TFL	18	37%	18	36%	13	26.5%	
Marital status (6 missing)	Single	25	38%	24	36%	17	26%	0.516
	Married/ cohabiting	9	35%	8	31%	9	35%	
	Widowed/ separated	70	33%	82	38%	63	29%	
Education (years)	(Median, IQ range)	3	0-38 (4)	2.5	0-12 (5)	3	0-39 (6)	0.40
Alcohol use (units)	(Median IQ range)	5	0-50 (10)	4	0-50 (14)	4	0-70 (10)	0.80

Table A4.4: Outcome scores for MoodGYM subjects according to their level of completion of the MoodGYM modules by time point

	Level of MoodGYM completion						p-value *
WSAS score	Less than half complete Score 0-10			More than half complete Score 11-20			
Baseline	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	
	235	19.9	8.1	77	19.7	7.8	0.811
6 week	107	16.2	9.5	63	15.7	8.6	0.741
12 week	57	15.9	11	43	13.4	8.4	0.230

* t-test to test whether the means at each time point differ according to number of MoodGYM modules that a subject completes.

We next looked at low and high completion levels. Table A4.4 shows that at baseline mean WSAS scores are similar regardless of level of completion of MoodGYM modules (there is no significant difference). For both levels of completion (above half and below half) the mean WSAS score appears to fall considerably with time. For those people who complete over half of the MoodGYM modules the mean scores fall from 19.7 at baseline to 13.4 at the 12 week follow-up. Similarly those people who complete less than half the modules have mean WSAS scores that fall with time, this reduction is not as large as for those who complete over half, from 19.9 at baseline to 15.9 at the 12 week follow-up. Although suggestive of a possible dose effect, there is no statistical difference between the mean WSAS scores by level of completion at each follow-up point.

This simple analysis shows no statistically significant evidence that outcome depends on level of participation (number of MoodGYM modules). A further analysis adjusting for baseline value leads to a similar conclusion (Table A4.5). In this random effects model which combines data from six and twelve weeks to give a combined effect size of completion level there was no statistically significant difference on WSAS scores based on MoodGYM completion level (above or below half of the modules) across the two time points; and there was no evidence for an interaction between time point and completion level so no evidence of a differential effect of completion level over time.

However, further analysis would be required to test this thoroughly. Such an analysis could be approached using Complier Average Causal Effect (CACE)[2] modelling and would require some assumptions about the level of compliance in the control arm. It would focus on the effectiveness for individuals actually using MoodGYM, rather than the Intention to Treat analysis in the main body of the report, which assesses the effect seen by employers across all those offered the intervention.

Table A4.5: Comparisons of completion level for WSAS scores

	Model One (n=195) *	**Model Two (n=195) ***
	Level of completion effect	**Level of completion effect and time interaction**
Outcome		
WSAS score	-1.15 (-3.2, 0.91); 0.274	-1.84 (-4.8, 1.2); 0.233

[2] Dunn, G, Maracy,M., Dowrick,C., Ayuso-Mateos,J.L., Dalgard,O.S., Page,H., Lehtinen,V., Casey,P., Wilkinson,C., Vazquez-Barquero,J.L., Wilkinson,G., and the ODIN Group. (2003). Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. ***British Journal of Psychiatry***, 183, 323-331. eScholarID:1d5368 | DOI:10.1192/bjp.183.4.323

APPENDIX 5 COMPARISON OF THE PARTICIPANTS WITH MORE SEVERE MENTAL HEALTH PROBLEMS

The main analysis includes 41 people whom we had originally deemed to be ineligible because they had more complex mental health problems such as psychosis or bipolar disorder. They did differ systematically from other participants, with more severe scores on the psychometric scales.

Table A5.1: Simple comparison of outcome (WSAS) scores according to eligibility status

	Meet all inclusion criteria			Meet secondary exclusion criteria			
Outcome	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**P-value***
WSAS							
Base	609	19.69	7.91	41	22.49	7.23	0.03
6 weeks	340	16.04	8.75	19	20.79	9.41	0.02
12 weeks	214	14.86	8.98	17	23.59	9.19	<0.001

* t-test to test whether the means at each time point differ according to whether participant met the eligibility criteria or not

Table A5.1 shows that mean WSAS scores were significantly smaller at each of the time points for those participants who met all the inclusion criteria compared to those who were strictly ineligible according to the secondary exclusion criteria.

Table A5.2: Simple comparison of outcome (WSAS) scores according to eligibility status by treatment arm

Control (n=19)

	Meet all inclusion criteria			Meet secondary exclusion criteria			
Outcome	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**P-value***
WSAS							
Base	300	19.89	7.79	19	22.42	5.24	0.06 **
6 weeks	179	16.32	8.62	9	20.67	7.57	0.14
12 weeks	122	15.68	8.45	7	20.29	10.78	0.17

MoodGYM (n=22)

	Meet all inclusion criteria			Meet secondary exclusion criteria			
Outcome	N	Mean	SD	N	Mean	SD	P-value*
WSAS							
Base	295	19.71	7.91	22	22.55	8.72	0.11
6 weeks	161	15.72	8.91	10	20.90	11.23	0.08
12 weeks	91	13.77	9.58	10	25.90	7.64	<0.001

* t-test to test whether the means at each time point differ according to whether participant met the eligibility criteria or not.

** t-test that allows for unequal variances

Table A5.2 shows that, for people on the control arm, baseline WSAS scores were significantly smaller for the participants who met all of the eligibility criteria, as compared to those who should have been excluded based on the secondary exclusion criteria. The difference for this comparison was not statistically significant for each of the follow-up points.

However, for the MoodGYM arm the opposite pattern was seen. At baseline WSAS scores were not significantly different for participants who met all of the eligibility criteria compared to those who should have been excluded based on the secondary exclusion criteria. However, at each of the follow-up points the mean WSAS scores were significantly smaller for people who met all of the eligibility criteria compared to those who should have been excluded based on the secondary exclusion criteria.

We went on to explore the impact of these differences on the final analysis (Table A5.3). Model one shows us the people who met the eligibility criteria had a significantly smaller mean WSAS score across follow-ups, controlling for their baseline score, when compared to participants who did not meet the eligibility criteria. The mean WSAS score was approximately 5 points smaller across follow-ups for those who met the strict eligibility criteria compared to those who did not.

Model two (Table A5.3) shows that the people who met the eligibility criteria had a significantly smaller mean WSAS score compared to the others who should have been excluded, controlling for the baseline WSAS score and also for the time of the follow-up. The mean WSAS score was approximately 5 points smaller for those who met the eligibility criteria across follow-ups, controlling for their baseline score and time of follow-up.

Model three shows that those people who met the eligibility criteria had a significantly smaller mean WSAS score compared to those that should have strictly been excluded, when controlling for the baseline WSAS score, the time of the follow-up and the treatment arm. The mean WSAS score was again approximately 5 points smaller for those that met the eligibility criteria.

It also shows that when we controlled for whether the participant should have been included or not, the baseline WSAS score and the time of the follow-up there was no significant difference in the treatment effect on WSAS scores (-0.76 (95% CI (-1.98, 0.45), p-value 0.216)). We looked at the effect of eligibility status on treatment effect more specifically below.

Table A5.3: Regression models to examine the effect of meeting the inclusion criteria compared to not on the outcome (WSAS) scores.

	Model One (n=588) *	**Model Two (n=588) ****	**Model Three (n=588) *****
	Effect of meeting eligibility criteria	**Effect of meeting eligibility criteria**	**Effect of meeting eligibility criterial**
Outcome			
WSAS score	-5.16 (-7.69, -2.63); <0.001	-5.26 (-7.79, -2.74); <0.001	-5.34 (-7.87, -2.81); <0.001

* Controlling for baseline WSAS score

** Controlling for baseline WSAS score and time (categorical – 6 weeks compared to 12 weeks)

*** Controlling for baseline WSAS score, time (categorical – 6 weeks compared to 12 weeks) and treatment arm

We ran the main analysis random effects model excluding these ineligible participants to assess whether our conclusions would differ. However, despite the differences identified here, when we excluded those who were strictly ineligible our conclusions did not alter. We still found that there was no statistically significant difference in the average treatment effect on the WSAS score over the two time points; and there was no evidence for an interaction between time point and intervention so no evidence of a differential effect over time.

Table A5. 4: Comparison of intervention groups for primary outcome measure: *Model comparisons*

Outcome	**Model One (n=588)** **Treatment effect * †** **(95% CI); P-value**	**Model Two (n=552)** **Treatment effect *†** **(95% CI); P-value**
WSAS	-0.470 (-1.837, 0.897); 0.50	-0.829 (-2.226, 0.568); 0.25

* MoodGYM – Control, based on 6 and 12 weeks combined; adjusted for time, baseline value of outcome measure and organisation.

† Refers to the estimated mean difference, combined across the two time points.

In Table A5.4, Model one is the model from our main analysis and Model two is this same model but excluding the participants that did not strictly meet the eligibility criteria. We found that once these ineligible participants were excluded there was a bigger difference in WSAS scores between MoodGYM and control arm (with those on MoodGYM having smaller outcome scores). However, again this difference was not statistically different.

APPENDIX 6: TABLE OF AUTHORS' CONTRIBUTIONS

	KB	AB	PS	LN	GT	ML	RP	PG	BG	PMc	PW	IM	RM	JS
1					X									X
2	X	X	x		X	x				X			X	X
3	X	X	x		x				x		x		X	X
4			x		x									X
5			x	X					X					X
6	X	X	x											X
7	X	X	X	X										X
8	X	X	X	X										
9	X	X		X			X					x		
10	X		x			X	X			X			X	
11			x					x						X
12										X		X		
13			x	X			X	x	x	x		X	x	X
14	x	x	X	X	X	x	X	x	x	x	x	x	X	X

KEY
1. Commissioning of study
2. Design of study
3. Implementation of study
4. Liaison with BOHRF
5. Liaison with employers
6. Portal development
7. Trial management
8. Data collection
9. Data management
10. Statistical analysis
11. Qualitative analysis
12. Economic analysis
13. Final report
14. General intellectual contribution

X – leading contribution, x supporting contribution

Authors

KB Kylie Bennett, Australia National University
AB Anthony Bennett, Australia National University
PS Pooria Sarrami Foroushani, formerly University of Nottingham, now University of New South Wales
LN Lesia Nel, Institute of Mental Health, Nottingham
GT Graham Thornicroft, Institute of Psychiatry, Kings College London
ML Morven Leese, Institute of Psychiatry, Kings College London
RP Rachel Phillips, Institute of Psychiatry, Kings College London
PG Paul Grime, University of Cambridge and Addenbrooke's Hospital
BG Bob Grove, Centre for Mental Health
PMc Paul McCrone, Institute of Psychiatry, Kings College London
PW Paul Walker, Institute of Psychiatry, Kings College London
IM Iris Molosankwe, Institute of Psychiatry, Kings College London
RM Richard Morris, University of Nottingham
JS Justine Schneider, University of Nottingham

REFERENCES

[1] Sanderson, K., Tilse, E., Nicholson, J. Oldenburg, B. and Graves, N. (2007). Which presenteeism measures are more sensitive to depression and anxiety? Journal of Affective Disorders 101, p66.

[2] Singleton N, Bumpstead R, O'Brien M, Lee A, Meltzer H (2001) Psychiatric morbidity among adults living in private households. ONS,

[3] Department of Health (2001). Treatment choice in psychological therapies and counselling. London: HMSO. See also Royal College of Psychiatrists' website on the topic < http://www.rcpsych.ac.uk/mentalhealthinformation/therapies/cognitivebehaviouraltherapy.aspx> accessed January 7, 2012.

[4] NICE (2006) Computerised cognitive behaviour therapy for depression and anxiety. Review of Technology Appraisal 51.

[5] Department of Health (2007) Improving Access to Psychological Therapies (IAPT) Programme: Computerised Cognitive Behavioural Therapy (cCBT) Implementation Guidance.UK Department of Health.

[6] Sarrami, P., Schneider, J. and Assareh, N. (2011) Meta-review of the effectiveness of Computerised CBT in treating depression. BMC Psychiatry. Available at: http://www.biomedcentral.com/1471-244X/11/131.

[7] Grime, PR (2004) Computerized cognitive behavioural therapy at work: a randomized controlled trial in employees with recent stress-related absenteeism, *Occupational Medicine* 54**:**353–359.

Mundt JC, Marks IM, Shear MK, Greist JH. (2002) The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry. 180, 461-4.

[8] http://MoodGYM.anu.edu.au accessed January 7, 2012.

[9] Christensen,H., Griffiths, K.M., Korten, A., et al (2004a) Comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a CBT website. Journal of Medical Internet Research, 6, 46.

[10] Christensen H, Griffiths KM, Jorm AF (2004) Delivering interventions for depression by using the internet: randomised controlled trial. BMJ. 328 (7434), 265.

[11] Proudfoot, J.,Goldberg, D. P., Mann, A., et al (2003) Computerised, interactive, multimedia cognitive behaviour therapy for anxiety and depression in general practice. Psychological Medicine, 33, 217-227.

[12] Proudfoot, J., Ryden, C., Everitt, B. et al. (2004) Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial British Journal of Psychiatry 185, 46-54.

[13] McCrone, P., Knapp,M., Proudfoot, J., et al (2004) Cost-effectiveness of computerised cognitive behavioural therapy for anxiety and depression in primary care: randomised controlled trial. British Journal of Psychiatry, 185.55-62.

[14] Grime, PR (2004) Computerized cognitive behavioural therapy at work: a randomized controlled trial in employees with recent stress-related absenteeism, *Occupational Medicine* 54,353–359.

[15] Mundt JC, Marks IM, Shear MK, Greist JH. (2002) The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry. 180, 461-4.

[16] Kroenke K, Spitzer RL. (2002) The PHQ-9: a new depression and diagnostic severity measure. Psychiatric Annals, 32, 509-521.

[17] Lerner, D., Adler, D.A., Chang, H., Berndt, E.R. et al. (2004a) The clinical and occupational correlates of work productivity loss among employed patients with depression. Journal of Occupational and Environmental Medicine, 46, S46-S55.

Lerner, D., Adler, D.A., Chang, H., Berndt, E.R. et al. (2004b) Unemployment, job retention and productivity loss among employees with depression. Psychiatric Services 55, 1372-1378.

[18] Schulz KF, Grimes DA (2002). "Generation of allocation sequences in randomised trials: chance, not choice". The Lancet 359 (9305): 515–9.

[19] Available at: http://www.workhealth.org/strain/jsquest.html. Accessed January 7, 2012.

[20] Beecham, J. and Knapp, M. (2001) Costing psychiatric interventions. In *Measuring Mental Health Needs (Second Edition)* G. Thornicroft (ed.) Royal College of Psychiatrists, London, 2001

[21] Curtis, L. (2010) Unit Costs of Health and Social Care. Canterbury: PSSRU.

[22] Schneider, J. ,Thornicroft, G. and Sarrami Foroushani, P. (in preparation) Acceptability of on-line self-help for stress: users' views of MoodGYM versus informational websites.

[23] Wade, A.G., (2010) Use of the internet to assist in the treatment of depression and anxiety: A systematic review. Primary Care Companion to the Journal of Clinical Psychiatry 12 (4) e1-e11.

[24] Pittaway, S., Cupitt, C., Palmer, D. et al. (2009) Comparative clinical feasibility study of three tools for delivery of cognitive behaviour therapy for mild to moderate depression and anxiety provided on a self-help basis. Mental Health in Family Medicine. 6, 145-154.

[25] Christensen, H., Griffiths, K.M., Mackinnon, A.J. and Brittliffe, K. (2006) Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. Psychological Medicine., 36, 1737-1746.