Behavioural and Cognitive Psychotherapy, 2018, 0, 1–15
doi:10.1017/S1352465818000498

Translating the Intention to Seek Treatment into Action: Does Symptom Monitoring Make a Difference? Results from a Randomized Controlled Trial

Q1

R. Shafran

UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK

A. Gyani*

School of Psychology and Clinical Language Sciences, University of Reading, Reading, Berkshire RG6 6AL, UK

J. Rostron

School of Systems Engineering, University of Reading, Reading, Berkshire RG6 6AY, UK

S. Allen

Department of Politics & International Relations, University of Reading, Reading, Berkshire RG6 6AA, UK

P. Myles-Hooton

School of Psychology and Clinical Language Sciences, University of Reading, Reading, Berkshire RG6 6AL, UK

H. Allcott-Watson

UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK

S. Rose

Berkshire Healthcare NHS Foundation Trust, 2nd, 3rd and 4th Floors Fitzwilliam House, Bracknell, Berkshire RG12 1BQ, UK

Correspondence to Alex Gyani, Behavioural Insights Team (Australia), Level 13, 167 Macquarie Street, Sydney, NSW 2000, Australia. E-mail: alex.gyani@bi.team

* Alex Gyani is now Principal Advisor (APAC ~~Research & Evaluation Lead~~) at the Behavioural Insights Team (Australia).

Background: Most people with common mental health problems do not seek evidence-based psychological interventions. **Aims:** The aim of this study was to investigate whether monitoring symptoms of depression and anxiety using an app increased treatment-seeking. **Method:** Three hundred and six people with significant levels of anxiety and depression, none of whom were currently receiving treatment, were randomly allocated to receive either (a) information about local psychological services only, (b) information plus regular symptom monitoring (every 6 days), or (c) information plus open symptom monitoring (monitoring when they felt like it). An app was used to provide information and monitor mood. **Results:** The proportion of participants who reported receiving treatment after starting the study was 7.2% (10/138) in the information only group, 8.1% (9/111) in the information plus regular monitoring group and 15.8% (9/57) in the information plus open monitoring group. There was a trend for participants who were able to monitor whenever they wished to be more likely to report receiving treatment than people who were only given information about their local treatment services. The impact of the intervention was greatest among participants who intended to seek treatment before taking part. Limitations were that only a small minority of those who downloaded the app completed the study and that the study relied on self-reported measures of treatment-seeking. **Conclusions:** Symptom monitoring can increase actual treatment-seeking in those with an intention to seek treatment.

Keywords: treatment seeking, anxiety, depression, eMental Health, monitoring

Introduction

Epidemiological studies have found that one in six people in England suffer from anxiety and/or depressive disorders at any point in time (McManus et al., 2009) and these debilitating disorders are undertreated (Kessler and Ustun, 2008; Wang et al., 2005). The Adult Psychiatric Morbidity Survey in England indicated that only 10% of people with such common mental health problems receive psychological interventions (McManus et al., 2009). When psychological therapy is received, it is often after long delays, is frequently suboptimal and is not evidence-based (Shafran et al., 2009; Wang et al., 2005, 2007). This is despite the fact that people presenting with common mental health disorders prefer psychological treatments to medication (Kwan et al., 2010; McHugh et al., 2013) and prefer psychological treatments that have more empirical support than those that do not (Tarrier et al., 2006).

One reason why common mental health problems are undertreated is that people often do not seek help (McManus et al., 2009). A number of interventions to increase treatment receipt and encourage people to seek treatment for common mental health disorders have been designed. A systematic review (Gulliver et al., 2012) evaluated these interventions in relation to help-seeking intent and help-seeking behaviour. The interventions in the review included (1) giving people information to increase mental health literacy, (2) giving people information about where to seek treatment, (3) giving people destigmatizing information and (4) giving people computerized cognitive behavioural therapy (CBT) with personalized feedback about the individual's symptoms. The review concluded that these interventions had a positive impact on intention to seek treatment but little effect on actual treatment seeking, with the exception of the computerized CBT intervention (Christensen et al., 2006), which could arguably be considered treatment in and of itself.

Intention to seek treatment is a potential predictor of treatment receipt (Gulliver et al., 2010, 2012; Schomerus et al., 2009). It is therefore possible that those with high intention to seek treatment would benefit from a relatively minor intervention that, according to theoretical

models, tips the balance for them to translate intention into action (Ajzen, 1991; Prochaska and DiClemente, 1982).

One potential method to encourage patients to seek treatment is for them to monitor their symptoms of anxiety and/or depression, receive feedback about the severity of their symptoms and to obtain information on the availability of treatment. Such monitoring of symptoms is increasingly common on the internet with a number of freely available applications (NHS Choices Reporting Team, 2012).

Weekly symptom monitoring is effective in improving treatment outcomes when it is administered either before or after a treatment session (Kordy et al., 2001; Lambert et al., 2003; Lutz et al., 2006). A study by Drake and colleagues investigated the impact of completing an online symptom monitoring tool (called Moodscope) daily in people currently receiving treatment (Drake et al., 2012). The majority of participants found that the symptom monitoring tool helped them manage their symptoms. When interviewed, some participants suggested that symptom monitoring might benefit other people experiencing these disorders who were not yet willing to see a clinician. However, as of yet the impact of symptom monitoring on seeking or receiving treatment is not known.

The primary aim of this study was to evaluate whether monitoring symptoms of anxiety and/or depression increased reported actual receipt of treatment. An application was created for Apple iOS devices (iPhone and iPad), called 'Mood Mate'. A cell phone is an ideal medium for monitoring symptoms as it is easily accessible 24 hours a day, and ~~application allowed participants to~~ monitor their symptoms of anxiety and depression. As this study is the first of its kind, the optimal frequency of monitoring to encourage treatment receipt was unknown and it was decided to allocate people to monitoring symptoms every 6 days or open monitoring, i.e. whenever the participant chose to. Six-day monitoring was selected because psychological treatment sessions are often conducted weekly. As a given day of the week may be a confounding factor, it was decided that every 6 days would be optimal. Limiting the use of the tool to every 6 days contrasts with tools like Moodscope, which can be completed whenever the participant wishes and might be beneficial. To increase ecological validity, participants in the open monitoring group were able to monitor whenever they liked.

Participants were randomized into three groups that were given: (a) information about local psychological services only, (b) information plus regular symptom monitoring (every 6 days) and (c) information plus open symptom monitoring (monitoring when they chose to). After 30 days, participants were asked whether they had received treatment since starting the study.

It was predicted that among people with significant levels of anxiety and depression: (1) participants who monitored their symptoms would be more likely to report receiving treatment over a 30-day period than participants who only received information about local treatments, and (2) monitoring symptoms would have a significant impact on treatment receipt in those that already intended to seek treatment. No specific hypothesis was made regarding differences between the 6-day and open monitoring groups as this was the first study of its kind.

Method

Recruitment

The application (app) was made available on the iTunes App Store and was on the front page for the first week of its release. We used social and traditional media methods of encouraging

people to download the app. This included a Twitter Feed, radio interviews and a press release. Advertisements focused on the monitoring of mood.

Participants

To be eligible to participate in the study participants were required to (1) have an iOS device to run the Mood Mate application, (2) be 18 years or older, (3) be based in England, (4) be reasonably proficient to read and understand the language of instruction which was English, and (5) score above threshold on either the PHQ-4 (Patient Health Questionnaire-4), a four-item measure of depression, or the Mini SPIN (Mini Social Phobia Inventory), a three-item screen for social anxiety disorder. The thresholds were set at 3 on either the anxiety or depression components of the PHQ-4 and 3 on the Mini SPIN, indicating that they were likely to have an anxiety or mood disorder (Kroenke et al., 2009; Weeks et al., 2007).

Participants who met the inclusion criteria on the screening and stated that they were already receiving treatment at the time of the study were included in the study but excluded from the analysis ($n = 1952$, 22.5%). The mean age of the 306 participants who remained in the analysis was 27.2 years old ($SD = 9.7$). The majority of participants were female ($n = 217$, 70.9%) and a large proportion of the sample had never received treatment ($n = 228$, 74.5%). The majority of participants were Caucasian ($n = 256$, 83.7%), 9.5% ($n = 29$) were Asian, 2.3% ($n = 7$) were Mixed Race, 2.3% ($n = 7$) were Black and 2.3% ($n = 7$) of participants stated that they were of an 'other' ethnicity, or did not wish to divulge their ethnicity. Sixty-two (20.3%) participants did not meet the threshold on the PHQ-4, but did score above the threshold on the Mini SPIN.

Procedure

Once participants downloaded the app they were shown information about the study and were screened with the Patient Health Questionnaire-4 (PHQ-4) which consists of the first two items of the PHQ-9 (referred to as the PHQ-2) and the first two items of the GAD-7 (see below). If potential participants scored below the threshold on the PHQ-4, which does not screen avoidance behaviours, they were asked to complete the Mini Social Phobia Inventory (Mini SPIN) to detect social anxiety. If participants scored below the threshold on both measures they were signposted to a helpline where they could receive support if they wanted it and they did not continue with the study.

Those participants who scored above the threshold on either screening measure were presented with a consent form. Participants who consented were then given the baseline questionnaire to establish their baseline characteristics. They were then automatically randomized by the application using a built-in random number generator into either the control group or one of two monitoring groups. Participants in the control group received information about common mental health conditions and possible treatments. They were informed of the location and contact details of all Improving Access to Psychological Therapy (IAPT) services in England including their websites. These are psychological treatment services for anxiety and depression, the majority of which accept self-referrals. The location of the closest IAPT service to the participant could be found using the global positioning system (GPS) location of their iPhone.

Participants in the monitoring groups received the same information as the control group. Additionally they were asked to monitor their anxiety and depression symptoms by completing the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001) and the Generalized Anxiety Disorder Assessment (GAD-7; Spitzer et al., 2006). Those in the 6-day monitoring group were asked to monitor their symptoms every 6 days, while those in the open monitoring group were able to monitor their symptoms as often as they wished. Both groups were sent reminders to monitor their symptoms every 6 days to ensure that all participants received the same number of reminders.

Participants in the monitoring groups could view their scores on the PHQ-9 and GAD-7 measures graphically over time whenever they accessed the app. Participants were informed of the clinical thresholds for both measures, so they could understand where their symptoms placed them with respect tp mild, moderate, moderately severe and severe depression on the PHQ-9 and mild, moderate and severe anxiety on the GAD-7.

All participants were asked to not delete the app before completing the post-intervention measure, which was administered after 30 days of app use. After completing the post-intervention measure, participants in the control group were offered the opportunity to use the application's symptom monitoring tools that had been accessible to participants in the experimental groups. All communication with participants was via the app only. In nine cases participants emailed the authors to state that the contact details of IAPT sites were not correct. These were corrected within 48 hours.

Measures and materials

Mood Mate. The app was developed by J.R., A.G. and S.A. and was free for participants to download onto iPhones. This app was purely developed as a research tool and had no commercial interest. An earlier prototype of the application was shown to focus groups consisting of service users in the local area, to ensure that it met the needs of potential participants. The app was available to download on the UK iTunes store from 10 October 2012 until 9 April 2013 and was promoted via blogs, radio and Twitter.

*The Patient Health Questionnaire-4 (*PHQ-4; Löwe *et al., 2010).* The PHQ-4 consists of the first two items of the PHQ-9 (referred to as the PHQ-2) and the first two items of the GAD-7 (referred to as the GAD-2). It is a screening tool for low mood and anxiety, rather than a validated diagnostic tool. The PHQ-4 has been studied in the general population and in Primary Care, with high scores on the measure being found to be associated with functional impairment, disability days, and healthcare use (Kroenke et al., 2009; Löwe et al., 2010). The threshold used in this study was a score of 3 or above on the PHQ-2 and/or the GAD-2, so all participants were likely to have at least mild anxiety and/or depression. Scores range from 0 to 12 and are categorized as normal (0–2), mild (3–5), moderate (6–8) and severe (9–12).

The Mini Social Phobia Inventory (Mini SPIN; *Connor et al., 2001*). The Mini SPIN is a reliable and valid three-item instrument for screening social anxiety disorder in adults (Seeley-Wait et al., 2009; Weeks et al., 2007). To be included in the study, participants had to score 3 or above, as a liberal screening tool for social phobia. Scores range from 0 to 12 and are not categorized into severity groups.

The Patient Health Questionnaire-9 (PHQ-9; *Kroenke et al., 2001*). The PHQ-9 is a self-administered nine-item questionnaire developed to measure low mood. The PHQ-9 has a sensitivity of 88% and a specificity of 88% for major depression (Kroenke et al., 2001). Scores range from 0 to 27 and are categorized into: healthy (0–5), mild (5–9), moderate (10–14), moderately severe (15–19) and severe (20–27).

The Generalized Anxiety Disorder Assessment (GAD-7; *Spitzer et al., 2006*). The GAD-7 is a self-administered seven-item questionnaire developed to screen for GAD. It has also been shown as a valid measure of anxious symptomatology across all anxiety disorders using a cut-off of 8 or greater (sensitivity 77%, specificity 82%) (Kroenke et al., 2007). Scores range from 0 to 21 and are categorized into: healthy (0–4), mild (5–9), moderate (10–14) and severe (15–21).

Baseline measure. Participants were asked eight questions which assessed: gender (binary choice), birth year, ethnicity (applying the 16 categories used by the Office of National Statistics), whether they had ever sought treatment for, and whether they had ever received treatment for, a mental health condition (binary choices). Current treatment-seeking intention for a mental health issue, how effective they believed psychological therapies were in treating common mental health disorders and how effective they believed medication was in treating common mental health disorders were assessed using a 0–100 point analogue scale, administered with sliders.

Post-measure. The outcome measure was designed to determine whether or not a participant had sought and/or received any treatment for a mental health condition since starting the study (binary choice). If participants indicated that they had received mental health treatment since starting the study, they were asked what their diagnosis was (if they were given one), what treatment they received and from whom they received treatment. These three items were assessed using multiple-choice questions with the possibility to respond in an open text box. Participants were also asked to ~~re-asses~~ how effective they believed psychological therapies were in treating common mental health disorders and how effective they believed medication was in treating common mental health disorders, assessed with 0–100 point analogue scale, administered with sliders.

Data analysis plan

To assess whether there were any differences between participants who completed the outcome measures and those who did not, a logistic regression was used. Participants were compared on the basis of their PHQ-4 severity, gender, age, ethnicity, previous receipt of treatment, prior intention to seek treatment and belief in the effectiveness of psychological and pharmacological treatments.

To assess whether there were any differences between the three groups, the characteristics of the participants were compared using Kruskal–Wallis tests for continuous variables (PHQ-4 severity, age, prior intention to seek treatment, belief in the effectiveness of psychological and pharmacological treatments) and Chi-squared analyses for categorical variables (gender, ethnicity, previous receipt of treatment). As eight tests were performed, a Bonferroni correction was used to adjust p values.

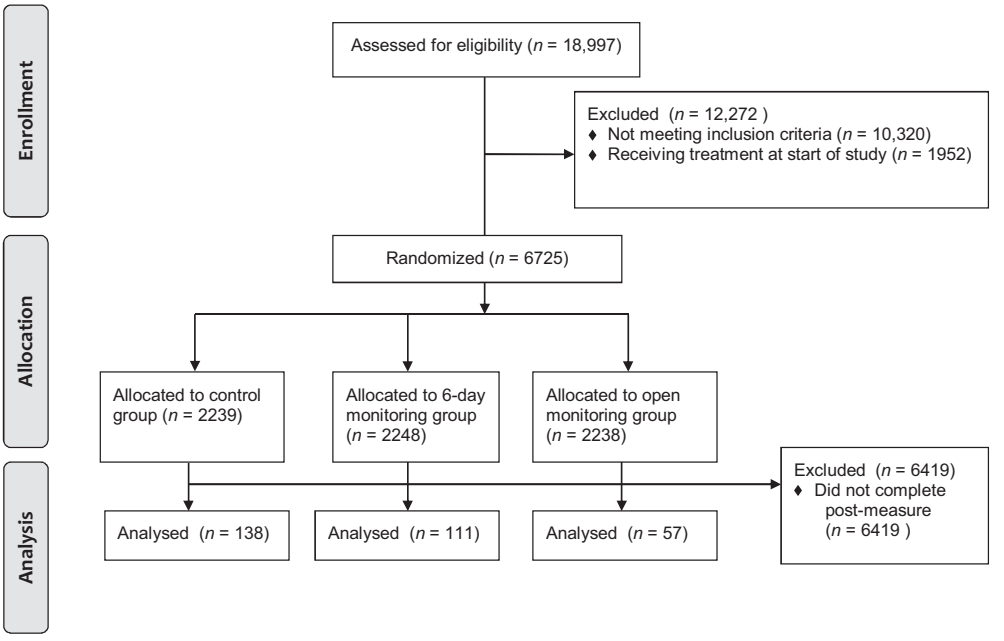

Figure 1. CONSORT diagram showing participant flow in the Mood Mate study

The primary outcome measure was reported receipt of treatment in the 30 days since starting the study. Two dummy variables coding participants' groups were used, with the information-only group acting as the reference group. Lemeshow and Hosmer's goodness of fit test was used to assess the fit of the models, and all models were shown to fit the data well ($p > .05$) (Lemeshow and Hosmer, 1982). Effect sizes are expressed as odds ratios (OR) and confidence intervals are reported. All analyses were conducted using SPSS v19.

Power analyses

Power analyses were conducted using G*Power 3.1 (Faul et al., 2007). A sample of 160 was needed to find a medium effect when comparing either of the monitoring groups with the control group, assuming equally balanced groups.

Results

Figure 1 shows the flow of participants through the study. Participants who were receiving treatment at the start of the study were excluded from the analysis which is why 8997 met the clinical threshold and were therefore given access to the tool, but only 6725 were included in the analysis. Furthermore, given the vast number of participants randomized but the small number of post-measures completed, it was decided that a completer analysis, rather than an intent-to-treat analysis, would be performed. Of the 8997 people that were randomized, 306 participants completed the post-measure.

Table 1. Baseline descriptive statistics of the participants in the three conditions and results of the comparisons between the groups

Variable	Control group ($n = 138$)		6-day monitoring group ($n = 111$)		Open monitoring group ($n = 57$)		χ^2 (d.f.)	Adjusted p value
PHQ-4 score (*SD*)	6.34	(2.68)	6.37	(2.76)	6.09	(2.85)	0.53 (2)	1
Age (*SD*)	26.57	(9.61)	27.85	(9.81)	27.23	(9.72)	1.01 (2)	1
Female *n* (%)	94	(68.12%)	84	(75.68%)	110	(68.42%)	1.92 (2)	1
Ethnicity *n* (%)							8.23 (8)	1
White	113	(81.88%)	98	(88.29%)	45	(78.95%)		
Mixed	3	(2.17%)	1	(0.90%)	3	(5.26%)		
Asian	16	(11.6%)	8	(7.21%)	5	(8.77%)		
Black	3	(2.17%)	3	(2.70%)	1	(1.75%)		
Other or unknown	3	(2.17%)	1	(0.90%)	3	(5.26%)		
Intention to seek treatment (*SD*)	29.07	(29.17)	34.52	(33.67)	30.70	(31.00)	0.91 (2)	1
Belief in the effectiveness of psychotherapy (*SD*)	57.58	(24.05)	54.24	(29.63)	49.86	(28.22)	2.61 (2)	1
Belief in the effectiveness of medication (*SD*)	44.38	(25.87)	41.68	(29.98)	38.81	(25.61)	1.79 (2)	1
Received treatment before *n* (%)	28	(20.29%)	38	(34.23%)	12	(22.05%)	7.02 (2)	.240

To investigate whether there was a difference between those who completed the post-measure and those who did not, logistic regression ~~was~~ performed. The model found that the baseline characteristics of participants described in Table 1 did not predict whether participants completed the outcome measure battery [$\chi^2(11) = 9.79$, $p = .550$]. The model explained 0.5% in variance between the participants who completed the outcome measure and those who did not. The descriptive statistics are shown in Table 2.

At baseline the mean score for all participants on the PHQ-4 was 6.30 ($SD = 2.74$); 6.21% were classed as normal on this measure ($n = 19$), 35.29% were mild ($n = 108$), 33.66% were moderate ($n = 103$) and 24.84% were severe ($n = 76$). The mean score on the Mini SPIN was 5.69 ($SD = 2.08$). Characteristics of the participants by group are shown in Table 1 along with the results of the Kruskal–Wallis tests and Chi-squared comparisons between the groups. These found no significant differences between the groups.

Effect of symptom monitoring on treatment receipt

The proportion of participants who reported receiving treatment after starting the study was 7.2% (10/138) in the control group, 8.1% (9/111) in the 6-day monitoring group and 15.8% (9/57) in the open monitoring group. Overall, 9.2% (28/306) of the participants started new treatment during the 30-day period.

Logistic regression found no significant difference between the proportion of people who reported receiving treatment in the control group and the 6-day monitoring group (OR = 1.13,

Table 2. Baseline characteristics of participants who completed the outcome measure and those that did not

Variable	Did not complete outcome measure ($n = 6419$)	Completed outcome measure ($n = 306$)
PHQ-4 score (*SD*)	6.43 (2.87)	6.30 (2.74)
Age (*SD*)	26.83 (9.26)	27.16 (9.69)
Female *n* (%)	4302 (67.0%)	217 (70.9%)
Ethnicity *n* (%)		
White	5344 (83.3%)	256 (83.7%)
Mixed	261 (4.1%)	7 (2.3%)
Asian	489 (7.6%)	29 (9.5%)
Black	159 (2.5%)	7 (2.3%)
Other or unknown	166 (2.6%)	7 (2.3%)
Intention to seek treatment (*SD*)	32.79 (31.96)	31.35 (31.20)
Belief in the effectiveness of psychotherapy (*SD*)	55.25 (28.27)	54.93 (27.03)
Belief in the effectiveness of medication (*SD*)	43.84 (28.19)	42.36 (27.38)
Received treatment before *n* (%)	1727 (26.9%)	78 (25.5%)

$p = .799$, 95% lower CI = 0.44, 95% upper CI = 2.88). However, there was a trend for people who were in the open monitoring group to be more likely to report receiving treatment than the control group (OR = 2.40, $p = .074$, 95% lower CI = 0.92, 95% upper CI = 6.27). Overall, this model was not found to be significantly better at predicting whether or not participants received treatment than a model that just contained the constant [$\chi^2(2) = 3.33$, $p = .189$] and only 2.4% of the variance in treatment receipt was explained, using Nagelkerke's R^2.

Post-hoc analysis on the role of intention to seek treatment

As the previous analysis did not find a significant difference between groups, *post-hoc* analyses were undertaken to investigate whether intention to seek treatment had a moderating role between symptom monitoring and treatment receipt. In order to investigate whether the impact of symptom monitoring was greater among those who had high intention to seek treatment, a median split was used to differentiate between participants who had low *versus* high intention to seek treatment at baseline.

Among those with low intention to seek treatment, people in the 6-day and open monitoring groups were no more likely to report treatment receipt than the control group (OR = 0.65, $p = .628$, 95% lower CI = 0.12, upper CI = 3.69; and OR = 1.15, $p = .876$, 95% lower CI = 0.20, upper CI = 6.62, respectively). Overall, this model was not found to be significantly better at predicting whether or not participants received treatment than a model that just contained the constant [$\chi^2(2) = 0.35$, $p = .839$]. Nagelkerke's R^2 indicated that only 0.7% of the variance in treatment receipt was explained.

Among participants who stated that their intention to seek treatment was above or equal to the median score, people in the 6-day monitoring group were no more likely to report treatment receipt than the control group (OR = 1.41, $p = .564$, 95% lower CI = 0.44, upper CI = 4.46).

However, the open monitoring group was significantly more likely to report receiving treatment than the control group (OR = 3.82, $p = .030$, 95% lower CI = 1.14, upper CI = 12.84). Overall, this model was not found to be much better at predicting whether or not participants were more likely to receive treatment than a model that just contains the constant, but a trend was found [$\chi^2(2) = 4.78$, $p = .092$]. Nagelkerke's R^2 indicated that 5.8% of the variance in treatment receipt was explained.

Type of treatment sought

Twenty-eight participants reported receiving treatment at the end of the study. For the majority of these participants treatment was received from their general practitioner (75.0%, $n = 21$), with one participant (3.6%) receiving treatment from an IAPT site and two (7.1%) receiving treatment from a private therapist. Four participants (14.2%) did not state from whom they received treatment.

Of the 28 participants who reported receiving treatment, 12 (42.9%) reported receiving a psychological treatment, nine (32.1%) reported receiving medication, two (7.1%) reported receiving both a psychological therapy and medication, and five (17.9%) reported an unclassified type of treatment. Of those who received psychological therapies eight received counselling (57.1%), five received CBT (35.7%) and one received psychodynamic therapy (7.1%).

For the 28 participants who reported receiving treatment, four were diagnosed with depression (14.2%), three with mixed anxiety and depression (10.7%), one with social phobia (3.6%) and one with post-traumatic stress disorder (3.6%). However, the majority either did not receive a diagnosis or stated they were not sure ($n = 15$, 53.5%). Two stated family loss (7.1%) and one preferred not to disclose the diagnosis (3.6%).

Discussion

This study investigated the effect of monitoring symptoms on treatment receipt. The proportion of participants who reported receiving treatment after starting the study was 7.2% (10/138) in the information group, 8.1% (9/111) in the 6-day monitoring group and 15.8% (9/57) in the open monitoring group. Half of the participants who reported receiving treatment received psychological treatments, but it was unclear whether these were evidence based. Among those that completed the final survey, participants who were able to monitor whenever they wished were more likely to report receiving treatment than people who were only given information about their local treatment services or were limited to using the app once every 6 days. The impact of the intervention was greatest among participants who intended to seek treatment before taking part in the study, consistent with the Transtheoretical Model of Change (Prochaska and DiClemente, 1982).

This study indicates that monitoring symptoms for people with a high intention to seek treatment might translate that intention into actual treatment-seeking behaviour, but if the effect exists, it is small. Previous research indicates that most people with common mental health disorders do not seek treatment, and those that do often wait a long time before doing so (McManus et al., 2009; Wang et al., 2007). Delaying treatment can lead to negative consequences for those with anxiety and depression (Brenes, 2007; Hawton and van Heeringen, 2009). Therefore, simple interventions to reduce the time it takes for people with anxiety and

depression to receive treatment are of potential value. The result of the current study ~~leads~~ some support to the notion that interventions may be more effective if targeted at an individual's current situation (Noar et al., 2007).

The majority of participants who received psychological treatments stated that they received counselling. Counselling is one of the NICE-recommended treatments for depression (NICE, 2011), but it is not recommended by NICE for any other mental health disorder. As the majority of participants did not report having received a diagnosis it is unclear whether the treatment they were given was evidence based. Further work should focus on whether an intervention can encourage people to receive treatments that are recommended by NICE, as opposed to another type of psychological treatment.

This study was not the first to trial an intervention to increase the number of people seeking or receiving treatment for anxiety and depression (Christensen et al., 2006). However, most interventions that have previously attempted to increase treatment receipt have only increased people's intention to seek treatment, rather than their behaviour (Gulliver et al., 2012). As the aim of these interventions has been to change the antecedents to behaviour, the effects on actual behaviour have frequently not been reported (Gulliver et al., 2012). The distinction between intention and behaviour is important for future behaviour change research. In this study, symptom monitoring was found to moderate the relationship between intention to seek treatment and treatment receipt. Rather than trying to change the antecedents of behaviour, interventions should also aim to increase the likelihood of intention being converted into action and future behaviour change research should consider the relationship between these two distinct constructs.

This study is subject to a number of limitations. Given the small number of people who completed the post-measure, it might be possible that this study's findings cannot be generalized to the wider population. Comparisons between the baseline characteristics of the groups indicate that there were no observed systematic differences between those who completed the post-measure and those who did not. It may also be the case that the effect of the intervention on treatment receipt may be a statistical artefact – the absolute numbers of people that reported receiving treatment were very similar. Analyses were undertaken to test whether or not participants who completed the outcome measure differed significantly on the measured variables by group. This was done to ascertain whether the differential drop-out rates had biased the analyses. There were no systematic differences between the groups on the measured variables, but this does not rule out the possibility of there being systematic biases between the groups on unmeasured variables. It is interesting that approximately twice as many participants completed the post-measure in the control and the 6-day monitoring group compared with the open monitoring group. It may be the case that continually having access to monitoring symptoms makes participants less inclined to complete post-measures. The high drop-out rate in this study also led to the analyses being less powerful than predicted, given the large initial sample size. High drop-out rates are common in eHealth trials – even up to 99.5% (see Eysenbach, 2005) – and thus the rate observed in this trial is not considered anomalous.

The design of the application may have impacted the length of time it was used by people. A number of authors have proposed certain design principles that have an impact on the length of time a user engages with an application (Fogg and Eckles, 2007). However, these design principles are rarely tested prospectively and robustly evaluated (Manzi, 2012). For this reason, further robust evaluation of the design of applications in relation to their effects on behaviour change should be undertaken.

This study did not collect information on suicidality in response to ethical concerns that if we were made aware of suicidality, we would be required to act. In order to act, we would have been required to identify individuals and monitor their responses. Given the nature of this study, it was not feasible to do so and we therefore omitted questions on suicidality for ethical reasons although such issues should be considered in future studies. Similarly it would be helpful for future studies to collect information on how often participants in the opening monitoring group monitored their symptoms and the monitoring behaviour of participants that did not complete the study. In addition, it is important to note that the timeframe of 30 days of monitoring was relatively arbitrary. It was chosen as it was considered that it would maximize engagement with the app and was sufficient to see fluctuations in mental health symptoms. However, it may have been too short to allow the full range of fluctuations to be captured and it was also inconsistent with the timeframe for the self-report measures of anxiety and depression which have a 2-week timeframe. Furthermore, the high drop-out rate would indicate a shorter period might be better for engagement and modifying the monitoring in some ways, for example personalizing messages and reminders, might also be beneficial for retention.

A further limitation is the study's dependency on the self-reported post-measures that are not diagnostic. As treatments were self-reported by the participants it is impossible to verify whether or not treatments actually took place. We specifically chose the wording of the question to reflect whether or not the participant reported receiving treatment instead of asking whether participants had sought treatment as we assumed there would be greater ambiguity about what behaviours can be considered as 'seeking' treatment. It is acknowledged that by wording the question in this way, participants who had sought treatment but were on a waiting list at time of study end would not be counted as having received treatment.

Future versions of this study could include the PHQ-9 and GAD-7 in the baseline and post-measure. This would allow us to establish whether symptom monitoring had an impact on participants' symptoms of anxiety or depression among participants who received treatment while using the application and those who did not. The work by Lambert and colleagues would suggest that participants who received psychotherapy and monitored their symptoms would have better outcomes than participants who received psychotherapy and do not monitor their symptoms (Kordy et al., 2001; Lambert, 2007, 2013; Lambert et al., 2003; Lutz et al., 2006).

The findings from this study indicate that symptom monitoring can encourage those who already have a high intention to seek treatment to actually seek it, provided that they are able to monitor their symptoms whenever they see fit. More generally, this study shows that cell phone applications can be used as a platform to run a randomized field trial, although high levels of drop-out are to be expected.

Acknowledgements

Q2 ~~Left blank for blind review.~~

Q3 *Ethical statements*: Ethical approval for this study was granted by ~~XXXXX~~.

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Conflicts of interest: The authors have no conflicts of interest with respect to this publication.

Q4 *Financial support*: ~~Left blank for blind review.~~

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211. doi: 10.1016/0749-5978(91)90020-T

Brenes, G. A. (2007). Anxiety, depression, and quality of life in primary care patients. *Primary Care Companion to the Journal of Clinical Psychiatry*, *9*, 437–443.

Christensen, H., Leach, L., Barney, L., Mackinnon, A. and Griffiths, K. (2006). The effect of web based depression interventions on self reported help seeking: randomised controlled trial [ISRCTN77824516]. *BMC Psychiatry*, *6*, 13. doi: 10.1186/1471-244X-6-13

Connor, K. M., Kobak, K. A., Churchill, L. E., Katzelnick, D. and Davidson, J. R. (2001). Mini-SPIN: a brief screening assessment for generalized social anxiety disorder. *Depression and Anxiety*, *14*, 137–140. doi: 10.1002/da.1055

Drake, G., Csipke, E. and Wykes, T. (2012). Assessing your mood online: acceptability and use of Moodscope. *Psychological Medicine*, *1*, 1–10. doi: 10.1017/S0033291712002280

Eysenbach, M. (2005). The law of attrition. *Journal of Medical Internet Research*, *7*. doi: 10.2196/jmir.7.1.e11

Faul, F., Erdfelder, E., Lang, A.-G. and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behaviour Research Methods and Instruments*, *39*, 175–191. doi: 10.3758/bf03193146

Fogg, B. and Eckles, D. (eds) (2007). *Mobile Persuasion: 20 Perspectives on the Future of Behavior Change*. Stanford: Stanford Captology Media.

Gulliver, A., Griffiths, K. M., Christensen, H. and Brewer, J. L. (2012). A systematic review of help-seeking interventions for depression, anxiety and general psychological distress. *BMC Psychiatry*,*12*, 81. doi: 10.1186/1471-244X-12-81

Gulliver, A., Griffiths, K. M. and Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatry*, *10*, 113. doi: 10.1186/1471-244x-10-113

Hawton, K. and van Heeringen, K. (2009). Suicide. *The Lancet*, *373*, 1372–1381.

Kessler, R. C. and Ustun, T. (2008). *The WHO World Mental Health Surveys*. Cambridge University Press: Cambridge.

Kordy, H., Hannöver, W. and Richard, M. (2001). Computer-assisted feedback-driven quality management for psychotherapy: the Stuttgart-Heidelberg Model. *Journal of Consulting and Clinical Psychology*, *69*, 173–183. doi: 10.1037//0022-006X.69.2.173

Kroenke, K., Spitzer, R. L. and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Spitzer, R. L., Williams, J. B. W. and Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics*, *50*, 613–621. doi: 10.1176/appi.psy.50.6.613

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O. and Löwe, B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*, 317–325. doi: 10.7326/0003-4819-146-5-200703060-00004

Kwan, B. M., Dimidjian, S. and Rizvi, S. L. (2010). Treatment preference, engagement, and clinical improvement in pharmacotherapy versus psychotherapy for depression. *Behaviour Research and Therapy*, *48*, 799–804. doi:10.1016/j.brat.2010.04.003

Lambert, M. J. (2007). Presidential address: what we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, *17*, 1–14. doi: 10.1080/10503300601032506

Lambert, M. J. (2013). Outcome in psychotherapy: the past and important advances. *Psychotherapy*, *50*, 42–51. doi: 10.1037/a0030682.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L. and Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice*, *10*, 288–301. doi: 10.1093/clipsy.bpg025

Lemeshow, S. and Hosmer, D. W. (1982). A review of goodness of fit statistics for use in the development of Logistic regression models. *American Journal Epidemiology*, *115*, 92–106. doi: 10.1093/oxfordjournals.aje.a113284

Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K. et al. (2010). A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders*, *122*, 86–95. doi: 10.1016/j.jad.2009.06.019

Lutz, W., Lambert, M. J., Harmon, S. C., Tschitsaz, A., Schürch, E. and Stulz, N. (2006). The probability of treatment success, failure and duration – what can be learned from empirical data to support decision making in clinical practice? *Science*, *232*, 223–232. doi: 10.1002/cpp.496

Manzi, J. (2012). *Uncontrolled: The Surprising Payoff of Trial-and-Error for Business, Politics, and Society*. New York, NY: Basic Books.

McHugh, R. K., Whitton, S. W., Peckham, A. D., Welge, J. A. and Otto, M. W. (2013). Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. *Journal of Clinical Psychiatry*, *74*, 595–602. doi: 10.4088/JCP.12r07757

McManus, S., Meltzer, H., Brugha, T., Bebbington, P. and Jenkins, R. (eds). (2009). Adult psychiatric morbidity in England, 2007: results of a household survey. Leeds: The NHS Information Centre for Health and Social Care.

National Institute for Health and Care Excellence (NICE) (2011). Common mental health disorders: identification and pathways to care (CG123). Retrieved from: http://www.nice.org.uk/nicemedia/live/13476/54604/54604.pdf (accessed 19 October 2013).

Noar, S. M., Benac, C. N. and Harris, M. S. (2007). Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. *Psychological Bulletin*, *133*, 673–693. doi: 10.1037/0033-2909.133.4.673

NHS Choices Reporting Team (2012). Mental Health Content Report November 2012. Retrieved from: http://www.nhs.uk/aboutNHSChoices/professionals/developments/Documents/Traffic%20reports/NHS%20Choices%20Mental%20Health%20Traffic%20-%20Nov%202012.pdf (accessed 26 July 2013).

Prochaska, J. O. and DiClemente, C. C. (1982). Transtheoretical therapy: toward a more integrative model of change. *Psychotherapy: Theory, Research and Practice*, *19*, 276–288. doi: 10.1037/h0088437

Schomerus, G., Matschinger, H. and Angermeyer, M. (2009). Attitudes that determine willingness to seek psychiatric help for depression: a representative population survey applying the Theory of Planned Behaviour. *Psychological Medicine*, *39*, 1855–1865. doi: 10.1017/S0033291709005832

Seeley-Wait, E., Abbott, M. J. and Rapee, R. M. (2009). Psychometric properties of the mini-social phobia inventory. *Primary Care Companion to the Journal of Clinical Psychiatry*, *11*, 231–236. doi: 10.4088/PCC.07m00576

Shafran, R., Clark, D. M., Fairburn, C. G., Arntz, A., Barlow, D. H., Ehlers, A. et al. (2009). Mind the gap: improving the dissemination of CBT. *Behaviour Research and Therapy*, *47*, 902–909. doi: 10.1016/j.brat.2009.07.003

Spitzer, R. L., Kroenke, K., Williams, J. B. W. and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*, 1092–1097. doi: 10.1001/archinte.166.10.1092

Tarrier, N., Liversidge, T. and Gregg, L. (2006). The acceptability and preference for the psychological treatment of PTSD. *Behaviour Research and Rherapy*, *44*, 1643–1656. doi: 10.1016/j.brat.2005.11.012

Wang, P. S., Angermeyer, M., Borges, G., Bruffaerts, R., Chiu, W. A., De Girolamo, G. et al. (2007). Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*, *6*, 177–185.

Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B. and Kessler, R. C. (2005). Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 603–613. doi: 10.1001/archpsyc.62.6.629

Weeks, J. W., Spokas, M. E. and Heimberg, R. G. (2007). Psychometric evaluation of the mini-social phobia inventory (Mini-SPIN) in a treatment-seeking sample. *Depression and Anxiety*, *24*, 382–391. doi: 10.1002/da.20250