Is the Devil in the Detail? A Randomized Controlled Trial of Guided Internet-Based CBT for Perfectionism

Roz Shafran[1*], Tracey Wade[2], Sarah Egan[3], Radha Kothari[1], Hannah Allcott-Watson[1], Per Carlbring[4], Alexander Rozental[4] and Gerhard Andersson[5,6]

[1] University College London, England

[2] Flinders University, Australia

[3] Curtin University, Australia

[4] Stockholm University, Sweden

[5] Linkoping University, Sweden

[6] Karolinska Institute, Sweden

[*] Author for correspondence: Roz Shafran, Chair in Translational Psychology, UCL Institute of Child Health, 30 Guilford Street, London WC1N 1EH, United Kingdom. Email: r.shafran@ucl.ac.uk

Abstract

An internet guided self-help cognitive-behavioural treatment (ICBT) for perfectionism was recently found to be effective (see this issue). Such studies stand in need of replication. The aim of this study was to report the outcomes and predictors of change when the treatment is delivered in a UK setting. A total of 120 people (Mean=28.9 years; 79% female) were randomised to receive ICBT or wait-list control over 12 weeks (trial registration: NCT02756871). While there were strong similarities between the current study and its Swedish counterpart, there were also important differences in procedural details. There was a significant impact of the intervention on the primary outcome measure (Frost Multidimensional Perfectionism Scale, Concern over Mistakes subscale) and also on the Clinical Perfectionism Questionnaire (between group effect sizes *d* =0.98 (95% CI: 0.60-1.36) and *d* =1.04 (95% CI: 0.66-1.43) respectively using intent-to-treat analyses). Unlike the Swedish study, there was significant non-engagement and non-completion of modules with 71% of participants completing fewer than half the modules. The number of modules completed moderated the rate of change in clinical perfectionism over time. In conclusion, the study indicates the intervention is effective in a UK setting but highlighted the importance of procedural details to optimise retention.

(200 words)

Keywords: moderation, non-engagement, completion, internet-based, cognitive-behaviour therapy, perfectionism.

HIGHLIGHTS (SEPARATE FILE)

- ICBT produced large between group effect sizes
- There was a modest but non-significant impact on depression and anxiety
- Non-engagement and non-completion rates were higher than the Swedish study (Rozental et al., 2017)
- The number of modules completed moderated the rate of change in clinical perfectionism over time.
- Those completing more modules had lower clinical perfectionism scores at the end of the intervention.

Introduction

Perfectionism is elevated across, and increases risk for, a range of psychological disorders including eating disorders, anxiety disorders and depression (Egan, Wade & Shafran, 2011). It can also have a direct negative impact on interpersonal relationships, physical health and daily functioning (Flett & Hewitt, 2002). A growing body of evidence shows that cognitive behavioral therapy (CBT) reduces perfectionism and psychological disorders, with medium to large effect sizes (Lloyd, Schmidt, Khondoker & Tchanturia, 2015). However, such face-to-face interventions can be difficult to access and costly to provide, so internet-based interventions have been developed.

Internet-delivered treatments have been found to be effective for a range of disorders and the treatment format facilitates cross-cultural adaptions and empirical investigations in different countries with minor differences in delivery (Andersson, 2016; Andersson & Cuijpers, 2009; Zachariae, Lyby, Ritterband, & O'Toole, 2016). The Swedish study (Rozental et al., 2017) on the internet-based treatment of perfectionism is highly encouraging for several reasons. It not only demonstrates that a 'low-intensity', high-tech intervention can have a positive impact on perfectionism, but it also significantly reduced anxiety and depression. The retention and completion rates of the 8 modules were high with 93% of participants completing at least half of the modules and 81% completing 75% or more in eight weeks. Although the findings are positive, there is a need for replication of findings in social science in general, and psychological research in particular (Open Science Collaboration, 2015; Yong, 2012).

The primary aim of this study was the same as the Rozental et al. (2017) Swedish study, i.e., to evaluate the impact of the internet-based intervention on perfectionism. Its secondary aims were (1) to compare the studies in terms of methods, retention and outcome

to the Swedish study and (2) establish the predictors and moderators of change. In terms of sequencing of the studies, thisstudy began before the Swedish one and finished afterwards, so they were running in parallel for the majority of the study period.

Method

Ethics, trial registration and protocol

Ethical approval for this study was granted by the University College London (UCL) Research Ethics Committee (Project ID: 6222:001) and the current study was registered as a clinical trial on ClinicalTrials.gov (NCT02756871). The protocol has been previously described (Kothari, Egan, Wade, Andersson, & Shafran, 2016).

Participants

An *a priori* power calculation was conducted for longitudinal designs (Hedeker, Gibbons, & Waternaux, 1999), with a two-tailed alpha of 0.05, three assessment points, a pre-post correlation for the primary outcome measure of 0.61, and attrition rates of 50%. Both the pre-post correlation and expected attrition rate were based upon a similar RCT of a web-based intervention for perfectionism (Egan van Noort et al., 2014). A sample size of 40 enrolled participants per group, with 20 participants completing per group, would provide 80% power at two-sided $p < 0.05$ to detect a large effect size (Cohen´s d=0.80) difference between the control and intervention groups. This use of a large effect size was justified by the results of a previous RCT conducted by Egan and colleagues (Egan, van Noort et al., 2014).

One hundred and fifty six people responded to advertisements via University email lists, social media and recruitment websites such as https://www.callforparticipants.com/ for people with significant perfectionism that interfered with daily functioning. Potential

participants were directed to the study website (https://www.overcomingperfectionism.co.uk/) to find out more about the study, read the information sheet, and give consent for participation. They then completed a battery of questionnaires, the results from a subset of which are reported in this manuscript.

Participants were required to be 18 years or over, score one standard deviation above published norms on the Concern over Mistakes subscale of the Frost Multidimensional Perfectionism Scale (FMPS; Frost, Marten, Lahart, & Rosenblate, 1990), i.e. a score of ≥ 29 (Suddarth & Slaney, 2001), agree to be randomized and be fluent in English. Participants were excluded if they disclosed suicidal thoughts or intent at any point from screening onwards. Currently receiving treatment for a mental health disorder was not an exclusion criterion and participants were not required to stop their other treatment (largely medication) to be included in this trial. Of those who responded, 35 were excluded as they did not meet the inclusion criteria of ≥29 on the FMPS Concern over Mistakes subscale (Frost et al., 1990). One more person declined to be randomized so was excluded. A flowchart of the recruitment process and treatment period, including the number of participants lost to each occasion, can be found in Figure 1.

Insert Figure 1 here

Procedure

The remaining 120 people were automatically randomized by a third party, unconnected to the study, using an internet-based randomisation program (https://www.sealedenvelope.com/simple-randomiser/v1/lists) to receive either the intervention (N=62) or waitlist control (N=58). Participants were randomized immediately

after completing the screening without being asked to confirm if they wanted to participate, with the exception of the last 7 people to enter the trial who were specifically asked to confirm that they still wanted to take part after they had completed the screening. This study reports on a subset of self-report questionnaire measures that are the same as/similar to those collected by the Swedish study (Rozental et al., 2017). Measures were collected at baseline and 12 weeks after randomisation, and all measures were administered online.

There were some key procedural differences between this study and the Swedish counterpart (Rozental et al., 2017). Unlike the Swedish study, this UK study:

1) Did not have a preliminary telephone interview assessment or use a diagnostic measure of anxiety or depression,
2) Automatically randomised participants after completion of the baseline measures (with the exception of the last 7 participants) without explicitly asking participants to confirm if they still wished to participate,
3) Asked participants in the intervention to complete the Clinical Perfectionism Questionnaire (CPQ) on a weekly basis and provided up to 3 automatically generated email reminders if the measure was not completed,
4) Did not provide guidance on specific days; in the UK all guidance had to be 'signed off' by RK which meant guidance could sometimes be delayed for up to a week,
5) Did not start all the participants in one 'batch' but rather entered participants as soon as they had completed screening measures,
6) While the interventions were similar and based on the treatment protocol for clinical perfectionism developed by Egan, Wade, Shafran and Antony (2014), the Swedish study changed the order of the modules slightly, provided more explanation and included more behavioural interventions,

7) Used the Depression Anxiety Stress Scale (Lovibond & Lovibond, 1996) to assess depression and anxiety rather than the Patient Health Questionnaire-9 (Löwe, Kroenke, Herzog, & Gräfe, 2004) and Generalized Anxiety Disorder-7 (Dear et al., 2011),
8) Allowed participants 12 weeks to complete the intervention instead of 8 weeks,
9) Paid participants £10 upon completion of the 12 week measure. Participants were told they would receive this payment as part of the information provided, prior to consenting to participate.

Measures

Frost Multidimensional Perfectionism Scale (FMPS; Frost et al., 1990): This self-report measure consists of 35 items grouped into six subscales: Concern over Mistakes (e.g. "I should be upset if I make a mistake"), Doubts about Actions (e.g. "I usually have doubts about the simple everyday things I do"), Personal Standards ("I set higher goals than most people"), Parental Expectations ("My parents set very high standards for me"), Parental Criticism ("My parents never tried to understand my mistakes"), and Organisation ("I try to be an organised person"). Participants respond on a five point scale ranging from 1 = "strongly disagree" to 5 = "strongly agree". The measure has been found to be both reliable and valid for use with non-clinical and clinical populations (Frost et al., 1990; Hewitt & Flett, 1991; Hewitt, Flett, Turnbull-Donovan, & Mikail, 1991). Cronbach's Alphas for each subscale indicated adequate to good internal consistency: Concern over Mistakes (9 items; α = .74), Doubts about Actions (4 items; α = .72), Personal Standards (7 items; α = .72), Parental Expectations (5 items; α = .92), Parental Criticism (4 items; α = .84) and Organisation (6 items; α = .92). The full scale was found to be highly reliable (35 items; α = .84). The Concern over Mistakes and Personal Standards subscales are widely considered to

be the most clinically relevant. This measure was amended in our study (but not in the Swedish study (Rozental et al., 2017)) to reflect participants' experience over the past month allowing us to measure change.

Clinical Perfectionism Questionnaire (CPQ; Fairburn, Shafran & Cooper, 2003): This self-report measure consists of 12 items (e.g. "Have you pushed yourself really hard to meet your goals?" and "Have you raised your standards because you thought they were too easy?"). Participants respond on a four point scale ranging from 0 = "not at all" to 3 = "all the time". This measure of clinical perfectionism was created by Fairburn, Cooper, and Shafran (2003) at the University of Oxford, and has been found to have good reliability and validity in two community samples and an eating disorder sample (Eganet al., 2016), in addition to adequate internal consistency (α = .74; this study). This measure was amended to allow for perfectionism in the domain of eating, shape and weight and, for the weekly measure, to reflect participants' experience over the past week rather than month.

Depression, Anxiety and Stress Scales (DASS; Lovibond & Lovibond, 1996) – Short form: The short form of the DASS is a 21 item self-report measure of depression, anxiety, and stress (e.g. "I found it hard to wind down"), rated on a four point scale ranging from 0 = "Did not apply to me at all" to 3 = "Applied to me very much or most of the time." It has been shown to be reliable, with good internal consistency (α = .91; this study), and has been validated for use among clinical and community samples (e.g., Crawford & Henry, 2003). The subscales were combined to form a measure of negative affect.

Statistical analyses

The primary outcome measure for the study was the FMPS Concern over Mistakes subscale (Frost et al., 1990). This was chosen since the CPQ was being completed weekly by the intervention group but not the control group and therefore was unsuitable to be the primary outcome measure. The secondary outcomes were clinical perfectionism, the remainder of the FMPS subscales and negative affect (depression, anxiety and stress). In order to compare change in our two groups over time, data were analysed in an ANCOVA using T2 (post-intervention and primary endpoint at 12 weeks) as the outcome variable adjusted for observations at T1 (baseline) in order to estimate the between group (intervention, control) effect on perfectionism and negative affect. Both completer and intent-to-treat (ITT) analyses were conducted. Where effects sizes and 95% confidence intervals are presented, these were calculated using the Campbell Collaboration effect size calculator (Wilson, n.d.).

We empirically assessed whether a number of baseline variables were predictive of missing values in outcome and also checked whether the number of modules completed (the control group were all scored as having completed no modules) was predictive of loss to follow-up (see Results). As less modules was a predictor of completion of the end of treatment measures, multiple imputation in Mplus version 7.11 (Muthén & Muthén, 1999-2010) was used to estimate missing values using Bayesian analysis (Enders, 2010; Rubin, 1987; Schafer, 1997). The imputation step of the procedure used all the outcome variables (baseline and post-treatment) and the number of modules completed (the control group were all scored as having completed no modules). Here the benefit of MI lies in its ability to incorporate post-randomization variables that are not part of the analysis model (treatment models completed) in the imputation step and so enable an analysis that is valid under a more realistic missing at random (MAR) assumption (Sterne et al., 2009). Ten data sets were imputed.

Clinically significant change was also used in order to determine the number of patients moving outside the dysfunctional range as a consequence of undergoing treatment, and was defined as having a score on the FMPS Concern over Mistakes subscale at post-treatment assessment that was within one standard deviation (6.39) of the mean in the general population (22.32) i.e., < 29 (Suddarth & Slaney, 2001). A Reliable Change Index (RCI) was computed, to show that the difference was considered reliable and not due to measurement error, using the formula $SE_{diff} = SD_1\sqrt{2}\sqrt{(1-r)}$, where SD_1 is the standard deviation at baseline (4.42 in this study) and r is the Cronbach Alpha coefficient of the measure (Jacobsen & Truax, 1991). Change scores are required to exceed 1.96 times the SE_{diff} (Evans, Margison, & Barkham, 1998), in the current study 3.19. Deterioration was determined using a negative change score exceeding the RCI (Rozental et al., 2014).

Examination of change in the weekly clinical perfectionism measures in the intervention group was conducted using Linear Mixed Models (LMM), which accounts for correlations and non-independence amongst observations and allows for an ITT analysis by using Restricted Maximum Likelihood (REML), a method that does not impute data, but uses each case's available data to compute estimates, thus providing the value of the parameter that is most likely to have resulted from the observed data (Nich & Carroll, 1997; Hesser, 2015). An unstructured covariance structure was utilised. LMM is argued to be more accurate than other methods if there are large amounts of missing data on the outcome variable (Von Hippel, 2007) and assumes that the data is missing at random (Han & Guo, 2014). A moderator analysis was conducted in LMM to investigate whether time, the number of modules completed, and the interaction term between these two was associated with significant variation in clinical perfectionism.

Results

Descriptives

One hundred and twenty people were entered into the trial and randomised; N=98 were female and N=62 were currently studying for a degree (University) level qualification. The mean age was 28.85 (SD = 7.99) ranging from 17 to 58 years; 34 (28%) were currently receiving treatment for a mental health problem and 25 of these were currently on medication for a mental health problem.

There were 62 (51.7%) people in the intervention group (49 females, 79%) and 58 (48.3%) in the control group (49 females, 84%). In the intervention group, the mean number of modules completed was 2.48 (SD=2.37) ranging from 0 to 8; 17 (27.5%) people did not complete any modules and were classified as non-engagers. A further 36 (58%) completed half or less of the modules (i.e., 1 to 4 modules), and 9 (14.5%) people completed more than four of the eight modules. There were no significant baseline differences between non-engagers and those who completed any modules across the primary or secondary outcome variables, or across demographic variables (i.e., age, sex, student status, currently receiving mental health treatment).

Missing data

There were 47 people (39%) who did not complete post-treatment measures; significantly more of these came from the treatment group, 31/62 (50%) compared to 16/58 (28%) in the control group, χ^2 (1) = 6.32, p=0.01. Baseline scores for the completers and non-completers are shown in Table 1. There were no differences between the groups with the exception of parental expectations and parental criticism; in both cases, non-completers had significantly higher levels at baseline than completers. However, within the treatment group, those people who did not complete post-treatment measures completed a significantly lower number of modules.

Insert Table 1 here

Completer and intent-to-treat analyses

The differences between the treatment and control group at post-treatment for completers on the primary and secondary outcome measures are summarised in Table 2. It can be seen that for both completer and ITT analyses, there were significant between group differences at the end of the intervention for all variables, with the exception of the negative affect variable (depression, anxiety, and stress) and organisation (ITT). Differences were in the hypothesised direction, with the intervention group having significantly lower levels of perfectionism than the control group. The ITT effect sizes ranged from moderate to large.

Insert Table 2 here

Clinically significant change on the primary outcome variable

Using ITT data, 31 (50%) and 7 (12%) of the treatment and control groups respectively achieved a FMPS Concern over Mistakes subscale score < 29. Clinically significant change *and* exceeding the RCI (Jacobsen & Truax, 1991) was attained by 31 (50%) of the patients receiving treatment compared to 6 (10%) in wait-list control. Patients in the treatment condition were over eight times more likely to experience such improvement compared to those in the control condition, OR=8.67 (95% CI: 3.25-23.11). Furthermore, deterioration on the FMPS Concern over Mistakes subscale was attained by only 1 of the

patients receiving treatment, in comparison to 8 in wait-list control, OR=8.55 (95% CI: 1.03-70.50).

Change in clinical perfectionism and number of modules completed

The number of modules completed is shown in Table 3. The weekly CPQ score was completed by 74% of the treatment group in the first week and this fell steadily each week to just 18% at the final time point. There were no significant predictors of missing status at this final time point (including age, number of modules completed, or any of the baseline outcome variables). Using ITT data, weekly change in clinical perfectionism was examined in the treatment group only (see Figure 2). There was a significant main effect of time, F (13, 28.43 = 28.37, p<0.001), showing a decline in perfectionism. Post-hoc analyses showed that baseline was significantly higher than all subsequent observations, with the biggest within group effect sizes occurring between baseline and week 1, and week 1 and 2. Change in clinical perfectionism between baseline and week 2 was significantly associated with change in FMPS Concern over Mistakes between baseline and post-treatment (r=.35, p < .05).

Insert Table 3 here

In order to examine whether the number of modules completed moderated rate of change in clinical perfectionism over time, we examined whether completing 0 (versus 1-8), or up to 1 (versus 2-8), 2 (versus 3-8), and 3 (versus 4-8) modules resulted in significantly less decreases in clinical perfectionism over time. A significant main effect of time (F [13,

35.15] = 16.60, $p<0.001$), number of modules (F [1, 92.73] = 3.98, $p= 0.049$), and a significant interaction between time and number of modules (F [3, 16.61] = 32.24, $p<0.001$) was only observed when up to 3 modules had been completed, shown graphically in Figure 3. Regardless of whether 0-3 (n = 44, 71%) or 4-8 (N = 18, 29%) modules were completed, both groups experienced a significant decrease in CPQ score over time, associated with a within group Cohen's *d* effect size of 0.96 (95% CI: 0.57-1.34) and 1.15 (95% CI: 0.77-1.53) respectively. Completion of more modules was associated with an overall lower mean CPQ score (M = 26.34, SE = 1.11) compared to fewer modules (M = 29.14, SE = 0.86). While there was no significant between group difference at baseline, different rates of change over time resulted in a significantly lower CPQ score at the end of the intervention for those who completed more modules (M = 22.34, SE = 1.67) than those who completed less (M = 27.71, SE = 1.45), with the mean for the former group commensurate to the community mean for the CPQ.

Insert Figure 2 and Figure 3 here

Given that our negative affect variable was the only one not to show significant change over time compared to the control group, we conducted a post-hoc analysis to investigate whether those participants who completed more modules showed greater reduction with respect to negative affect. In the ITT sample, the correlation between decreased negative affect and number of modules completed was .22 ($p= .08$). In the completer sample, this correlation was significant, $r= .36$ ($p= .048$).

Predictors of treatment outcome

The ability of negative affect at baseline to predict treatment outcome with reference to FMPS Concern over Mistakes and Personal Standards at post-treatment assessment was examined, adjusting for baseline values of the outcome variables. Neither Concern over Mistakes (p=0.25) nor Personal Standards (p=0.38) predicted outcome.

Discussion

The results demonstrated that the intervention for perfectionism had a large positive impact despite high rates of non-engagement and non-completion. Furthermore, not only were the effect sizes for our primary outcome variable large for completers (d=1.61 (95% CI: 1.07-2.14)) and intent-to-treat analyses (d=0.98 (95% CI: 0.60-1.36)), but the treatment group were 8 times more likely than the control group to report a reliable and clinically significant change. Similarly, the mean score at end of treatment on the Clinical Perfectionism Questionnaire was within the normal range for both the ITT and completer groups (Egan et al., 2016). The ITT results should be treated with some caution, as the use of MI and REML can over-estimate the efficacy of treatment and complicate the relationships between different variables and treatment outcome. Nevertheless, the findings compare favourably with both with effect sizes reported in meta-analyses (Lloyd et al., 2015) and more recent studies in which there was a large effect size for group face-to-face CBT for perfectionism (Cohen's d = 1.23) (Egan, van Noort et al., 2014; Handley, Egan, Kane, & Rees, 2015) and a smaller (but still large) effect size for pure online self-help for Frost's Concern over Mistakes subscale (Cohen's d = 0.84) (Egan, van Noort et al., 2014). The intervention also had a small but non-significant impact on depression and anxiety (d= 0.19 (95% CI: -0.17-0.55).The average effect size for internet delivered treatments for depression is between group Cohen's

d=0.21 when support is provided (Johansson & Anderson, 2012) and *d*=0.91 for the treatment of generalized anxiety disorder (Richards, Richardson, Timulak, & McElvaney, 2015).

The impact of the intervention on perfectionism is particularly encouraging since treatment for perfectionism is not widely available – it is not a diagnostic disorder and there are relatively few clinicians trained in the protocol. The strength of the impact of the intervention suggested by our ITT analyses is all the more surprising because while it replicates the Swedish study, only 14.5% people completed five or more modules in this study compared to 96% completing at least half the modules in the Swedish study. The reasons for such a difference in completion rates are likely to be attributable to a range of factors, including failure of the UK study to have an initial telephone conversation or to confirm that participants wanted to take part after completing the screening measures. Other methodological differences may also have contributed. Asking participants in detail about the acceptability of the intervention and procedures in both studies might have helped better understand the reasons for the discrepancy in attrition rates. However, while there was a significant association between the number of modules completed and the outcome, benefit slows down over time on the weekly measure of perfectionism, and large effect size decreases in perfectionism (≥ 0.96) were experienced whether 0-3 or 4-8 modules were completed. Taken together, the findings raise the possibility that a briefer intervention could be developed alongside the longer one, mirroring the work on the prevention of perfectionism in the classroom which is effective in both an 8-session and 2-session format (Fairweather-Schmidt & Wade, 2015; Nehmy & Wade, 2015).

The content of these shorter interventions is suggested by our results where the change was most striking at the very early stages of the intervention with an effect size between baseline and the first week double that of subsequent weekly changes. It is notable that the first session has a strong focus on psychoeducation. As is the case in CBT for bulimia

nervosa and panic disorder, early change was a strong predictor of overall outcome (Agras et al., 2000; Lutz et al., 2014). Anecdotally, participants found the behavioural experiments (module 4) the most challenging and there was a sense in this study that the 'all or nothing' thinking characterising perfectionism made the behavioural task particularly difficult to implement and led to an increase in attrition rates. Nevertheless, a longer version for those highly motivated and keen to obtain maximum benefit might also be warranted.

This study did not find a significant impact of the intervention on any of the DASS subscales. This is a failure to replicate the Swedish findings. The reasons for this are not clear – our participants had relatively high scores on the DASS (within clinical levels) whereas the mean levels of depression and anxiety in the Swedish study were below clinical cut-offs. Our results suggest that this may be an area of psychopathology that does benefit from doing more modules. More work remains to be done to establish the impact of the intervention on related psychopathology.

Many studies do not report the details of the procedure despite the value of detail (Vlaescu, Alasjö, Miloff, Carlbring, & Andersson, in press). In our case, the screening process to identify participants high in perfectionism did not produce baseline scores in perfectionism any higher than those found in the Swedish study that simply asked people to identify themselves as having difficulties with perfectionism. The differences in study procedure were very likely responsible for the difference in engagement and retention. In particular, we consider that meeting participants and checking that they still wish to participate after completion of the screening measures and prior to randomising, are critical in boosting retention (Andersson, Carlbring, Berger, Almlöv, & Cuijpers, 2009). It is likely that having met the person providing the guidance helps build the foundations for the relationship that can help participants maintain their motivation during the challenges presented by the treatment. It is also possible that completion of weekly questionnaires (with reminder emails

for non-completion) is off-putting for participants who have chosen an internet-based treatment partly due to the pressure of busy lives. Multiple emails and demands may be counter-therapeutic despite the benefits of session-by-session monitoring in face-to-face interventions (Lambert et al., 2003). Further consideration of the value (or otherwise) of sessional measures is warranted.

The findings of this study regarding the efficacy and acceptability of this intervention should be interpreted with caution, owing to both the large amount of missing data and the non-systematic variation in procedures from the Swedish study, which makes it difficult to know exactly what has led to reduced engagement. However, in the 15 years since the cognitive behavioural theory and treatment of perfectionism was first developed (Shafran, Cooper, & Fairburn, 2003), it is the case that there is now an effective intervention that can be delivered in multiple formats. Questions remain regarding which of the components are fundamental to change, the specificity of the intervention, the long-term maintenance of change and which factors impact engagement, acceptability and efficacy. These questions are important avenues for further research to maximise engagement and treatment benefit.

Acknowledgements (alphabetical order). Julia Bowen, Tuhina Bhattacharyya, Harriet Clarkson, Sophie Cripps, Jovita Leung, Lauren Robinson, Jean Stafford, George Vlaescu, Martha von Werthern, Chloe Yu Shu.

Funding: The research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors but was supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London.

REFERENCES

Agras, W. S., Crow, S. J., Halmi, K. A., Mitchell, J. E., Wilson, G. T., & Kraemer, H. C. (2000). Outcome predictors for the cognitive behaviour treatment of bulimia nervosa: data from a multisite study. *American Journal of Psychiatry, 157* 1302-1308.

Andersson, G., & Cuijpers, P. (2009). Internet-based and other computerized psychological treatments for adult depression: a meta-analysis *Cognitive Behaviour Therapy, 38,* 196-205.

Andersson, G., Carlbring, P., Berger, T., Almlöv, J., & Cuijpers, P. (2009). What makes internet therapy work? *Cognitive Behaviour Therapy, 38*(SUPPL.1), 55-60. doi:10.1080/16506070902916400

Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42*(2), 111-131.

Dear, B. F., Titov, N., Sunderland, M., McMillan, D., Anderson, T., Lorian, C., & Robinson, E. (2011). Psychometric Comparison of the Generalized Anxiety Disorder Scale-7 and the Penn State Worry Questionnaire for Measuring Response during Treatment of Generalised Anxiety Disorder. *Cognitive Behaviour Therapy, 40*(3), 216-227. doi:10.1080/16506073.2011.582138

Egan, S. J., Shafran, R., Lee, M., Fairburn, C. G., Cooper, Z., Doll, H. A., . . . Watson, H. J. (2016). The reliability and validity of the clinical perfectionism questionnaire in eating disorder and community samples. *Behavioural and Cognitive Psychotherapy, 44,* 79-91.

Egan, S. J., van Noort, E., Chee, A., Kane, R. T., Hoiles, K. J., Shafran, R., & Wade, T. D. (2014). A randomised controlled trial of face to face versus pure online self-help

cognitive behavioural treatment for perfectionism. *Behaviour Research and Therapy, 63*, 107-113.

Egan S.J., Wade T. D. & Shafran R. (2011). Perfectionism as a transdiagnostic process: A clinical review. *Clinical Psychology Review, 31*, 203–212.

Egan, S.J., Wade, T.D, Shafran, R., Antony, M. M. (2014). Cognitive Behavioral Treatment of Perfectionism. *Guilford Press.*

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Press.

Fairburn, C., Cooper, Z., & Shafran, R. (2003). *Clinical Perfectionism Questionnaire*. Department of Psychiatry. University of Oxford, United Kingdom. *Unpublished.*

Fairweather-Schmidt, A.K., & Wade, T.D. (2015). Piloting a perfectionism intervention for pre-adolescent children. *Behaviour Research and Therapy*, 73, 67

Flett, G. & Hewitt, P. (Eds). Perfectionism: Theory, Research, and Treatment. American Psychological Association. Washington, DC. 2002.

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive therapy and research, 14*(5), 449-468

Handley, A. K., Egan, S. J., Kane, R. T., & Rees, C. S. (2015). A randomised controlled trial of group cognitive behavioural therapy for perfectionism. *Behaviour research and therapy*, *68*, 37-47.

Hedeker, D., Gibbons, R. D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics, 24*(1), 70-93

Hedeker, D., Gibbons, R. D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics, 24*(1), 70-93.

Henry, J.D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44,* 227-39.

Hesser, H. (2015). Modeling individual differences in randomized experiments using growth models: Recommendations for design, statistical analysis and reporting of results of internet interventions. *Internet Interventions, 2*(2), 110-120. doi:http://dx.doi.org/10.1016/j.invent.2015.02.003

Hewitt, P. L., & Flett, G. L. (1991). Dimensions of perfectionism in unipolar depression. *Journal of Abnormal Psychology, 100*(1), 98-101

Hewitt, P. L., Flett, G. L., Turnbull-Donovan, W., & Mikail, S. F. (1991). The Multidimensional Perfectionism Scale: Reliability, validity, and psychometric properties in psychiatric samples. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *3*(3), 464.

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

Johansson, R., & Andersson, G. (2012). Internet-based psychological treatments for depression. Expert Review of Neurotherapeutics, 12, 861-870.

Kothari, R., Egan, S., Wade, T., Andersson, G., & Shafran, R. (2016). Overcoming Perfectionism: Protocol of a Randomized Controlled Trial of an Internet-Based Guided Self-Help Cognitive Behavioral Therapy Intervention. *JMIR research protocols*, *5*(4), e215.

Lloyd, S., Schmidt, U., Khondoker, M., & Tchanturia, K. (2015). Can psychological interventions reduce perfectionism? A systematic review and meta-analysis. *Behavioural and cognitive psychotherapy*, *43*(06), 705-731.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & 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.

Lovibond, S., & Lovibond, P. F. (1996). *Manual for the depression anxiety stress scales*: Psychology Foundation of Australia.

Löwe, B., Kroenke, K., Herzog, W., & Gräfe, K. (2004). Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of Affective Disorders, 81*(1), 61-66.

Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., Woods, S. W. & Barlow, D. H. (2014). Patterns of early change and their elationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology, 82,* 287-297.

Nehmy, T., & Wade, T.D. (2015). Reducing the onset of negative affect in adolescents: Evaluation of a perfectionism program in a universal prevention setting. *Behaviour Research and Therapy*, 67, 55-63.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349 (6251*), 910-911. DOI:10.1126/science.aac4716

Richards, D., Richardson, T., Timulak, & McElvaney, J. (2015). Internet Interventions, 3, 272-282.

Rozental, A., Andersson, G., Boettcher, J., Ebert, D. D., Cuijpers, P., Knaevelsrud, C., Ljótsson, B., Kaldo, V., Titov, N., & Carlbring, P. (2014). Consensus statement on defining and measuring negative effects of Internet interventions. *Internet Interventions, 1(1),* 12-19.

Rozental, A., Shafran, R., Wade, T. D., Nordgren, L. B., Carlbring, P., Landstrom, A., Roos, S., Skoglund, M., Thelander, E., Trosell, L., Örtenholm, A., & Andersson, A. (2017). A Randomized Controlled Trial of Internet-Based Cognitive Behavior Therapy for Perfectionism Including an Investigation of Outcome Predictors. *Behaviour Research and Therapy*.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Sterne, J.A., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., … Carpenter, J.R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal,* 339, b2393.

Suddarth, B. H., & Slaney, R. B. (2001). An investigation of the dimensions of perfectionism in college students. *Measurement and Evaluation in Counseling and development, 34*(3), 157-165

Vlaescu, G., Alasjö, A., Miloff, A., Carlbring, P., & Andersson, G. (in press). Features and functionality of the Iterapi platform for internet-based psychological treatment. Internet Interventions. doi: 10.1016/j.invent.2016.09.006

Wilson, D. B., Ph.D. (n.d.). Practical Meta-Analysis Effect Size Calculator [Online calculator]. Retrieved 12, 8, 2016, from https://www.campbellcollaboration.org/effect-size-calculato.html

Yong, E. (2012). In the wake of high profile controversies, psychologists are facing up to problems with replication. *Nature* 483, 298-300.

Zachariae, R., Lyby, M. S., Ritterband, L. M., & O'Toole, M. S. (2016). Efficacy of internet-delivered cognitive-behavioral therapy for insomnia–A systematic review and meta-analysis of randomized controlled trials. *Sleep medicine reviews, 30*, 1-10.

Table(s)

Table 1

Baseline comparison of non-completers and completers

Variable	Non-completers, N=47 M (SD)	Completers, N=73 M (SD)	Odds Ratio (95% Confidence Interval)
Sex (N, % female)	37 (79)	61 (84)	1.37 (0.54-3.49)
Student (N, %)	23 (60)	39 (53)	0.84 (0.40-1.74)
Receiving mental health treatment (N, %)	15 (32)	19 (26)	1.33 (0.60-2.95)
Age	27.69 (7.15)	29.72 (8.43)	0.97 (0.92-1.02)
Concern over mistakes	36.55 (5.30)	35.66 (4.92)	1.05 (0.97-1.15)
Personal standards	30.06 (3.30)	29.45 (3.81)	1.05 (0.95-116)
Doubt about actions	16.13 (2.89)	15.64 (3.01)	1.06 (0.93-1.20)
Organisation	23.87 (5.09)	24.64 (5.45)	0.97 (0.91-1.04)
Parental criticism	12.28 (4.04)	10.48 (4.30)	**1.11 (1.01-1.21)**
Parental expectations	16.30 (4.86)	13.68 (5.60)	**1.10 (1.02-1.18)**
Clinical perfectionism	36.55 (5.30)	35.66 (4.92)	1.04 (0.96-1.12)
Depression, anxiety, stress	28.34 (13.22)	26.88 (11.45)	1.01 (0.90-1.04)
N modules completed [a]	1.10 (1.33)	3.87 (2.39)	**0.44 (0.29-0.67)**

Note: Variables that are significantly different are bolded; [a] treatment group only

Table 2

Between group differences at post-treatment: the first line reports completer analyses and the shaded second line reports intent to treat analyses.

Variable	Baseline Covariate	Treatment group N=62 31 completers	Control group N=58 42 completers	Between group effect size *d* (95% CI)
		Mean (SE)	Mean (SE)	
Concern over mistakes	36.57	26.09 (1.17)	36.31 (0.99)	1.61 (1.07-2.14)
	36.92	27.56 (1.40)	36.62 (0.82)	1.01 (0.63-1.39)
Personal Standards	29.47	24.11 (0.71)	29.23 (0.60)	1.33 (0.82-1.84)
	29.69	25.28 (0.79)	29.01 (0.63)	0.67 (0.31-1.04)
Doubt about actions	15.61	13.05 (0.51)	15.04 (0.43)	0.72 (0.24-1.20)
	15.83	13.62 (0.54)	15.19 (0.50)	0.39 (0.03-0.75)
Organisation	24.64	22.20 (0.43)	24.52 (0.36)	1.00 (0.51-1.49)
	24.34	22.46 (0.79)	24.38 (0.58)	0.36 (-0.004-0.72)
Parental criticism	10.56	8.72 (0.44)	10.75 (0.37)	0.85 (0.37-1.44)
	11.18	9.51 (0.72)	11.51 (0.61)	0.39 (0.03-0.75)
Parental expectations	13.79	11.14 (0.65)	14.57 (0.55)	0.97 (0.48-1.46)
	14.71	12.99 (0.86)	15.32 (0.81)	0.36 (0.001-0.72)

Clinical	35.66	24.93 (0.99)	32.77 (0.85)	1.44 (0.92-1.96)
perfectionism	36.01	25.45 (1.32)	33.42 (0.84)	0.92 (0.55-1.30)
Depression,	27.03	19.69 (1.98)	24.01 (1.67)	0.40 (-0.07-0.87)
Anxiety, stress	27.45	21.15 (2.55)	24.37 (1.67)	0.19 (-0.17-0.55)

Note: Cohen's *d*, and CI=confidence interval

Table 3: Number of Modules Completed by those in the Intervention Group (n=62)

Number of Modules Completed	Number of Participants (%)
0	17 (27.4)
1	9 (14.5)
2	8 (12.9)
3	10 (16.1)
4	9 (14.5)
5	1 (1.6)
6	2 (3.2)
7	2 (3.2)
8	4 (6.5)

Figure 1: Flow of participants through the trial

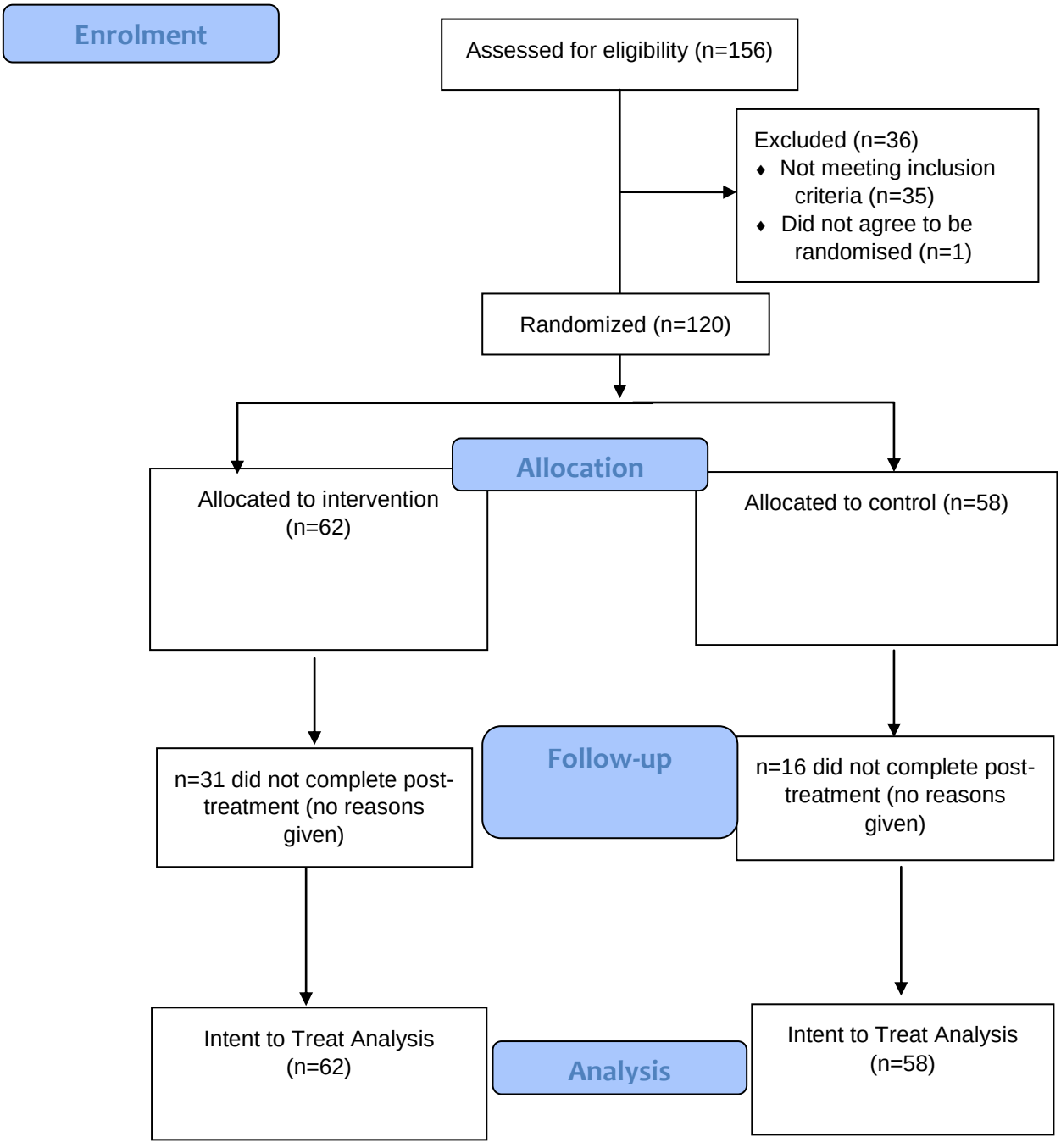
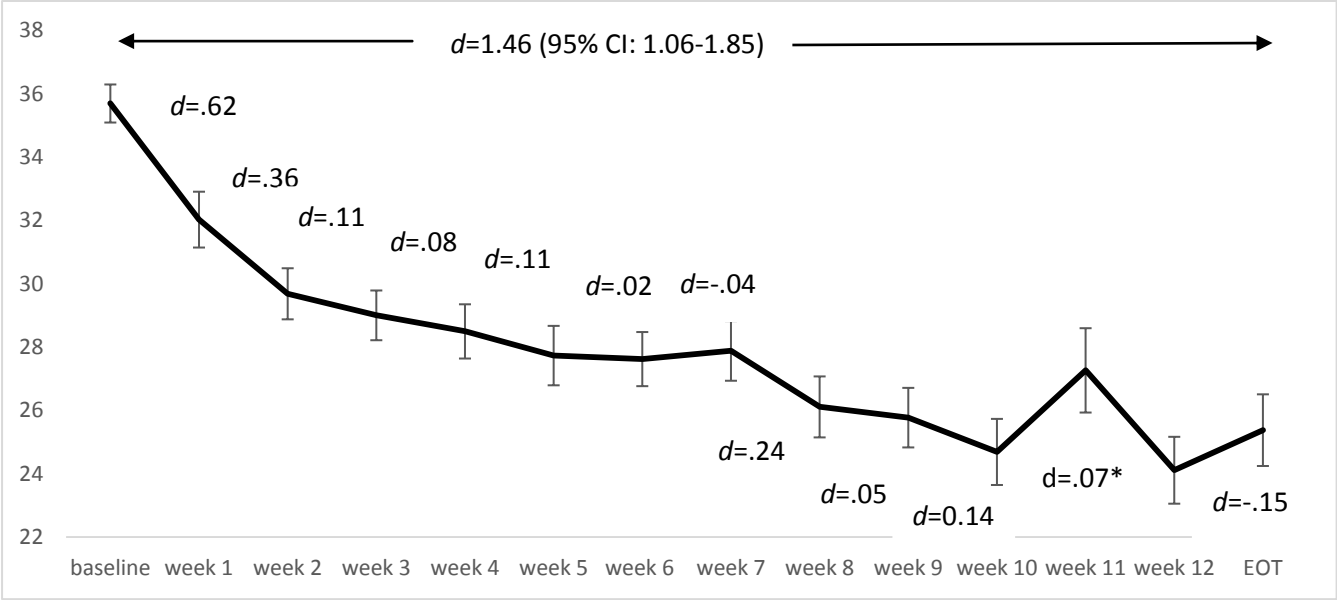


Figure 2

Weekly CPQ score (ITT) and within group effect sizes, Cohen's d



* Within group effect size for week 10 to week 12

Figure 3

*Weekly CPQ score (and standard error bars) for those who completed up to 3 modules, and those who completed more than 3 modules: * denotes a significant between group difference*