Investigating the use of training and implementation intentions to tackle therapist drift in Cognitive Behavioural Therapy

Heidi Trivasse

A thesis submitted in partial fulfilment of the requirements for the award of Doctor of Clinical Psychology at the University of Sheffield

The University of Sheffield

Faculty of Science

Clinical Psychology Unit, Department of Psychology

Submission date: May 2019

Declaration

I declare that this thesis has been submitted for the award of Doctorate in Clinical Psychology at the University of Sheffield. It has not been submitted for any other qualification or to any other academic institution.

Word Count

Literature review

Excluding tables, references and appendices ……...7972

Including tables and references, excluding appendices ……..13732

Research report

Excluding tables, references and appendices ……...8013

Including tables and references, excluding appendices …….12655

Total

Excluding tables, references and appendices …….15948

Including tables and references, excluding appendices ……..26365

Acknowledgments

Firstly, I would like to thank all the people who gave their valuable time to participate in this study, and the organisations and colleagues who supported recruitment. Thank you to Liz Benson, for your hard work, commitment and support in the joint parts of this endeavour! Thank you to my supervisors, Professor Glenn Waller and Dr Thomas Webb, for your expertise, guidance, and availability.

A huge thank you to my wonderful cohort! Thank you for the constant laughter, wisdom, love and support – I'm so grateful to have been on this journey with you. Thank you to my family and friends for your positivity and encouragement since before clinical psychology training began. Finally, thank you to my partner, Somi, for your unconditional love, support, patience, and unwavering belief in me.

Summary

Despite substantial evidence for the effectiveness of Cognitive Behavioural Therapy (CBT), many clinicians fail to implement CBT appropriately. Waller (2009) described therapists' failure to deliver key elements of CBT protocols as 'therapist drift'. Such drift often leads to poorer outcomes for patients and might explain, in part, why CBT does not benefit all patients. This thesis aimed to investigate how to tackle therapist drift and involved: (i) a meta-analysis investigating the effects of exposure therapy training, and (ii) an empirical study investigating the effectiveness of asking therapists to form implementation intentions to support them to adhere more closely to CBT protocols for treating adults with eating disorders – more specifically, increasing their weighing behaviour.

The first part of this thesis reports a review of 14 studies that examined the effectiveness of training clinicians in exposure therapy for anxiety. Meta-analyses were conducted to estimate the effect of training on five outcome variables: knowledge, attitudes, intentions, self-efficacy, and behaviour. Training had large-sized positive effects on clinicians' attitudes and knowledge, medium-sized positive effects on intentions and self-efficacy, but no effect on behaviour. The findings suggest that although clinicians may have the knowledge, confidence and intentions to use exposure therapy, they fail to translate their intentions into action. These findings are in line with research on the intention-behaviour gap, and suggest that future research might consider incorporating volitional interventions (such as implementation intentions) into training to bridge this gap. However, the paucity of available studies and heterogeneity of the effects found among the primary studies indicate that these findings should be interpreted with caution.

The second part of this thesis reports a randomised-controlled trial investigating the possibility that prompting therapists to set goal intentions (to weigh patients with eating disorders) and form implementation intentions (to support these intentions) might help them

to increase their weighing behaviour. Eighty-four therapists actively using CBT with adults with eating disorders were randomised to an intervention or a 'usual practice' control condition. Therapists in the intervention condition received information about the importance of weighing and a volitional help-sheet to support them to form implementation intentions. Therapists completed the outcome measures at baseline, post-intervention, and follow-up. No significant differences were found in weighing behaviour over time for either condition. However, therapists in the intervention condition weighed significantly more patients than those in the control condition at post-intervention – an effect that approached significance at follow-up. There was no significant impact of anxiety or intentions to weigh on the effect of the intervention. The findings of the study suggest that, although implementation intentions did not increase therapists' weighing behaviours, they might help to protect against therapist drift.

Taken together, the two studies provide evidence for the use of implementation intentions to support therapeutic practice. Forming implementation intentions might not only be beneficial for people trying to change their own behaviour but might also support therapists supporting others to change their behaviour. Implementation intentions are a relatively quick, easy and cheap intervention. Therefore, even small effects might warrant the investment of time and effort.

Contents

PART ONE: Literature Review

Does training clinicians in exposure therapy for anxiety disorders change their knowledge, attitudes, or practice? A meta-analysis

Abstract

Objectives

Despite substantial evidence for the effectiveness of exposure therapy in the treatment of anxiety, many clinicians fail to implement it appropriately. The current review aimed to investigate the effectiveness of exposure therapy training in improving clinicians' implementation of exposure therapy.

Methods

A systematic literature search of four databases (PsycINFO, Medline, Scopus, and ProQuest Dissertations and Theses) was conducted. Studies were selected if they evaluated the impact of an exposure therapy training intervention, using quantitative methods from which an effect size could be derived. Meta-analyses were conducted to estimate the effect of training on five outcome variables: knowledge, attitudes, intentions, self-efficacy, and behaviour. Moderator analyses were conducted to explore heterogeneity.

Results

Fourteen studies were included in the meta-analyses. Findings demonstrated that training had large-sized positive effects on clinicians' attitudes (d_+=0.87) and knowledge (d_+=1.05), medium-sized positive effects on intentions and self-efficacy (d_+=0.53 and 0.70, respectively), but no effect on behaviour (d_+=-0.02). The nature of the study's comparison condition and the outcome measure used to measure attitudes significantly moderated the effect of training on clinicians' self-efficacy and knowledge, and attitudes, respectively.

Conclusion

In line with research indicating the 'intention-behaviour gap', findings suggest that although clinicians might have the knowledge, confidence and intentions to use exposure therapy, they fail to translate their intentions into action. Consequently, future research might consider incorporating volitional interventions (e.g., implementation intentions) into training

to bridge this gap. However, the paucity of available studies and the heterogeneity of the effects indicate that these findings should be interpreted with caution.

Practitioner points and limitations

- Training may help clinicians to improve their knowledge, attitudes, confidence and intentions related to implementing exposure therapy.
- Training alone may not be enough to support clinicians to translate their intentions to use exposure therapy into practice.
- The findings of this review are based on a limited number of studies and demonstrated high heterogeneity, and so should be interpreted with caution.
- Studies included in this review were predominantly rated as ‘fair’ quality, suggesting that additional high quality research is required.

Keywords

Exposure therapy, training, intention-behaviour gap, meta-analysis

Introduction

The National Institute for Health and Care Excellence (NICE) recommend Cognitive Behavioural Therapy (CBT) for the treatment of a range of anxiety-based disorders including Obsessive Compulsive Disorder (OCD; NICE, 2005), Social Anxiety Disorder (SAD; NICE, 2013), Generalised Anxiety Disorder and Panic Disorder (GAD; PD; NICE, 2011), and Post-Traumatic Stress Disorder (PTSD; NICE, 2018). Although the manualised content of CBT differs according to the specific anxiety disorder being treated, two treatment elements are present in all recommendations – cognitive restructuring and behavioural change (Waller, 2009). The behavioural aspect of CBT for anxiety-based disorders emphasises the use of exposure therapies.

Exposure therapies encompass a range of therapies (e.g., client- or clinician-led in vivo exposure, interoceptive exposure), which aim to reduce and resolve anxiety through clients' repeated and prolonged confrontation with anxiety-provoking stimuli (Richard & Lauterbach, 2007). Research has consistently shown that exposure therapies are effective. For example, Deacon and Abramowitz (2004) reviewed 19 meta-analytic studies and found that exposure therapies had consistently strong effects on a range of anxiety disorders, and were as effective as combined cognitive and behavioural therapies (with the possible exception of SAD, Ougrin, 2011). Furthermore, exposure therapies have been identified as the active ingredient in the management of anxiety (Gunter & Whittal, 2010). For example, a meta-analysis of 16 randomised-controlled trials (RCTs) found that the effects of exposure therapies and CBT on anxiety were comparable. Adding a cognitive component to exposure-based interventions did not significantly increase the effect size, suggesting that exposure therapies confer a standalone benefit, which is not necessarily enhanced by the cognitive element of CBT (Adams, Brady, Lohr, & Jacobs, 2015).

Barriers to implementing exposure therapy

Despite extensive evidence for the effectiveness of exposure therapies, evidence suggests that many clinicians do not use exposure when treating clients with anxiety (Becker, Zayfert, & Anderson, 2004; Hipol & Deacon, 2012; van Minnen, Hendricks, & Olff, 2010). Even when clinicians do use exposure therapies, they might adapt them or use them in an overly-cautious manner, potentially making them less effective (Abramowitz, Deacon, & Whiteside, 2011; Deacon, Lickel, Farrell, Kemp, & Hipol, 2013; Stobie, Taylor, Quigley, Ewing, & Salkovskis, 2007) – for instance, prematurely terminating the exposure or choosing less distressing exposure tasks, rather than the prolonged and intense exposure recommended in manuals (Farrell, Deacon, Kemp, Dixon, & Sy, 2013). Freiheit, Vye, Swan, and Cady (2004) asked psychologists who regularly treated anxiety disorders which exposure and non-exposure therapy interventions they used. Even though 71% endorsed CBT as their theoretical orientation, only 38%, 12%, and 7-31% used exposure to treat OCD, PD, and SAD, respectively. Waller (2009) termed clinicians' failure to deliver treatments adequately, or at all, 'therapist drift'.

Therapist drift means that clients might not receive effective or competently delivered therapy (Deacon et al., 2013; Farrell, Kemp, Blakey, Meyer, & Deacon, 2016; Waller & Turner, 2016), and so might not obtain the benefits (Gunter & Whittal, 2010). Clinician-level factors that influence and limit the dissemination and implementation of exposure therapies appear to fall into two main areas – (i) clinicians' lack of training, knowledge and confidence in implementing exposure therapies, and (ii) clinicians' anxiety and negative beliefs about exposure therapies.

Lack of training, knowledge and confidence. Becker et al. (2004) found that only 12-28% of clinicians had received training in exposure therapies for PTSD and other anxiety disorders, and 60% reported that limited training was the most important factor preventing

them from using exposure therapies. Similarly, in a qualitative study exploring the barriers to delivering exposure-based CBT for anxiety, clinicians identified the primary barrier as a lack of training and thus a lack of competency and confidence in delivering exposure therapies (Wolitzky-Taylor et al., 2018).

Clinician anxiety and negative beliefs. Exposure therapies have been described as having a 'public relations problem' (Richard & Gloster, 2007), in that clinicians are concerned that exposure therapies can be harmful, intolerable for clients, and potentially unethical (Olatunji, Deacon, & Abramowitz, 2009). Concerns about exposure therapies stem from the fact that exposure therapies require clinicians to purposefully evoke distress in their clients, rather than soothe it; which, on the surface might appear to conflict with the ethical mandate that clinicians do no harm to their clients (Gunter & Whittal, 2010). Relatedly, many clinicians believe that exposure can exacerbate clients' symptoms, prompt drop out, and impact negatively on themselves through vicarious distress or increasing risk of litigation, despite evidence that this is not the case (Olatunji et al., 2009; Rosqvist, 2005).

In addition to negative beliefs about exposure therapies, clinicians' own anxiety is often a barrier to implementing exposure therapies. Specifically, clinicians have reported experiencing high levels of anxiety when conducting exposure therapy (Pittig, Kotter, & Hoyer, 2019; Schare & Wyatt, 2013; Waller & Turner, 2016), and Schumacher et al. (2014; 2015) found high levels of physiological stress responses in clinicians during exposure. In turn, clinician distress has been linked to more cautious exposure therapy delivery (Deacon, Lickel, Farrell, Kemp, & Hipol, 2013), and less time allocated in sessions to exposure work (Scherr, Herbert, & Forman, 2015).

Reducing therapist drift in exposure therapy

Despite clinicians' reservations about exposure therapies, exposure appears to be thought of positively by clients (Brown, Deacon, Abramowitz, Dammann, & Whiteside,

2007; Deacon & Abramowitz, 2004; Olatunji et al., 2009), and positive perceptions appear to remain even when clients find the exposure experience unpleasant. For example, clients receiving exposure for panic disorder perceived it to be useful despite disliking it (Cox, Fergus, & Swinson, 1994). Consequently, it may be more important to target clinician-level drift, rather than client-level factors, to increase the use of exposure as a treatment for anxiety.

Targeting clinician factors, through education, training, and the promotion of positive beliefs about exposure therapy, has been suggested as an effective way to improve the implementation of and adherence to exposure therapies (Farrell, Deacon, Dixon, & Lickel, 2013; Waller & Turner, 2016). Several key principles for exposure therapy training have been outlined. First, trainees should be provided with information regarding the underlying empirical and theoretical principles of exposure therapy, along with examples from actual practice in which intense and prolonged exposure does not lead to negative consequences. The combination of psychoeducation and case examples aims to balance empirical and emotional appeals, and to reduce dissonance between clinicians' cognitions and affect. Second, training should seek to violate clinicians' expectations that exposure is unsafe, intolerable, or unethical. Clinicians may be encouraged to engage in exposure exercises to tackle their own anxiety. Finally, Farrell et al. (2013) suggest that training should encourage clinicians to defend the position that exposure is safe, tolerable, and ethical, through written and verbal exercises. Encouraging clinicians to summarise their learning in their own words is suggested to support sustained learning and attitude change.

Rationale and aims of the current meta-analysis

The review above points to a clear discrepancy between the (strong) evidence-base for exposure therapies, and the relative lack of adherence to and implementation of exposure therapies. It has been suggested that training clinicians in exposure therapy may be a solution

to this therapist drift (Farrell et al., 2013; Waller & Turner, 2016). Given the implications of therapist drift for the care and treatment that clients receive, and for their subsequent clinical outcomes (Deacon et al., 2013; Farrell et al., 2016; Gunter & Whittal, 2010; Waller & Turner, 2016), it is important to understand whether training interventions are an effective solution to therapist drift and to help clinicians change their behaviour.

To the author's knowledge, no systematic literature review has synthesised the empirical research assessing the effectiveness of exposure therapy training interventions. Consequently, it is currently difficult to know whether – and to what extent – training is effective. Therefore, the current review used meta-analytic methods to ascertain the effectiveness of exposure therapy training on clinicians' knowledge and use of exposure therapy for anxiety (i.e., clinicians' behaviour).

The current review also aimed to understand the impact of training on the social-cognitive precursors to changes in clinicians' behaviour. The Theory of Planned Behaviour (TPB; Ajzen, 1991; Figure 1) proposes that intentions are the proximal determinant of an individual's behaviour, and are in turn, the function of three determinants: (i) attitudes, (ii) subjective norms, and (iii) perceived behavioural control. Attitudes are the individual's evaluation of performing the behaviour, subjective norms are the perception of others' views of the behaviour, and perceived behavioural control is the individual's confidence in their ability to perform the behaviour. Inspired by the TPB, the current review also aimed to investigate the effectiveness of training interventions on clinicians' attitudes towards exposure therapy, subjective norms, intentions to use exposure therapy, and self-efficacy or confidence in their ability to use exposure therapy. Self-efficacy has been identified as equivalent to or as a key component of perceived behaviour control (Ajzen, 2002; Bandura, 1977; 1989; 1997), and so measures of self-efficacy will be used to investigate this social-cognitive precursor.

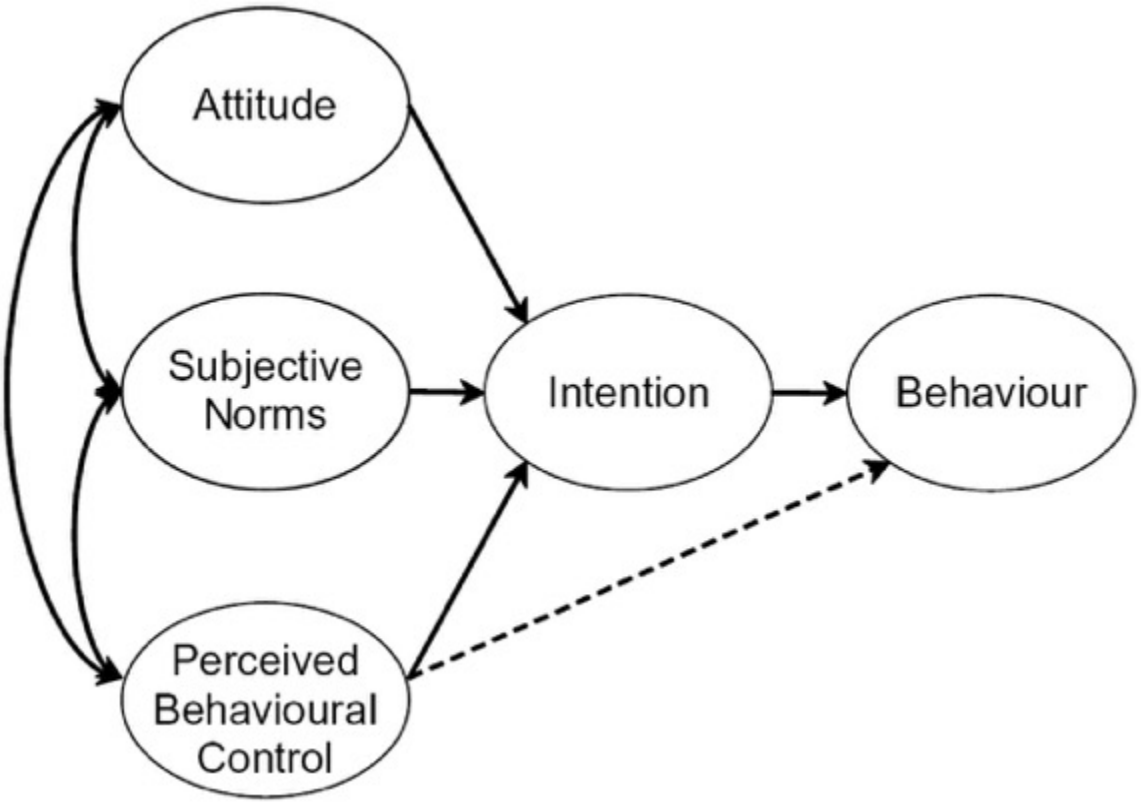


Figure 1: Theory of Planned Behaviour

What factors influence the effectiveness of training? Finally, the current review aimed to investigate factors that might influence how effective training clinicians in exposure therapy is. James, Blackburn, Milne, and Reichfelt (2001) found that previous cognitive therapy experience was associated with clinician competence following cognitive therapy training. Godin, Bélanger-Gravel, Eccles, and Grimshaw (2008) found the following factors moderated the relationship between clinicians' intentions and behaviours: type of professional; study quality; and type of outcome measure (e.g., self-reported vs. objectively assessed). Where possible, these variables were investigated as moderators of the effect of the intervention, along with elements of the study design (e.g., repeated-measures vs. independent-groups).

Method

Literature search strategy

Electronic literature searches of four databases: PsycINFO, Medline, Scopus, and ProQuest Dissertations and Theses were conducted in January 2019. A combination of the search terms outlined in Table 1 were used to identify relevant records. To ensure a comprehensive search, the search terms included synonyms, which were mapped onto relevant subject headings and 'exploded' when possible to include related subject headings.

Table 1: Search terms used in electronic literature searches

Filter 1: Therapist	Filter 2: Training	Filter 3: Exposure	Filter 4: CBT
therapist*	train*	exposure	CBT
clinician*	teach*	"exposure therapy"	"behavio* therapy"
"mental health practitioner*"	dissemin*	"graded exposure"	"cognitive behavio* therapy"
"cbt therapist*"	"overcom* barrier*"	"prolonged exposure"	
"behavio* therapist*"			
psychotherapist*			

Note: The Boolean operator 'OR' was used to combine search terms within each filter, and the operator 'AND' was used to combine filters. The Boolean search modifier '' was used to search for words that begin or end with the truncated search term.*

Inclusion and exclusion criteria

To be eligible for inclusion in this review, studies were required to meet the following criteria. No restrictions were applied regarding the date of publication.

1. Includes an intervention(s) to train clinicians in exposure therapy.
2. Can be accessed in English.
3. Has an experimental research design (including pre- post-intervention designs) used to investigate the effectiveness of the training intervention, and includes quantitative methodology and data analysis.

4. Includes sufficient data to enable calculation of an effect size reflecting the effect of the training intervention.

An outline of the search strategy and flow of studies through the review is presented in Figure 2 (adapted from Moher, Liberati, Tetzlaff, & Altman, 2009). Electronic searches identified 3303 records, which were screened by title and abstract to establish potential relevance. Of the 3303 records, 3253 records (98%) along with 17 duplicates (1%) were removed. Remaining full-texts were read in greater depth and checked against the inclusion criteria, and a further 22 records were removed. Reference lists of relevant reviews and papers were hand-searched, and an additional three records were identified for inclusion. In total, 14 records proved suitable for inclusion in the review.

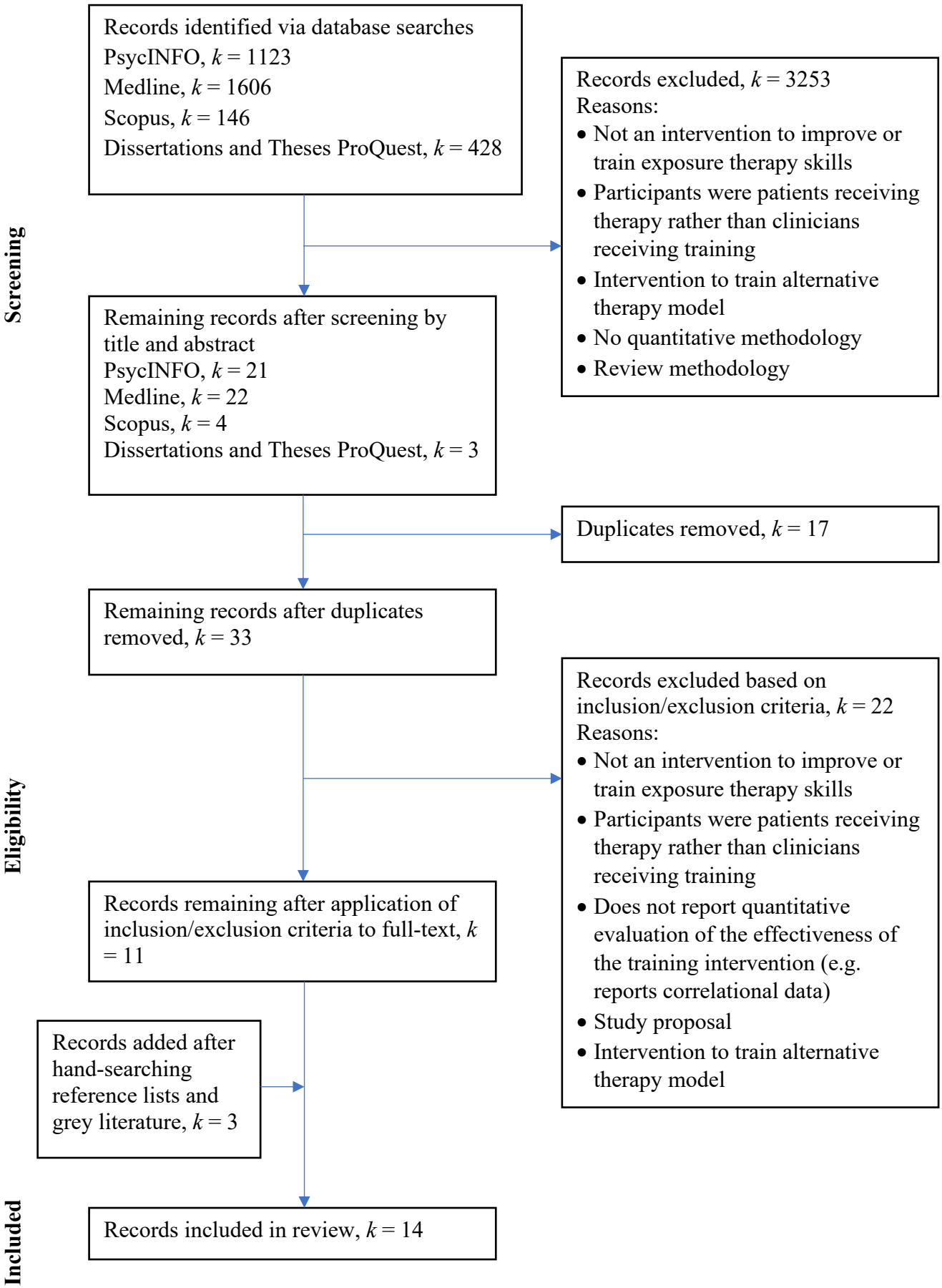


Figure 2: PRISMA diagram outlining results of systematic search strategy

Data extraction and quality assessment

The following data were extracted from each study: (i) participant characteristics; (ii) study characteristics; (iii) description of the training intervention and comparison condition (where applicable); (iv) outcome variables and related outcome measures; (v) means, standard deviations, or test statistics for each outcome variable. Further details of the extracted data are presented in Table 2. Table 3 summarises the characteristics of the primary studies.

Table 2: Details of the data extracted from each primary study

Study characteristic	**Extracted data**	**Details**
Participant characteristics	Age Gender Ethnicity Profession and/or current workplace Education level Experience of exposure therapy	Characteristics were extracted according to available data reported in the primary study. Characteristics were reported across conditions (where applicable) to account for the combination of studies with independent-groups and repeated-measures designs.
Study characteristics	Design Sample size	Design (e.g., RCT, repeated-measures, quasi-randomised design). Total sample size and sample size for each condition or at pre- and post-time-points (where applicable) to calculate effect sizes for meta-analyses.
Information about intervention and comparison conditions (where applicable)	Length of intervention Method of delivery (e.g., face-to-face or online) Key topics and activities covered in training	Descriptions of conditions were extracted according to available data reported in the primary study. Intervention conditions were identified as the more intensive condition (e.g., face-to-face training identified as intervention condition relative to online training; training with additional elements identified as intervention condition relative to training without additional elements).
Outcome variables	Relevant outcome variables including: knowledge, attitudes, subjective norms,	Outcome variable identified by the description of the construct measured in the primary study.

	self-efficacy, intentions, and behaviour	
Outcome measures	Outcome measure used to measure each relevant outcome variable	Name of outcome measure or description of measure provided by the study author if created for the purposes of the study.
Relevant results	Means Standard deviations Test statistics	Relevant quantitative results required to calculate effect size. Means and standard deviations where possible – test statistics when means and standard deviations not available.

Study quality was assessed using the Downs and Black checklist (1998; Appendix A) for assessing methodological quality of randomised and non-randomised studies in healthcare interventions. Studies were scored ('yes', 'no', or 'unable to determine', scored, '1', '0', and '0', respectively) on a range of criteria – reporting, external validity, internal validity, and power of the study. Item five was scored a maximum of two points ('yes', 'partial', or 'no/unable to determine', scored, '2', '1', and '0', respectively). In accordance with previous studies (O'Connor et al., 2015), item 27 was simplified, so that a score of '1' was achieved if the study performed a power calculation, or '0' where no power analysis was conducted or was not reported. Therefore, the overall highest possible score was 28. A total score was derived for each paper (with higher scores reflecting better quality) and score ratings were given the following qualitative labels (O'Connor et al., 2015): 'excellent' (24-28), 'good' (19-23), 'fair' (14-18), and 'poor' (<14). Due to the limited number of studies within the scope of this review, the quality appraisal was not used to exclude studies, but rather to reduce bias and to identify key limitations of the literature for discussion (McDonagh, Peterson, Raina, Chang, & Shekelle, 2013).

A trainee clinical psychologist, blind to the author's rating, acted as an independent rater, repeating the quality assessment process for a random subset (k = 9, 70%) of the included studies. Inter-rater reliability using a two-way random-effects intra-class correlation

coefficient (ICC; Koo & Li, 2016), indicated good inter-rater reliability, *ICC* = .89, (95% CI [.86; .92]). Disagreements and final quality scores were resolved and agreed through discussion.

Table 3: Characteristics of primary studies included in meta-analyses

Authors (year)	**Design**	***N***	**Participant characteristics**	**Comparison condition**	**Intervention**	**Outcome variable**	**Measure(s) of outcome variable**	**Effect size (*d*) for outcome variable**	**Quality score (/28)**
Chin, Bernecker, Buchanan, Cunningham, Schumacher, & Coffey (2017)	Repeated-measures	53	• Community practitioners from substance abuse treatment facilities. • 57.1% master's degree level or educational specialist degree. • 42.9% current counsellor or psychosocial rehabilitation worker. • 74.5% female. • 40.4% 31-45years. • 75.5% Caucasian	No comparison condition	Eight-hour Prolonged Exposure Therapy workshop – didactic training, demonstration videos, and experiential activities such as role-plays.	Attitudes	Five self-report items related to attitudes towards prolonged exposure. Participants responded: 'true', 'false' or 'don't know'.	1.23	13
						Intentions	Three self-report items related to intent to use prolonged exposure measured on 10-point Likert scale.	0.77	
Deacon, Farrell, Kemp, Dixon, Sy, Zhang, & McGrath (2013)	Repeated-measures	162	• Mental health professionals. • Majority master's level in social work or counselling. • Mean (SD) age 51.2 (13.0) years. • 75.9% female. • 29.6% had provided exposure therapy in the last year. • Limited experience or training in exposure therapy.	No comparison condition	Seven-hour didactic workshop on the nature and exposure-based treatment of anxiety disorders. Modifying negative beliefs about exposure therapy was briefly addressed but was not the main focus.	Attitudes	TBES	1.50	15

Farrell, Kemp, Blakey, Meyer, & Deacon (2016)	Quasi-randomised design	49	• Endorsed theoretical orientations: 65.3% CBT, 14.3% family/systems, 10.2% humanistic/client-centred, 6.1% psychodynamic, 2.0% interpersonal, and 2.0% "other". • Mean (SD) age = 51.5 (10.5) years (31-73 years). • 65.3% female. • 93.6% Caucasian • 75.5% Master's degree, 18.4% Ph.D. • Mean 18.7 years of experience (SD = 9.6) in clinical practice.	Standard training workshop – eight-hour didactic instruction on the theory and practice of exposure therapy.	Enhanced training workshop – Standard training workshop plus: summaries of empirical findings that refute concerns about exposure therapy, case presentations and video-based client testimonials selected as emotion-based appeals, simulated interoceptive exposure exercise.	Knowledge	Question: "How thorough is your understanding of the theory and practice of exposure therapy?", (0 = not at all thorough, 10 = extremely thorough)	0.04	15
						Attitudes	TBES	0.77	
						Behaviour	ETCV (self-report)	0.69	
Gega, Norman, & Marks (2007)	RCT	92	• Mental health nursing students. • Mixed gender, age and ethnic origin. • No past training. Participants were in one of three pathways of study: 65% in a 3-year diploma, 11% in a 3-year degree, and 24% in an accelerated 2-year diploma. Students' participation and their results in the RCT were part of extra-curricular skills training and did not count towards examination results.	'Fearfighter' online individual training on exposure therapy, including 'step-by-step' how to devise personalised exposure, and how to troubleshoot difficulties.	Didactic lecture on exposure therapy, including how to do it, and coping techniques to use during exposure.	Knowledge	Two 10-item MCQs.	0.20	18
						Behaviour	Five questions on two case scenarios. Participants answered using short text and were scored for accuracy and completeness. (Assessed)	-0.23	

			• 75% English not first language.						
Harned, Dimeff, Woodcock, & Skutch (2011) Exposure therapy online training (ET OLT)	Pilot RCT	23	• Minimal prior experience of exposure therapy. • Mean (SD) age = 41.37 (11.47) years. • 82.6% female. • 73.9% Caucasian, 6.5% African American, 8.7% Asian American, 8.7% Hispanic/Latino, 2.2% Other. • Range of professions with average 7.45 years worked as treatment provider. • Education: 8.7% BA/BS, 56.5% MA/MA, 28.3% Psy.D/Ph.D/M.D, 6.5% Ph.D dissertation. • 67.4% CBT theoretical orientation.	Control online training (OLT) – 'Dialectical Behaviour Therapy validation principles and strategies'. Comparable in quality, length, and design, and containing no overlapping content.	Exposure therapy online training (ET OLT) – Empirical and theoretical foundations of exposure therapy and how to conduct exposure therapy, activities for participants to construct exposure hierarchy and exposure task, and section highlighting the importance of minimizing avoidance during exposure and providing tips and practice exercises for recognizing and addressing avoidance.	Knowledge	27-item MCQ	2.03	23
						Attitudes	ATETS	0.74	
						Intentions	Adapted Readiness to Change Questionnaire 'action' subscale.	1.04	
						Self-efficacy	30-item self-efficacy questionnaire. All items began with, "I feel confident to" and were rated on 5-point Likert scale (1 = not confident, 5 = very confident).	1.59	
						Behaviour	4-item measure of the application of course content in clinical practice. (Self-report)	0.43	
Harned, Dimeff, Woodcock, & Skutch (2011)	Pilot RCT	23	• Minimal prior experience of exposure therapy. • Mean (SD) age = 41.37 (11.47) years.	Control online training (OLT) – 'Dialectical Behaviour Therapy	ET OLT plus motivational interviewing (ET OLT+MI) – In addition to ET OLT (see above,	Knowledge	27-item MCQ	3.22	23
						Attitudes	ATETS	0.95	

Exposure therapy online training + motivational interviewing (ET OLT+MI)			• 82.6% female. • 73.9% Caucasian, 6.5% African American, 8.7% Asian American, 8.7% Hispanic/Latino, 2.2% Other. • Range of professions with average 7.45 years worked as treatment provider. • Education: 8.7% BA/BS, 56.5% MA/MA, 28.3% Psy.D/Ph.D/M.D, 6.5% Ph.D dissertation. • 67.4% CBT theoretical orientation.	validation principles and strategies'. Comparable in quality, length, and design, and containing no overlapping content.	Harned et al., 2011 ET OLT), participants participated in 1–2 brief (up to 20- min) Motivational Interviewing-based phone calls to decrease ambivalence about adopting exposure therapies due to attitudinal barriers.	Intentions Self-efficacy Behaviour	Adapted Readiness to Change Questionnaire 'action' subscale 30-item self-efficacy questionnaire. All items began with, "I feel confident to" and were rated on 5-point Likert scale (1 = not confident, 5 = very confident). 4-item measure of the application of course content in clinical practice. (Self-report)	0.69 2.06 0.07	
Harned, Dimeff, Woodcock, Kelly, Zavertnik, Contreras, & Danner (2014) Online training + Motivational enhancement (OLT+ME)	RCT	90	• Minimal prior experience of exposure therapy. • Mean (SD) age = 37.4 (10.3) years. • 71.3% female. • 72.1% Caucasian, 8.4% African American, 4.5% Asian American, 6.1% Hispanic/Latino, 8.9% multiracial. • Education: 11.8% BA/BS/RN, 67.1% MA/MS, 21.1% Psy.D/Ph.D/M.D/ABD.	Online 'Foundations of Exposure Therapies' training (OLT) – 10-hour online didactic training course with simulated clinical scenarios.	OLT plus motivational enhancement (OLT+ME) – In addition to OLT, participants received two brief ME interventions to address attitudinal barriers to learning and using exposure therapy, including a five-minute video at the start of the OLT, and an online module at the end of the OLT with a virtual exposure therapy consultant.	Knowledge Attitudes Self-efficacy	49-item MCQ ATETS Adapted 27-item self-efficacy subscale of the Behavioral Anticipation and Confidence questionnaire. All items began with, "I feel confident to" and were rated on 5-point Likert scale	-0.14 0.41 0.19	23

			• Range of professions with average 4.8 years worked since degree.				(1 = not confident, 5 = very confident).		
						Behaviour	ETCS. (Self-report) Structured role plays coded by trained research assistant. (Assessed)	0.04	
Harned, Dimeff, Woodcock, Kelly, Zavertnik, Contreras, & Danner (2014) Online training + Motivational enhancement + Learning Community (OLT+ME+LC)	RCT	91	• Minimal prior experience of exposure therapy. • Mean (SD) age = 37.4 (10.3) years. • 71.3% female. • 72.1% Caucasian, 8.4% African American, 4.5% Asian American, 6.1% Hispanic/Latino, 8.9% multiracial. • Education: 11.8% BA/BS/RN, 67.1% MA/MS, 21.1% Psy.D/Ph.D/M.D/ABD. • Range of professions with average 4.8 years worked since degree.	Online 'Foundations of Exposure Therapies' training (OLT) – 10-hour online didactic training course with simulated clinical scenarios.	OLT+ME, plus learning community (OLT+ME+LC) – In addition to OLT+ME (see above, Harned et al., 2014), participants attended eight 1-hour LC meetings via an online conferencing platform facilitated by an experienced exposure therapy clinician. The first five meetings targeted knowledge acquisition and practice, and the next three meetings focused on increasing use of and clinical proficiency of exposure therapy. Homework assignments were given after each meeting.	Knowledge	49-item MCQ	0.29	23
						Attitudes	ATETS	0.43	
						Self-efficacy	Adapted 27-item self-efficacy subscale of the Behavioral Anticipation and Confidence questionnaire. All items began with, "I feel confident to" and were rated on 5-point Likert scale (1 = not confident, 5 = very confident).	0.26	
						Behaviour	ETCS. (Self-report) Structured role plays coded by trained research assistant. (Assessed)	0.16	

Kaye (2018)	Quasi-randomised design	99	• 84% female. • 79% Caucasian. • 70% endorsed CBT orientation. • Average level of knowledge about exposure therapy was 44.88 (SD = 22.32) on scale 0-100	Standard exposure therapy training (ST) workshop, incorporating features to target clinicians' negative beliefs about exposure therapy.	ST plus acceptance-based techniques (ST+ABT) – ST workshop, plus role-play practice in developing fear hierarchies, and techniques intended to reduce clinicians' experiential avoidance of anxiety and discomfort during exposure.	Knowledge	49-item MCQ	1.29	20
						Attitudes	TBES	1.52	
						Intentions	5-items measuring likelihood of implementing different types of exposure	-0.15	
						Behaviour	ETCV. (Self-report) Exposure sessions videoed and coded by trained research assistants. (Assessed)	-0.28	
Kobak, Wolitzky-Taylor, Craske, & Rose (2017) Exposure module of CBT training intervention	Repeated-measures	70	• Community clinicians. • 83% female. • 85% Caucasian, 10% African American, and 5% other or mixed racial categories. • Mean (SD) age = 47.7 (13.7) years, (23-86 years). • Mean (SD) years of clinical experience 12.2 (9.5). Range: 1–35 years.	No comparison condition	Exposure module within web-based tutorial on CBT concepts and skills, including: differences between cognitive restructuring and exposure therapy, goals and critical factors in exposure therapy, therapist's role, designing exposure task, identifying and addressing avoidance of exposure therapy, and	Knowledge	46-item MCQ	2.65	18

			• 72% social workers, 24% psychologists, and 4% marriage and family therapists. • 40% received some type of prior formal training in CBT.		rationale for the frequency, timing and duration of exposure sessions.				
McDonough & Marks (2002)	RCT	37	• 3rd year medical students from King's College Hospital medical school. • 5 weeks into their 6-week clinical attachment in psychiatry. • 54% female.	'Fearfighter' online individual training on exposure therapy, including 'step-by-step' how to devise personalised exposure, and how to troubleshoot difficulties.	Interactive face-to-face exposure therapy tutorial in small groups	Knowledge	15-item MCQ	0.67	17
Reid, Guzick, Balkhi, McBride, Geffken, &	Repeated-measures	42	• 60% doctoral students in a clinical, counselling, or school psychology, 19% pre-doctoral interns in	No comparison condition	Didactic presentation teaching basic principles of exposure therapy, challenges, and	Attitudes	TBES	0.68	16

McNamara (2017)			clinical psychology, 12% post- doctoral associates, and 10% master's students in mental health counselling. • 81% female. • Mean age 27.10 years (SD 4.10). • 71% Caucasian, 14% Black, 7% Hispanic/Latino, 5% Indian/Middle Eastern, 2% Asian.		commonly held negative beliefs. Licensed supervisors systematically and regularly assess trainee competency. Supervisor-to-trainee and trainee-to-trainee feedback is provided by live modelling during treatment sessions and through individual and group supervision.	Behaviour	ETDS. (Self-report) ETCV. (Assessed)	-0.33	
Ruzek, Eftekhari, Rosen, Crowley, Kuhn, Foa, Hembree, & Karlin, (2016)	Repeated-measures	943	• 46.5% clinicians in specialised outpatient PTSD clinics, 26.6% clinicians in general mental health outpatient clinics, 6.9% clinicians in PTSD residential programs, and 16.6% clinicians in other clinic types. • 57.5% doctoral-level psychologists, 35.8% master's-level clinicians • 66.4% female. • 61.4% described theoretical orientation as CBT. • 74.4% never received formal training in prolonged exposure.	No comparison condition	Four-day interactive prolonged exposure therapy (PET) training workshop, including didactic teaching, supervised role-play, videos demonstrating core elements of PET, and discussion.	Attitudes Self-efficacy	Questionnaire assessing the degree to which clinicians placed value on specific exposure therapy treatment goals. Question: "How confident are you in your ability to effectively deliver the following aspects of prolonged exposure?" on 7-point scale.	0.17 0.55	17

Waller, D'Souza Walsh, Wright (2016)	Repeated-measures	34	• 88% female. • 85.3% Caucasian. • Mean (SD) age = 39.0 (10.4) years. • Range of professions: 38% clinical psychology, 21% dietetics, 12% psychiatry, 6% nursing, 6% social work, 3% family therapy, 3% occupational therapy, 3% counselling, 3% psychotherapy, and 3% art therapy. One participant did not state their profession.	No comparison condition	90-minute didactic teaching session covering the theory and evidence of exposure therapy for eating disorders.	Attitudes	TBES	1.15	19

Note: ATETS = Attitudes Towards Exposure Therapy Scale (Harned et al., 2011); TBES = Therapist Beliefs about Exposure Scale (Deacon et al., 2013); ETCS = Exposure Therapy Clinical use Survey (Harned et al., 2014); ETCV= Exposure Therapy Case Vignette (Deacon et al., 2013); ETDS = Exposure Therapist Delivery Scale (Reid et al., 2017); MCQ=Multiple-choice Questionnaire

Computing effect sizes from the primary studies

Key outcome variables were identified within each study including: attitudes towards exposure therapy, knowledge about exposure therapy, intentions to use exposure therapy, subjective norms, clinicians' self-efficacy regarding their use of exposure therapy, and use of exposure therapy. When possible, Cohen's *d* effect sizes were calculated for each outcome variable, using means and standard deviations reported in the paper. Due to the variability in follow-up time-points across studies, effect sizes were calculated using data from the first post-intervention time-point. Meta-Essentials workbooks were used to complete the meta-analyses (Suurmond, van Rhee, & Hak, 2017). In one study (Harned et al., 2014), the reported standard errors were converted to standard deviations (see Appendix B for equation), prior to inclusion in the analyses. Chin et al. (2017) did not report the means and standard deviations for the outcome measure measuring attitudes, and so the Chi-squared test statistic was converted into Cohen's *d* using the Psychometrica Effect Size Calculator (Lenhard & Lenhard, 2016). The effect size for studies with repeated-measures designs was computed according to Morris and DeShon's (2002) recommendations for combining repeated-measures and independent-groups data within a meta-analysis (e.g., where it was not reported, a correlation of 0.50 was assumed between pre- and post-measures for studies with repeated-measures designs).

Where multiple measures were used to assess the same outcome variable, effect sizes were calculated separately for each measure using the Psychometrica Effect Size Calculator (Lenhard & Lenhard, 2016). Effect sizes were then averaged to provide one effect size reflecting each outcome to be included in the main analyses (Card, 2012). Where studies compared more than one intervention group with a control group (e.g., Harned et al., 2011; 2014), both comparisons were included as separate studies and the sample size of the

comparison conditions (against which both intervention groups were compared) was halved to avoid violating the assumption of independence.

Meta-analytic strategy

Sample-weighted average effect sizes (d_+) were combined using a random effects model, as studies were likely to be "different from one another in ways too complex to capture by a few simple study characteristics" (Cooper, 1986, p. 526). Effect sizes were interpreted in line with Cohen's (1992) recommendations (d = 0.20 representing 'small', d = 0.50 representing 'medium', and d = 0.80 representing 'large'). 'Q' and 'I^2' statistics were used to estimate the heterogeneity of effect sizes from the primary studies. A significant Q statistic indicates that the variability exceeds what would be expected based on sampling error (Higgins, Thompson, Deeks, & Altman, 2003). The I^2 statistic indicates the percentage of variation across the studies that is not explained by chance (Higgins, et al., 2003). I^2 values between 0-40% were considered 'low', 40-60% 'moderate', 60-90% 'substantial', and above 90% 'considerable' heterogeneity (Cochrane Consumers and Communication Review Group, 2016).

Moderator analyses were conducted to explore factors that might account for any heterogeneity in the effect sizes within each primary meta-analysis (Borenstein, Hedges, Higgins, & Rothstein, 2009), provided that there were at least two studies that represented each level of the moderator. When possible, studies were split into subgroups based on elements of the study design, the nature of outcome measures used, participant characteristics (e.g., previous experience, profession), and study quality. An analogue to Analysis of Variance (ANOVA) was conducted using Meta-Essentials workbooks (Suurmond et al., 2017) for each subgroup analysis to assess whether the subgroup variables could account for variability within the primary meta-analyses.

Publication bias was examined visually using funnel plots (Light & Pillemer, 1984; Field & Gillett, 2010), and statistically using Egger's regression (Egger, Davey, Smith, Schneider, & Minder, 1997). Orwin's (1983) formula was used to determine the fail-safe N, as an additional measure of publication bias. This provides an estimate of the number of studies with a null finding that would be required to overturn the statistical significance of findings.

Results

Study characteristics

The characteristics of the primary studies are outlined in Table 3. Study sample sizes ranged from 23 to 943 participants. Studies were predominantly conducted in the USA, with one study conducted in the UK (Gega et al., 2007) and one study conducted at an international conference (Waller et al., 2016). All studies had a majority female and Caucasian sample, although this is likely to be representative of mental health clinicians in the studied countries (e.g., Memon et al., 2016; Morison, Trigeorgis, & John, 2014). Studies included a range of professionals, such as psychologists, mental health nurses, medical students, and a range of community mental health clinicians. Studies varied in their methodology, including six repeated-measures designs, two quasi-experimental designs, and six RCTs. Intervention conditions included didactic teaching sessions, face-to-face workshops, and online training. Most training sessions included information and activities aimed at tackling clinicians' negative beliefs about exposure therapy, and some used simulated scenarios to support clinicians to apply their learning. Most studies with a comparison condition utilised an active comparison condition, which included an aspect of exposure therapy training. Two studies utilised a passive comparison condition, in which participants engaged in online training about 'Dialectical Behaviour Therapy validation

principles and strategies'. This passive comparison condition was comparable to the training intervention in quality, length, and design, and contained no overlapping content.

Outcome measures. Most studies measuring attitudes towards exposure therapy used the Therapist Beliefs about Exposure Scale (TBES) or the Attitudes Towards Exposure Therapy Scale (ATETS). However, one study created a questionnaire to assess the value that clinicians placed on exposure therapy treatment goals (Ruzek et al., 2016) and one study created a questionnaire that asked clinicians to indicate whether they believed that a series of statements relating to attitudes towards exposure therapy were true or false (Chin et al., 2017). Most studies used multiple-choice questionnaires to assess clinicians' knowledge of exposure therapy, except one study (Farrell et al., 2016), which asked clinicians to rate their knowledge of exposure therapy theory and practice. Self-efficacy was typically measured using questionnaires designed by the study authors, or an adapted version of the self-efficacy subscale of the Behavioural Anticipation and Confidence questionnaire (Dimeff et al., 2009). Behaviour was assessed using self-report measures including: the Exposure Therapy Case Vignette; Exposure Therapy Delivery Scale; Exposure Therapy Clinical Use Survey; or a questionnaire designed by the authors. Behaviour was also assessed through coding of exposure therapy sessions, or clinicians' responses to a series of case scenarios.

Quality appraisal

Quality assessment ratings are presented in Appendix C. Quality scores ranged from 13 ('poor') to 23 ('good'). The quality of most studies was rated as 'fair' (k = 7, 50%), six studies (43%) were rated as 'good', and one study (1%) was deemed to be 'poor'. The main strengths of the studies were the clear reporting of aims, outcome measures and interventions. Most studies used appropriate statistical analyses, reported their proposed analyses, and did not employ 'data dredging' (Smith, 2002). Eight studies (57%) reported power analyses, but it was not possible to determine whether six (46%) of the studies were adequately powered.

Regarding external validity, most studies did not report the population from which participants were recruited, nor the representativeness of the sample in relation to the intended population. However, studies that recruited to different groups reported that participants were recruited from the same population, which improves internal validity. Given the active nature of the interventions, it was not possible for any study to blind participants to the condition, and studies did not report whether participants were blind to the fact that other participants received a different intervention. Four studies (29%) attempted to blind the researchers measuring the main outcomes, but most studies used self-report measures, which limits the internal validity. Compliance with the intervention was only reported in three studies (21%), and so it is unclear how closely participants engaged with and adhered to the training. No studies provided a list of possible adverse events.

Impact of training on outcome variables

Random effects meta-analyses were conducted to examine the effects of training on five outcome variables: knowledge, attitudes, intentions, self-efficacy, and behaviour. None of the studies identified as suitable for inclusion in the review examined the effect of training on subjective norms. Table 4 reports a summary of the findings of these analyses. Positive effect sizes indicate an improvement in the outcome. For example, a positive effect size would indicate more positive attitudes towards exposure therapy.

Table 4: Summary of primary meta-analyses results

Outcome variable	***N***	***k***	***d*$_+$ (SE)**	**95% CI**	***p***	***Q* (*p*)**	***I*^{2} (%)**
Knowledge	607	9	1.05 (0.40)	0.14; 1.96	.008	126.77 (*<.001*)	93.69
Attitudes	1565	11	0.87 (0.14)	0.56; 1.18	<.001	244.23 *(< .001*)	95.91
Intentions	190	4	0.53 (0.26)	-0.31; 1.37	.043	17.27 (*.001*)	82.63
Self-efficacy	1134	5	0.70 (0.32)	-0.19; 1.59	.030	15.86 (*.003*)	74.78
Behaviour	543	8	-0.02 (0.12)	-0.31; 0.27	.853	13.98 (*.052*)	49.93

Note: N = total number of participants; k = number of studies; d_+ = sample-weighted average effect size; SE = standard error; CI = confidence interval

Knowledge. Nine studies evaluated the impact of training on knowledge. The effect sizes ranged from d = -0.14 (Harned et al., 2014) to d = 3.22 (Harned et al., 2011). The sample weighted average effect size was d_+ = 1.05, 95% CI [0.14, 1.96], p = .008 derived from a total sample size of N = 607 (see Figure 3 for forest plot). This indicates that the training interventions had, on average, a large-sized positive effect on clinicians' knowledge of exposure therapy. The homogeneity statistic was significant, $Q(8) = 126.77$, $p < .001$, indicating that the effect sizes from the primary studies were varied, and the I^2 statistic (93.69%) also indicated considerable heterogeneity. Visual inspection of the funnel plot (Figure 4) suggested some asymmetry and thus risk of publication bias. However, Egger's regression was not significant (p = .068). Orwin's (1983) fail-safe N analysis indicated that 181 studies with trivial effect sizes (i.e., d = 0.05) would be required to overturn the conclusion that training clinicians in exposure therapy has a positive impact on their knowledge of exposure therapy.

#	Study name	Effect size	CI Lower limit	CI Upper limit	Weight
1	Farrell (2016)	0.04	-0.54	0.62	11.44%
2	Gega (2007)	0.20	-0.12	0.52	12.19%
3	Harned (2011) ET OLT	2.03	0.93	3.13	9.47%
4	Harned (2011) ET OLT+MI	3.22	1.87	4.57	8.42%
5	Harned (2014) OLT+ME	-0.14	-0.64	0.36	11.70%
6	Harned (2014) OLT+ME+LC	0.29	-0.21	0.79	11.70%
7	Kaye (2018)	1.29	1.03	1.55	12.31%
8	Kobak (2017)	2.65	2.15	3.15	11.70%
9	McDonough (2002)	0.67	-0.02	1.36	11.07%
10					
11					

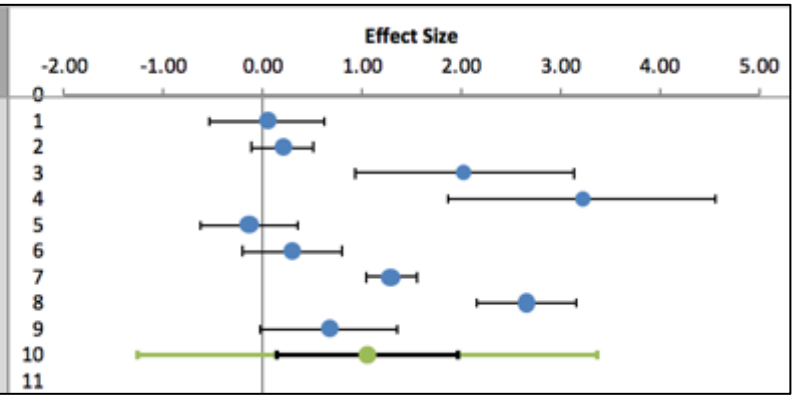


Figure 3: Forest Plot presenting the effects of training on 'Knowledge' outcome variable

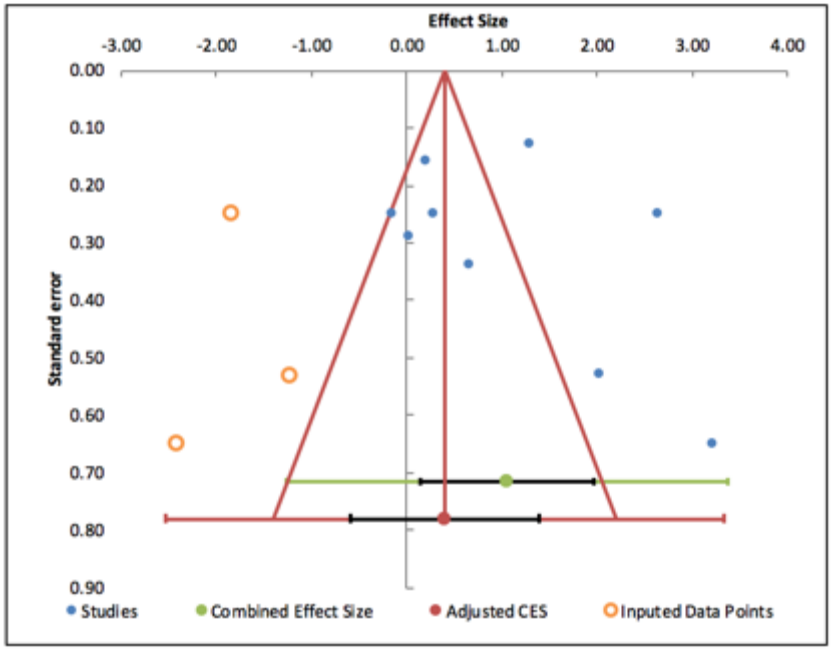


Figure 4: Funnel Plot for 'Knowledge' outcome variable meta-analysis

Attitudes. Eleven studies evaluated the impact of training on attitudes. Effect sizes ranged from $d = 0.17$ (Ruzek et al., 2016) to $d = 1.52$ (Kaye, 2018). The sample weighted average effect size was $d_+ = 0.87$, 95% CI [0.56, 1.18], $p < .001$, derived from a total sample size of $N = 1565$ (see Figure 5 for forest plot). This indicates that training had, on average, a large-sized positive effect on clinicians' attitudes towards exposure therapy. The homogeneity statistic was significant, $Q(10) = 244.23$, $p < .001$, indicating that the effect sizes from the primary studies were varied, and the I^2 statistic (95.91%) also indicated considerable heterogeneity. Visual inspection of the funnel plot (Figure 6) did not show asymmetry and Egger's regression was not significant ($p = .789$). Orwin's (1983) fail-safe N analysis indicated that 182 studies with trivial effect sizes would be required to overturn the conclusion that training clinicians in exposure therapy has a positive impact on their attitudes towards exposure therapy.

#	Study name	Effect size	CI Lower limit	CI Upper limit	Weight
1	Chin (2017)	1.23	0.80	1.67	9.38%
2	Farrell (2016)	0.77	0.17	1.37	8.72%
3	Harned (2011) ET OLT	0.74	-0.19	1.67	7.35%
4	Harned (2011) ET OLT+MI	0.95	0.00	1.90	7.26%
5	Harned (2014) OLT+ME	0.41	-0.09	0.91	9.13%
6	Harned (2014) OLT+ME+LC	0.43	-0.09	0.95	9.05%
7	Kaye (2018)	1.52	1.22	1.82	9.81%
8	Reid (2017)	0.68	0.34	1.02	9.70%
9	Ruzek (2016)	0.17	0.11	0.23	10.22%
10	Waller (2016)	1.15	0.70	1.60	9.36%
11	Deacon (2013)	1.50	1.28	1.72	10.01%
12					

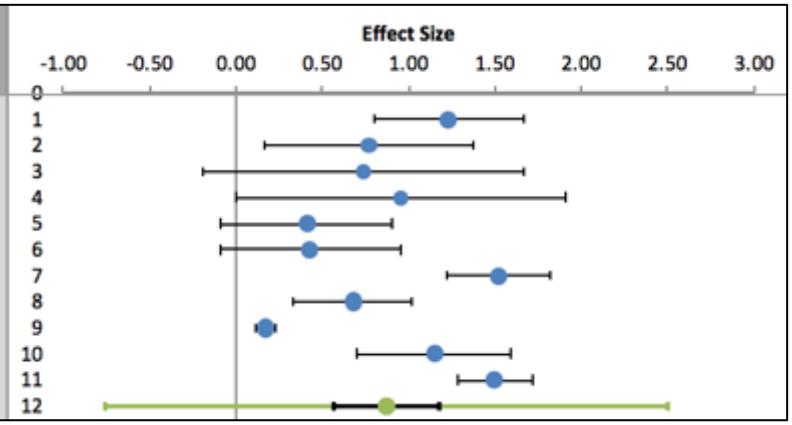


Figure 5: Forest Plot presenting effects of training on 'Attitudes' outcome variable.

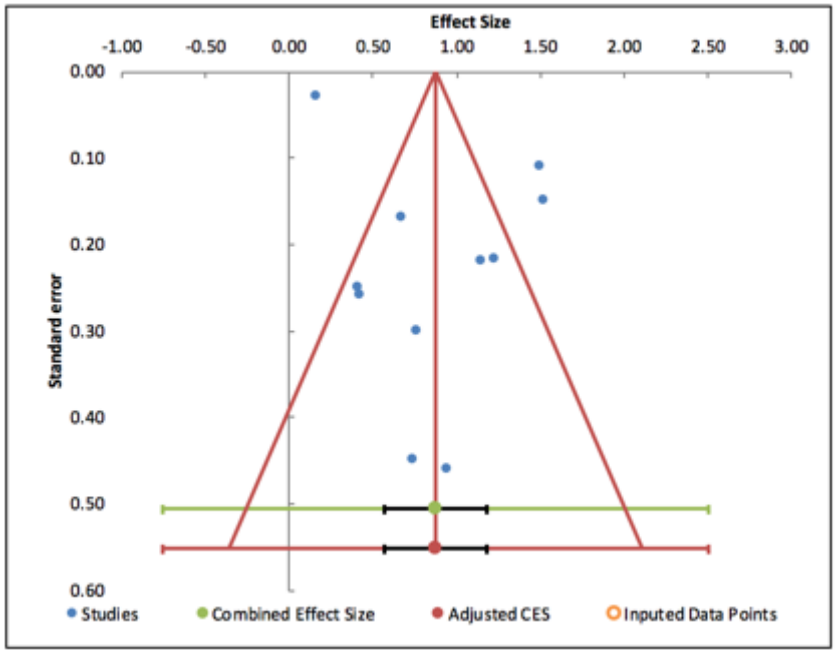


Figure 6: Funnel Plot for 'Attitudes' outcome variable meta-analysis

Intentions. Four studies evaluated the impact of training on intentions to use exposure therapy. Effect sizes ranged from d = -0.15 (Kaye, 2018) to d = 1.04 (Harned et al., 2011, ET OLT). The sample weighted average effect size was d_+ = 0.53, 95% CI [-0.31, 1.37], p = .043, derived from a total sample size of N = 190 (see Figure 7 for forest plot). This indicates that the training interventions had, on average, a medium-sized positive effect on clinicians' intentions to use exposure therapy. The homogeneity statistic was significant, $Q(3) = 17.27$, p = .001, indicating that the effect sizes from the primary studies were varied, and the I^2 statistic (82.63%) also indicated substantial heterogeneity. Visual inspection of the funnel plot (Figure 8) did not show asymmetry and Egger's regression was not significant (p = .542). Orwin's (1983) fail-safe N analysis indicated that 39 studies with trivial effect sizes would be required to overturn the conclusion that training clinicians in exposure therapy has a positive impact on their intentions to use exposure therapy.

#	Study name	Effect size	CI Lower limit	CI Upper limit	Weight
1	Chin (2017)	0.77	0.55	0.99	32.39%
2	Harned (2011) ET OLT	1.04	0.09	1.99	18.90%
3	Harned (2011) ET OLT+MI	0.69	-0.24	1.62	19.27%
4	Kaye (2018)	-0.15	-0.55	0.25	29.45%
5					
6					

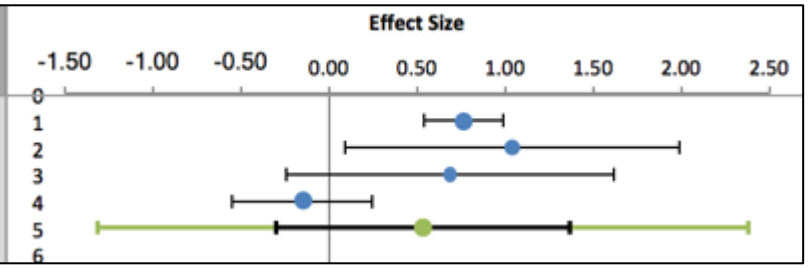


Figure 7: Forest Plot presenting effects of training on 'Intentions' outcome variable.

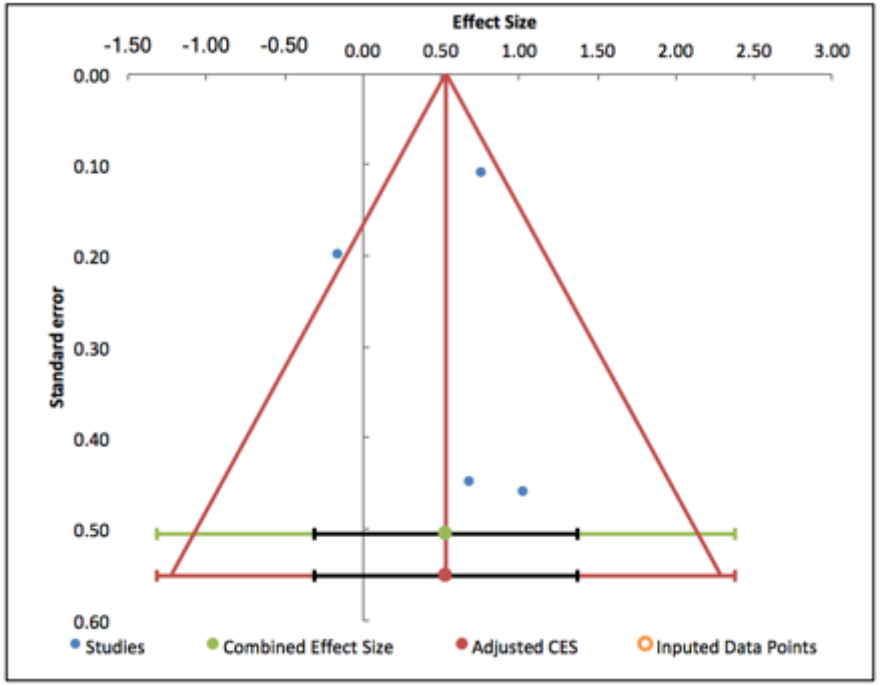


Figure 8: Funnel Plot for 'Intentions' outcome variable meta-analysis

Self-efficacy. Five studies evaluated the impact of training on self-efficacy associated with using exposure therapy. Effect sizes ranged from $d = 0.19$ (Harned et al., 2014, OLT+ME) to $d = 2.06$ (Harned et al., 2011, ET OLT+MI). The sample weighted average effect size was, $d_+ = 0.70$, 95% CI [-0.19, 1.59], $p = .030$ derived from a total sample size of $N = 1134$ (see Figure 9 for forest plot). This indicates that the training interventions had, on average, a medium-to-large positive effect on clinicians' belief in their ability to use exposure therapy. The homogeneity statistic was significant, $Q(4) = 15.86$, $p = .003$, indicating that the effect sizes from the primary studies were varied, and the I^2 statistic (74.78%) also indicated substantial heterogeneity. Visual inspection of the funnel plot (Figure 10) suggested some asymmetry and thus risk of publication bias. However, Egger's regression was not significant ($p = .072$). Orwin's (1983) fail-safe N analysis indicated that 65 studies with trivial effect sizes would be required to overturn the conclusion that training clinicians in exposure therapy has a positive impact on their self-efficacy when using exposure therapy.

#	Study name	Effect size	CI Lower limit	CI Upper limit	Weight
1	Harned (2014) OLT+ME	0.19	-0.31	0.69	22.57%
2	Harned (2014) OLT+ME+LC	0.26	-0.24	0.76	22.57%
3	Harned (2011) ET OLT	1.59	0.55	2.63	12.08%
4	Harned (2011) ET OLT+MI	2.06	0.96	3.16	11.22%
5	Ruzek (2016)	0.55	0.49	0.61	31.57%
6					
7					

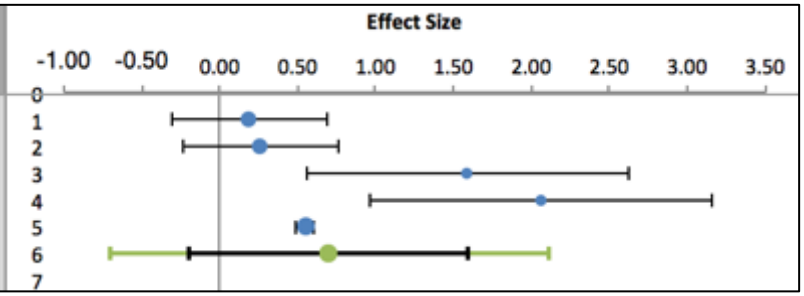


Figure 9: Forest Plot presenting effects of training on 'Self-efficacy' outcome variable.

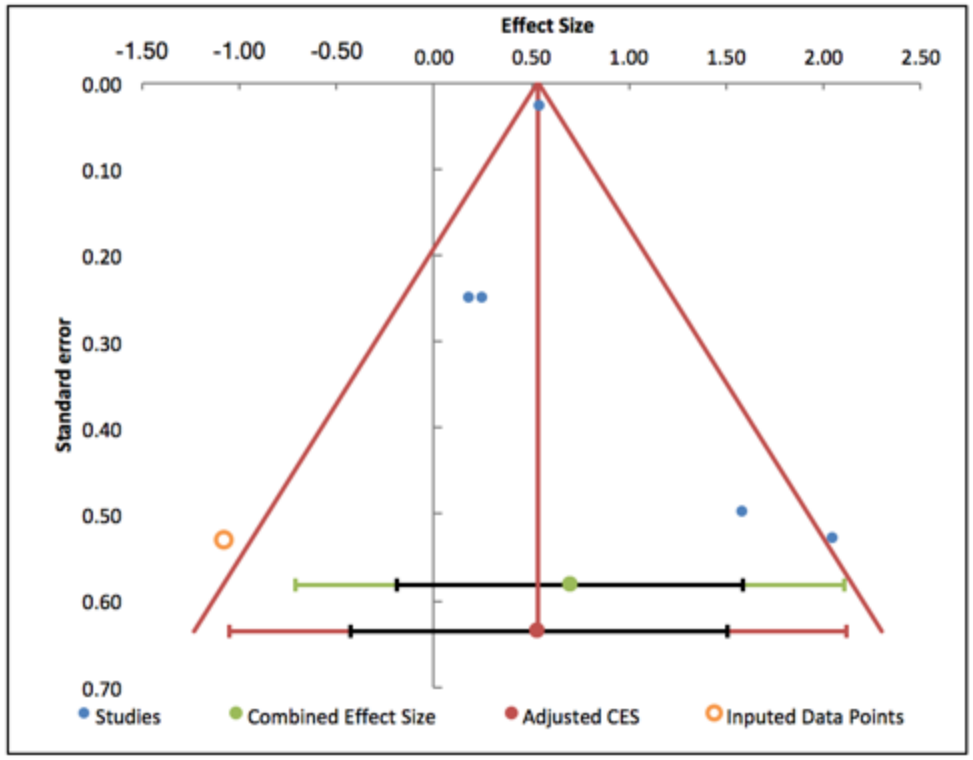


Figure 10: Funnel Plot for 'Self-efficacy' outcome variable meta-analysis

Behaviour. Eight studies evaluated the impact of training on clinicians' use of exposure therapy. Effect sizes ranged from d = -0.33 (Reid et al., 2017) to d = 0.69 (Farrell et al., 2016). The sample weighted average effect size was, d_+ = -0.02, 95% CI [-0.31, 0.27], p = .853 derived from a total sample size of N = 543 (see Figure 11 for forest plot). This indicates that the training interventions had, on average, no effect on clinicians' use of exposure therapy. The homogeneity statistic was almost significant, $Q(7) = 13.98$, p = .052, indicating that the effect sizes from the primary studies were varied, and the I^2 statistic (49.93%) also indicated moderate heterogeneity. Visual inspection of the funnel plot (Fig. 12) suggested some asymmetry, but Egger's regression was not significant (p = .072), suggesting no publication bias.

#	Study name	Effect size	CI Lower limit	CI Upper limit	Weight
1	Farrell (2016)	0.69	0.11	1.27	10.49%
2	Gega (2007)	-0.23	-0.53	0.07	19.11%
3	Harned (2011) ET OLT	0.43	-0.48	1.34	5.82%
4	Harned (2011) ET OLT+MI	0.07	-0.84	0.98	5.82%
5	Harned (2014) OLT+ME	0.04	-0.46	0.53	12.46%
6	Harned (2014) OLT+ME+LC	0.16	-0.34	0.65	12.46%
7	Kaye (2018)	-0.28	-0.68	0.12	15.49%
8	Reid (2017)	-0.33	-0.65	-0.01	18.35%
9					
10					

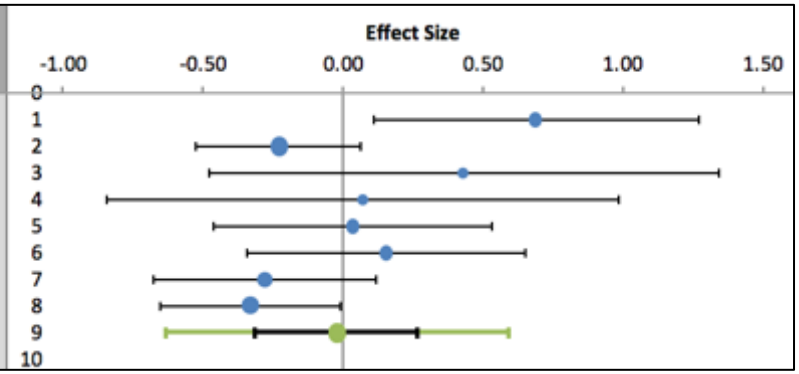


Figure 11: Forest Plot presenting effects of training on 'Behaviour' outcome variable.

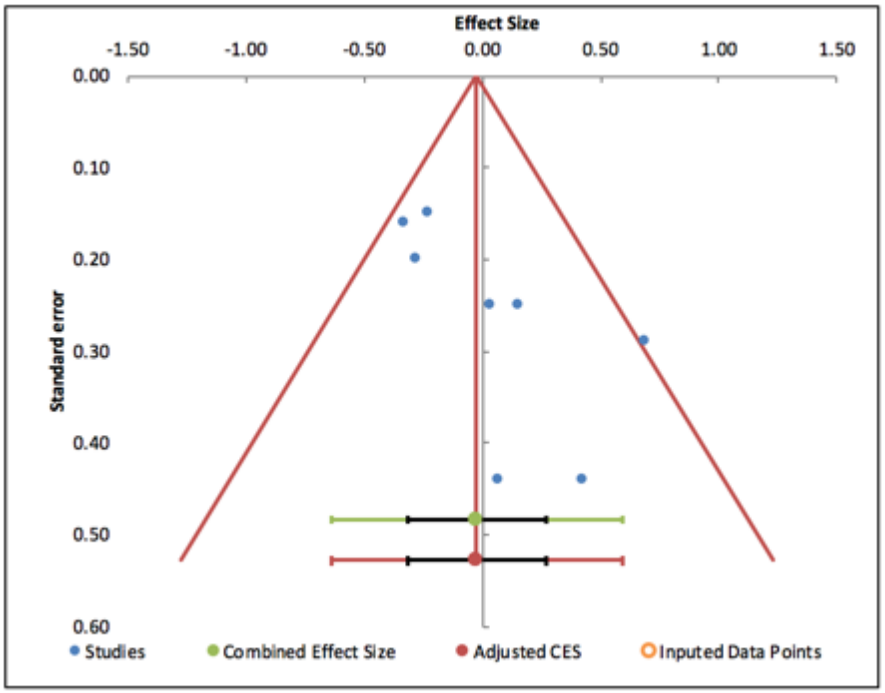


Figure 12: Funnel Plot for 'Behaviour' outcome variable meta-analysis

Moderator analyses

Moderator analyses were conducted to investigate variables that could help explain the heterogeneity within the effect sizes from the primary studies (see Table 5 for summary; see Appendix D for forest plots). An insufficient number of studies examining the impact of training on intentions to use exposure therapy were available to conduct any moderator analyses. However, a sufficient number of studies were available to consider whether the following variables: study quality, study design, nature of the comparison condition, attitude outcome measure, and behaviour outcome measure moderated the effect of training on knowledge, attitude, self-efficacy, and behaviour.

Moderators of the effect of training on knowledge.

Study quality. Four studies evaluating the effect of training on knowledge were deemed 'fair' quality and five studies were deemed 'good'. Studies rated as fair and good both typically reported large-sized effects, $d_+ = 0.89$, (95% CI [-1.03; 2.81]), and $d_+ = 1.19$ (95% CI [-0.41; 2.79]), respectively. However, the variance between subgroups was not significant, $Q\text{b}(1) = 0.15$, $p = .697$, suggesting that the quality of the study did not explain a significant amount of the heterogeneity found within the effect sizes from the primary studies.

Study design. Seven studies evaluating the effect of training on knowledge used an independent-groups design and two used a repeated-measures design. Studies with a repeated-measures design typically reported a large-sized effect of training on knowledge, $d_+ = 1.95$, (95% CI [-6.68; 10.59]), whereas studies with an independent-groups design typically reported a medium-to-large-sized effect, $d_+ = 0.68$ (95% CI [-0.33; 1.69]). The variance between subgroups was not significant, $Q\text{b}(1) = 3.01$, $p = .083$, suggesting that the design of the study did not explain a significant amount of the heterogeneity found within the effect sizes from the primary studies.

Nature of the comparison condition. Five studies evaluating the effects of training on knowledge employed an active comparison condition and two employed a passive comparison condition. Studies with active comparison conditions typically reported a small-sized effect of training on knowledge, $d_+ = 0.18$, (95% CI [-0.12; 0.48]), while studies with passive comparison conditions typically reported a large-sized effect, $d_+ = 2.57$ (95% CI [-4.96; 10.09]). The variance between subgroups was significant, $Q\text{b}(1) = 15.72$, $p < .001$, suggesting that the nature of the comparison condition used within the study moderated the heterogeneity found within the primary meta-analysis.

Moderators of the effect of training on attitudes.

Study quality. Four studies evaluating the effect of training on attitudes were deemed 'fair' quality and six were deemed 'good'. Studies rated as fair typically reported medium-to-large-sized effects, $d_+ = 0.78$, (95% CI [-0.11; 1.67]), and studies rated as good typically reported large-sized effects, $d_+ = 0.89$ (95% CI [0.41; 1.38]). The variance between subgroups was not significant, $Q\text{b}(1) = 0.06$, $p = .804$, suggesting that the quality of the study did not explain a significant amount of the heterogeneity found within the effect sizes from the primary studies.

Study design. Five studies evaluating the effect of training on attitudes used independent-group designs and six used repeated-measures designs. Studies with a repeated-measures design typically reported a large-sized effect of training on attitudes, $d_+ = 1.04$, (95% CI [0.48; 1.59]), while studies with an independent-group design typically reported a medium-sized effect, $d_+ = 0.57$ (95% CI [0.30; 0.85]). However, the variance between subgroups was not significant, $Q\text{b}(1) = 1.64$, $p = .200$, suggesting that the design of the study did not explain a significant amount of the heterogeneity found within the effect sizes from the primary studies.

Nature of the comparison condition. Three studies evaluated the effect of training on attitudes against active comparison conditions and two studies against passive comparison conditions. Studies with active comparison conditions typically reported a medium-sized effect of training on attitudes, $d_+ = 0.51$, (95% CI [0.04; 0.98]), while studies with passive comparison conditions typically reported a large-sized effect, $d_+ = 0.84$ (95% CI [-0.49; 2.18]). However, the variance between subgroups was not significant, $Q\text{b}(1) = 0.86$, $p = .355$, suggesting that the nature of the comparison condition used within the study did not explain a significant amount of the heterogeneity found within the effect sizes from the primary studies.

Nature of the attitude measure. Four studies used the ATETS to measure clinicians' attitudes towards exposure therapy and five studies used the TBES. Studies using the ATETS typically reported a medium-sized effect of training on attitudes, $d_+ = 0.52$, (95% CI [0.18; 0.87]), while studies using the TBES typically reported a large-sized effect, $d_+ = 1.16$ (95% CI [0.66; 1.65]). The variance between subgroups was significant, $Q\text{b}(1) = 6.94$, $p = .008$, suggesting that the type of attitude outcome measure moderated the heterogeneity found within the primary meta-analysis.

Moderators of the effect of training on self-efficacy.

Nature of the comparison condition. Two studies evaluated the effect of training on self-efficacy against active comparison conditions and two studies evaluated the effect of training against passive comparison conditions. Studies with active comparison conditions typically reported a small-sized effect, $d_+ = 0.23$, (95% CI [-0.22; 0.67]), while studies with passive comparison conditions typically reported a large-sized effect, $d_+ = 1.81$ (95% CI [-1.17; 4.79]). The variance between subgroups was significant, $Q\text{b}(1) = 15.39$, $p < .001$, suggesting that the nature of the comparison condition used within the study explains some of the heterogeneity in the effect sizes found within the primary studies.

Moderators of the effect of training on behaviour.

Study quality. Three studies evaluating the effect of training on behaviour were deemed 'fair' quality and five were deemed 'good'. Studies rated as fair and good both typically reported no effects, d_+ = -0.02, (95% CI [-1.33; 1.29]), and, d_+ = -0.02 (95% CI [-0.33; 0.29]), respectively. There was no variance between subgroups, Qb(1) = 0.00, p = .999, suggesting that study quality did not explain the heterogeneity found within the effect sizes from the primary studies.

Nature of the comparison condition. Five studies evaluated the effect of training on behaviour against active comparison conditions and two studies evaluated the effect of training against passive comparison conditions. Studies with active comparison conditions typically reported no effect, d_+ = 0.02, (95% CI [-0.44; 0.48]), and studies with passive comparison conditions typically reported small-sized effects, d_+ = 0.25 (95% CI [-2.04; 2.54]), respectively. The variance between subgroups was not significant, Qb(1) = 0.43, p = .512, suggesting that the nature of the comparison condition did not explain a significant amount of the heterogeneity found within the effect sizes from the primary studies.

Nature of the behaviour outcome measure. Seven studies used self-report measures to assess clinicians' use of exposure therapy and five studies used objective assessment measures. Studies using self-report and objective assessment measures both typically reported no effect of training on behaviour, d_+ = 0.01, (95% CI [-0.32; 0.34]), and, d_+ = -0.15 (95% CI [-0.50; 0.19]), respectively. The variance between subgroups was not significant, Qb(1) = 0.82, p = .365, suggesting that the nature of the outcome measure did not explain a significant amount of the heterogeneity found within the effect sizes from the primary studies.

Table 5: Summary of moderator analyses

Moderator	Outcome	Subgroups	*N*	*k*	*d* (SE)	95% CI	*Q* (*p*)	I^2 (%)	*Q*b (*p*)
Study quality									
	Knowledge	Fair	317	4	0.89 (0.60)	-1.03; 2.81	75.63 (*<.001*)	96.03	-
		Good	290	5	1.19 (0.58)	-0.41; 2.79	49.15 (*<.001*)	91.86	-
									0.15 (*.697*)
	Attitudes	Fair	1196	4	0.78 (0.28)	-0.11; 1.67	144.65 (*<.001*)	97.93	-
		Good	324	6	0.89 (0.19)	0.41; 1.38	22.76 (*<.001*)	78.04	-
									0.06 (*.804*)
	Behaviour	Fair	260	3	-0.02 (0.30)	-1.33; 1.29	9.94 (*.007*)	79.87	-
		Good	283	5	-0.02 (0.11)	-0.33; 0.29	3.32 (*.506*)	0.00	-
									0.00 (*.999*)
Study design									
	Knowledge	Repeated-measures	169	2	1.95 (0.68)	-6.68; 10.59	23.29 *(<.001*)	95.71	-
		Independent-groups	438	7	0.68 (0.41)	-0.33; 1.69	36.27 *(<.001*)	83.46	-
									3.01 (*.083*)
	Attitudes	Repeated-measures	1325	6	1.04 (0.22)	0.48; 1.59	239.66 (*<.001*)	97.91	-
		Independent-groups	240	5	0.57 (0.10)	0.30; 0.85	1.97 (*.742*)	0.00	-
									1.64 (*.200*)
Comparison condition									
	Knowledge	Active	392	5	0.18 (0.11)	-0.12; 0.48	4.16 (*.385*)	3.79	-
		Passive	46	2	2.57 (0.59)	-4.96; 10.09	2.01 (*.156*)	50.33	-
									15.72 (*<.001*)
	Attitudes	Active	194	3	0.51 (0.11)	0.04; 0.98	1.01 (*.605*)	0.00	-
		Passive	46	2	0.84 (0.10)	-0.49; 2.18	0.11 (*.744*)	0.00	-
									0.86 (*.355*)
	Self-efficacy	Active	145	2	0.23 (0.04)	-0.22; 0.67	0.04 (*.843*)	0.00	-

		Passive	46	2	1.81 (0.23)	-1.17; 4.79	0.42 (*.519*)	0.00	- 15.39 *(<.001*)
	Behaviour	Active	455	5	0.02 (0.17)	-0.44; 0.48	10.06 (*.039*)	60.24	-
		Passive	46	2	0.25 (0.18)	-2.04; 2.54	0.33 (*.563*)	0.00	- 0.43 (*.512*)
Outcome measure									
	Attitudes	TBES	386	5	1.16 (0.18)	0.66; 1.65	22.26 (< *.001*)	82.03	-
		ATETS	191	4	0.52 (0.11)	0.18; 0.87	1.43 (*.699*)	0.00	- 6.94 (*.008*)
	Behaviour	Self-reported	381	7	0.01 (0.13)	-0.32; 0.34	11.04 (*.087*)	45.63	-
		Assessed	447	5	-0.15 (0.12)	-0.50; 0.19	7.36 (*.118*)	45.68	- 0.82 (*.365*)

Note: N = total number of participants included in subgroup analysis; k = number of studies included in subgroup analysis; SE = standard error; CI = confidence interval; TBES = Therapist Belief about Exposure Scale; ATETS = Attitudes Towards Exposure Therapy Scale; Qb = variance between subgroups

Discussion

This is the first review to examine the effectiveness of training clinicians in exposure therapy for anxiety-based disorders. Following a systematic literature search, 14 studies were identified that examined the effects of training on clinicians' knowledge, attitudes, intentions, self-efficacy, and use of exposure therapy. The findings of a series of meta-analyses suggested that training had large-sized effects on clinicians' knowledge ($d_+ = 1.05$) and attitudes towards exposure therapy ($d_+ = 0.87$), and medium-sized effects on clinicians' intentions ($d_+ = 0.53$) and self-efficacy associated with using exposure therapy ($d_+ = 0.70$). However, no effect of training was found on clinicians' behaviour ($d_+ = -0.02$). Taken together, these findings suggest that, although clinicians had the knowledge, confidence and intentions to use exposure therapy following training, they did not transform their intentions into action.

This 'intention-behaviour' gap is well-evidenced across many areas (e.g., health and educational goals), with evidence suggesting that intentions only account for around 28% of the variance in behaviour (Sheeran, 2002). Similarly, Webb and Sheeran's (2006) meta-analysis showed that a medium-to-large-sized change in intention leads only to a small-to-medium-sized change in behaviour, suggesting that positive effects of interventions on intentions may not translate into changes in behaviour. Findings of the current meta-analyses suggest that, in addition to the previously identified barriers to implementing exposure therapy, which are typically motivational (e.g., a lack of training, knowledge, and confidence, along with anxiety and negative attitudes; Waller & Turner, 2016), clinicians' lack of use of exposure therapy may be a volitional problem.

Implications for theory and clinical practice

The findings of the current meta-analyses suggest that training clinicians in exposure therapy goes some way to toward reducing the gap between evidence (i.e., that exposure

therapy is effective to treat anxiety) and practice (i.e., therapist drift in the use of exposure in treating clients with anxiety disorders). However, the current findings also indicate that, while training has a positive effect on knowledge, attitudes, intentions, and self-efficacy, training alone may not be enough to help clinicians to change their behaviour and utilise exposure therapy in practice (i.e., there was no effect of training on behaviour). Consequently, further volitional behaviour change strategies may need to be integrated into training in order to bridge this intention-behaviour gap (Sheeran, 2002).

Gollwitzer (1993; 1999) proposed that to increase the likelihood of goal attainment, intentions need to supplemented by planning. Gollwitzer developed the idea of 'implementation intentions' ('if-then' plans), which specify when, where and how individuals will strive towards goals. Supplementing goal intentions with implementation intentions has been found to promote goal realisation more than the strength of intentions alone (Gollwitzer & Sheeran, 2006; Toli, Webb, & Hardy, 2016). In light of the current findings, prompting clinicians to form implementation intentions might be a relatively easy and effective way to help clinicians to translate their intentions into actions following a training intervention. For example, as part of training, clinicians might be prompted to form 'if-then' plans to implement the exposure-based techniques that they intend to use.

Supervision might also be a useful resource to help clinicians to translate changes in motivation following training into action. Specifically, evidence suggests that in order to achieve a desired outcome, related behaviours must be planned, initiated, maintained, and restarted when setbacks occur, until it is habitual (Abraham & Sheeran, 2000; Bagozzi & Edwards, 2000; Bandura, 1997; Sniehotta, Scholz, & Schwarzer, 2005). Supervisors could support clinicians following training by providing opportunities for reminders, and problem-solving during setbacks.

Is it possible to explain variability in the effects of training?

Moderator analyses were conducted to investigate variation in the effects of training. Two factors were found to play a role in determining the apparent effect of training on outcomes.

Nature of the comparison condition. Significant differences were found in the effects of training on knowledge and self-efficacy (but not attitudes or behaviour) between studies comparing the effect of training against an active comparison condition and a passive comparison condition. As might be expected, training typically had a smaller effect on clinicians' knowledge and self-efficacy when compared against an active comparison condition than when compared against a passive comparison condition. The smaller effect size found for studies using active comparison conditions might suggest that clinicians benefit only slightly more from training when additional elements are added, such as an additional focus on negative attitudes, or exposure tasks for clinicians. However, given the evidence for the benefit of exposure therapy training, it may be more ethical to use an active comparison condition to provide all participants with a form of exposure therapy training, whilst using the experimental design to explore the effectiveness of specific elements of training (i.e., a volitional element; Danaher, & Seeley, 2009).

Nature of the outcome measure. The type of outcome measure used to measure attitudes was found to explain some of the heterogeneity found in the primary meta-analysis. Specifically, studies using the ATETS typically reported a medium-sized effect of training on clinician's attitudes towards exposure therapy, whereas those using the TBES typically reported a large-sized effect. The TBES is a more widely used and well-established questionnaire than the ATETS, which may explain the difference in results. Although the ATETS has been shown to have strong internal consistency (Harned et al., 2011), it has only been used in two studies to date (Harned et al., 2011; 2014). In contrast, the TBES has been shown to have excellent internal consistency, high test-retest reliability, and has been used in

several studies outside of those included in this review (e.g., Farrell, Deacon, Kemp, Dixon, Sy, 2013; Schumacher, Schopka, Heinrich, Knaevelsrud, 2019; Whiteside, Deacon, Benito, Stewart, 2016). Therefore, the TBES might provide a more sensitive, reliable and valid estimate of the effects of training on attitudes.

Strengths and limitations

Strengths of the present review include a comprehensive and systematic search of four databases, which allowed a large number of studies to be identified and screened. Unpublished studies were also searched for and included, and statistical analyses indicated that there was no publication bias within each primary meta-analysis. Publication bias is present in other areas of research and can impact on the ability to accurately synthesise and describe the evidence in a given area (DeVito, & Goldacre, 2019). Thus, the current finding indicating no publication bias allows for greater confidence in the estimates of the effect of training on outcomes provided by the current review. Although it was decided to only include studies that could be accessed in English, no studies were excluded on the basis of language. The meta-analytic methodology allowed for the average magnitude of the effect of exposure therapy training on each outcome variable to be aggregated, which is more robust and generalisable than data from individual studies (Lipsey & Wilson, 2001). Another key strength is the use of a theoretical framework (namely, the TPB) to organise the measures used to assess the effects of training and to compare the effects of training on beliefs and behaviour. Indeed, the differential impact of training on the different outcomes and the theoretical and clinical implications of this attests to the value of this approach. Furthermore, studies recruited a wide range of clinicians from a variety of clinically-relevant and ecologically valid settings, which increases the generalisability of the findings.

It is also important to recognise some potential limitations, however. First, the literature search and selection and extraction of data were only conducted by one researcher. Second-

coding these decisions would have reduced the risk of bias, and potentially increased the validity and reliability of the search and selection procedures (Sampson et al., 2009). Second, this review is limited by the relatively small number of studies (namely, 14) that have investigated the effects of exposure therapy training, despite attempts to broaden inclusion criteria, such as including studies with varied interventions (e.g., didactic workshops, online training). This also limited the number of studies that could be included in the moderator analyses, and thus the results should be interpreted with caution. In particular, only two studies with passive comparison conditions were included when investigating the nature of the comparison condition as a moderator variable. It is also worth noting that most studies had a relatively small sample size, which limits their statistical power. Coyne, Thombs, and Hagedoorn (2010) suggest that small, underpowered trials may overestimate the effect size and are often susceptible to methodological issues. Coyne et al. (2010) propose excluding such trials as meta-analyses cannot control for such biases. Future research should therefore aim to fulfil the need for adequately powered studies.

The quality appraisal showed that the primary studies varied in quality. Studies with repeated-measures designs scored lower, in part as a consequence of the lack of a comparison condition and randomisation (Downs & Black 1998). The use of self-report measures of behaviour in most studies might also have threatened the validity of the studies and increased the risk of social desirability bias. However, subgroup analyses showed that there was no significant difference on the effect of training on behaviour between studies using self-report measures and those assessing clinician's behaviour more objectively. There was a lack of blinding in most studies, and objective measures taken by researchers who are blind to group allocation would have reduced the potential impact of demand effects. Due to the limited number of studies, subgroup analyses were not able assess the moderating effect of all aspects of the study design within all primary meta-analyses. Consequently, it is not known

how other methodological features may have impacted on results. Finally, in order to include studies that only measured outcomes at the post-training time-point, the current review used post-training scores to estimate effect sizes. However, this method increases the risk of selection bias in the review compared to using both pre- and post-training scores (Morris & DeShon, 2002). Additionally, post-training scores do not include the statistical adjustments made to control for bias, such as intent-to-treat analysis.

Recommendations for future research

The findings of the present review suggest that exposure therapy is effective, but that clinicians still require support to translate their intentions into action. As such, future research could investigate the use of volitional interventions, such as implementation intentions, to help bridge this gap. Indeed, implementation intentions interventions have been found to be beneficial in increasing clinicians' use of specific clinical procedures (Casper, 2008), and in reducing stress and increasing work engagement (Gollwitzer, Mayer, Frick, & Oettingen, 2018). Moreover, future research could explore the role of supervision in providing ongoing support for clinicians following exposure therapy training, and whether this helps clinicians translate their intentions to use exposure therapy into practice.

It is also worth noting that none of the studies identified as suitable for inclusion in the meta-analyses examined the effect of training on subjective norms – namely, beliefs about whether important others would approve or disapprove of the clinician using exposure therapy. Subjective norms are a key aspect of the TPB, and therefore future research could include subjective norm measures to assess the impact of training clinicians in exposure therapy on norms and the impact of normative beliefs on behaviour. With this in mind, it might also be valuable to expand the consideration of subjective – or injunctive – norms, to also consider descriptive norms, which reflect what important others actually do (Cialdini, Reno, & Kallgren, 1990; Sheeran, & Orbell, 1999). For example, although clinicians might

believe that others would approve of their using exposure therapy, they may also see that relatively few of their colleagues do so, with the consequence that this undermines their good intentions.

Finally, most studies included in the review were rated ‘fair’ in quality, and thus future high quality research is required to assess the robustness of the current findings. Further research would also allow insight into the mechanisms of training. For example, RCTs comparing active training interventions could help to investigate which elements of exposure therapy training are more effective.

Conclusion

The findings of this review indicate that training in exposure therapy typically improves clinician’s knowledge, attitudes, intentions, and self-efficacy with respect to using the technique. However, training did not typically have an impact upon clinician’s use of exposure therapy in practice. These findings are in line with research indicating the intentions-behaviour gap, and indicate the need for volitional interventions to bridge this gap, such as implementation intentions and supervision to support clinicians following exposure therapy training. However, the paucity of available studies and the heterogeneity found within the effect sizes from the primary studies suggest that these findings should be interpreted with caution.

References

References preceded by an asterisk (*) represent studies included in the systematic review.

Abraham, C., & Sheeran, P. (2000). Understanding and changing health behaviour: from health beliefs to self-regulation. In P. Norman, C. Abraham, & M. Conner (Eds.), *Understanding and changing health behaviour* (pp. 3–24). Amsterdam, The Netherlands: Harwood.

Abramowitz, J. S., Deacon, B. J., & Whiteside, S. P. H. (2011). *Exposure therapy for anxiety: Principles and practice.* New York, NY: Guilford Press.

Adams, T. G., Brady, R. E., Lohr, J. M., & Jacobs, W. J. (2015). A meta-analysis of CBT components for anxiety disorders. In B. Deacon (Ed.) *The behaviour therapist.* New York, NY: Association for Behavioral and Cognitive Therapies.

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.

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the Theory of Planned Behavior. *Journal of Applied Social Psychology, 32,* 665-683. doi: 10.1111/j.1559-1816.2002.tb00236.x.

Bagozzi, R. P., & Edwards, E. A. (2000). Goal setting and goal pursuit in the regulation of body weight. In P. Norman, C. Abraham, & M. Conner (Eds.), *Understanding and changing health behaviour* (pp. 261–297). Amsterdam, The Netherlands: Harwood

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-215. doi: 10.1037/0033-295X.84.2.191.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44,* 1175-1184. doi: 10.1037/0003-066X.44.9.1175.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Becker, C.B., Zayfert, C., & Anderson, E. (2004). A survey of psychologists' attitudes towards and utilization of exposure therapy for PTSD. *Behaviour Research and Therapy, 42,* 277–292. doi: 10.1016/S0005-7967(03)00138-4.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons.

Brown, A. M., Deacon, B. J., Abramowitz, J. S., Dammann, J., & Whiteside, S. P. (2007). Parents' perceptions of pharmacological and cognitive-behavioral treatments for childhood anxiety disorders. *Behaviour Research and Therapy, 45,* 819-828. doi: 10.1016/j.brat.2006.04.010.

Card, N. A. (2012). *Applied meta-analysis for social science research.* New York, NY: The Guildford Press.

Casper, E. S. (2008). Using implementation intentions to teach practitioners: changing practice behaviors via continuing education. *Psychiatric Services, 59,* 747-752. doi: 10.1176/appi.ps.59.7.747.

*Chin, E. G., Bernecker, S., Buchanan, E. M., Cunningham, S., Schumacher, J. A., & Coffey, S. F. (2017). A novel dissemination effort for prolonged exposure: practice and dissemination curriculum. *Psychological Trauma: Theory, Research, Practice, and Policy, 11*, 314-318. doi: 10.1037/tra0000362.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology, 58*, 1015-1026. doi: 10.1037/0022-3514.58.6.1015.

Cochrane Consumers and Communication Review Group. (2016). Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Group reviews: planning the analysis at protocol stage. Available from http://cccrg.cochrane.org.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159. doi: 10.1037/0033-2909.112.1.155.

Cooper, H. M. (1986). *Integrating research: A guide for literature reviews*. London, UK: Sage.

Cox, B. J., Fergus, K. D., & Swinson, R. P. (1994). Patient satisfaction with behavioral treatments for panic disorder with agoraphobia. *Journal of Anxiety Disorders, 8*, 193-206. doi: 10.1016/0887-6185(94)90001-9.

Coyne, J. C., Thombs, B. D., & Hagedoorn, M. (2010). Ain't necessarily so: review and critique of recent meta-analyses of behavioral medicine interventions in health psychology. *Health Psychology, 29*, 107–116. doi: 10.1037/a0017633.

Danaher, B. G., & Seeley, J. R. (2009). Methodological issues in research on web-based behavioral interventions. *Annals of Behavioral Medicine, 38*, 28-39. doi: 10.1007/s12160-009-9129-0.

Deacon, B. J., & Abramowitz, J. S. (2004). Cognitive and behavioral treatments for anxiety disorders: a review of meta-analytic findings. *Journal of Clinical Psychology, 60,* 429-441. doi: 10.1002/jclp.10255.

*Deacon, B. J., Farrell, N. R., Kemp, J. J., Dixon, L. J., Sy, J. T., Zhang, A. R., & McGrath, P. B. (2013). Assessing therapist reservations about exposure therapy for anxiety disorders: the Therapist Beliefs about Exposure Scale. *Journal of Anxiety Disorders, 27,* 772-780. doi: 10.1016/j.janxdis.2013.04.006.

Deacon, B. J., Lickel, J. J., Farrell, N. R., Kemp, J. J., & Hipol, L. J. (2013). Therapist perceptions and delivery of interoceptive exposure for panic disorder. *Journal of Anxiety Disorders, 27*, 259-264. doi: 10.1016/j.janxdis.2013.02.004.

DeVito, N. J., & Goldacre, B. (2019). Catalogue of bias: publication bias. *British Medical Journal Evidence-Based Medicine*, *24*, 53–54. doi: 10.1136/bmjebm-2018-111107.

Dimeff, L. A., Koerner, K., Woodcock, E. A., Beadnell, B., Brown, M. Z., Skutch, J. M., Paves, A. P., Bazinet, A., & Harned, M. S. (2009). Which training method works best? A randomized controlled trial comparing three methods of training clinicians in dialectical behavior therapy skills. *Behaviour Research and Therapy, 47,* 921-930. doi: 10.1016/j.brat.2009.07.011.

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions. *Journal of Epidemiology and Community Health, 52*, 377-384. doi:10.1136/jech.52.6.377.

Egger, M., Davey., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-634. doi:10.1136/bmj.315.7109.629.

Farrell, N. R., Deacon, B. J., Dixon, L. J., & Lickel, J. J. (2013). Theory-based training strategies for modifying practitioner concerns about exposure therapy. *Journal of Anxiety Disorders, 27*, 781-787. doi: 10.1016/j.janxdis.2013.09.003.

Farrell, N. R., Deacon, B. J., Kemp, J. J., Dixon, L. J., Sy, J. T. (2013). Do negative beliefs about exposure therapy cause its suboptimal delivery? An experimental investigation. *Journal of Anxiety Disorders, 27,* 763-771. doi: 10.1016/j.janxdis.2013.03.007.

*Farrell, N. R., Kemp, J. J., Blakey, S. M., Meyer, J. M., & Deacon, B. J. (2016). Targeting clinician concerns about exposure therapy: a pilot study comparing standard vs. enhanced training. *Behaviour Research and Therapy, 85,* 53-59. doi: 10.1016/j.brat.2016.08.011.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *63*, 665-694. doi: 10.1348/000711010X502733.

Freiheit, S.R., Vye, C., Swan, R., & Cady, M. (2004). Cognitive-behavioral therapy for anxiety: is dissemination working? *The Behavior Therapist, 27,* 25–32.

*Gega, L., Norman, I. J., & Marks, I. M. (2007). Computer-aided vs. tutor-delivered teaching of exposure therapy for phobia/panic: randomized controlled trial with pre-registration nursing students. *International Journal of Nursing Studies, 44,* 397-405. doi: 10.1016/j.ijnurstu.2006.02.009.

Godin, G., Bélanger-Gravel, A., Eccles, M., & Grimshaw, J. (2008). Healthcare professionals' intentions and behaviours: a systematic review of studies based on social cognitive theories. *Implementation Science*, *3*, 36-48. doi:10.1186/1748- 5908-3-36.

Gollwitzer, P. M. (1993). Goal achievement: the role of intentions. *European Review of Social Psychology, 4,* 141-185. doi: 10.1080/14792779343000059.

Gollwitzer, P. M. (1999). Implementation intentions: strong effects of simple plans. *American Psychologist, 54,* 493-503. doi: 10.1037/0003-066X.54.7.493.

Gollwitzer, P. M., Mayer, D., Frick, C., & Oettingen, G. (2018). Promoting the self-regulation of stress in health care providers: an internet-based intervention. *Frontiers in Psychology, 15,* 1-11. doi: 10.3389/fpsyg.2018.00838.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: a meta-analysis of effects and processes. *Advances in Experimental Social Psychology, 38,* 69-119. doi: 10.1016/S0065-2601(06)38002-1.

Gunter, R. W., & Whittal, M. L. (2010). Dissemination of cognitive-behavioral treatments for anxiety disorders: overcoming barriers and improving patient access. *Clinical Psychology Review, 30,* 194-202. doi: 10.1016/j.cpr.2009.11.001.

*Harned, M. S., Dimeff, L. A., Woodcock, E. A., & Skutch, J. M. (2011). Overcoming barriers to disseminating exposure therapies for anxiety disorders: a pilot randomized

controlled trial of training methods. *Journal of Anxiety Disorders, 25,* 155-163. doi: 10.1016/j.janxdis.2010.08.015.

*Harned, M. S., Dimeff, L. A., Woodcock, E. A., Kelly, T., Zavertnik, J., Contreras, I., & Danner, S. M. (2014). Exposing clinicians to exposure: a randomized controlled dissemination trial of exposure therapy for anxiety disorders. *Behavior Therapy, 45*, 731–744. doi: 10.1016/j.beth.2014.04.005.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 23,* 557-560. doi: 10.1136/bmj.327.7414.557.

Hipol, L. J., & Deacon, B. J. (2012). Dissemination of evidence-based practices for anxiety disorders in Wyoming: a survey of practicing psychotherapists. *Behavior Modification, 37,* 170-188. doi: 10.1177/0145445512458794.

James, I. A., Blackburn, I-M., Milne, D. L., & Reichfelt, F. K. (2001). Moderators of trainee therapists' competence in cognitive therapy. *British Journal of Clinical Psychology, 40,* 131-141. doi: 10.1348/014466501163580.

*Kaye, J. L. (2018). Targeting therapists' experiential avoidance during exposure delivery: an experimental investigation to improve the quality of exposure-based interventions. (Unpublished doctoral dissertation). Drexel University, Philadelphia, USA.

*Kobak, K. A., Wolitzky-Taylor, K., Craske, M. G., & Rose, R. D. (2017). Therapist training on cognitive behavior therapy for anxiety disorders using internet-based technologies. *Cognitive Therapy and Research, 41,* 252-265. doi:10.1007/s10608-016-9819-4.

Koo, T. M., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*, 155-163. doi: 10.1016/j.jcm.2016.02.012.

Lenhard, W., & Lenhard, A. (2016). *Calculation of effect sizes.* Dettelbach, Germany: Psychometrica. Available at https://www.psychometrica.de/effect_size.html. doi: 10.13140/RG.2.1.3478.4245.

Light, R. & Pillemer, D. (1984). *Summing up: The science of reviewing research*. Cambridge, UK: Harvard University Press.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*McDonough, M., & Marks, I. M. (2002). Teaching medical students exposure therapy for phobia/panic: randomized, controlled comparison of face-to-face tutorial in small groups vs. solo computer instruction. *Medical Education, 36,* 412–417. doi: 10.1046/j.1365-2923.2002.01210.x.

McDonagh, M., Peterson, K., Raina, P., Chang, S., & Shekelle, P. (2013). Avoiding bias in selecting studies. Methods guide for comparative effectiveness reviews. In: *Methods guide for effectiveness and comparative effectiveness reviews.* Rockville, MD: Agency for Healthcare Research and Quality.

Memon, A., Taylor, K., Mohebati, L. M., Sundin, J., Cooper, M., Scanlon, T., de Visser, R. (2016). Perceived barriers to accessing mental health services among black and minority ethnic (BME) communities: a qualitative study in Southeast England. *British Medical Journal Open, 6,* 1-9. doi: 10.1136/bmjopen-2016-012337.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal, 339*, 332-336. doi:10.1136/bmj.b2535.

Morison, L., Trigeorgis, C., & John, M. (2014). Are mental health services inherently feminised? *The Psychologist, 27,* 414-416. Retrieved from: https://thepsychologist.bps.org.uk/volume-27/edition-6/are-mental-health-services-inherently-feminised

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105-125. doi: 10.1037/1082-989X.7.1.105.

National Institute for Health and Care Excellence. (2005). *Obsessive Compulsive Disorder and Body Dysmorphic Disorder: treatment (CG31).* Retrieved from https://www.nice.org.uk/guidance/cg31

National Institute for Health and Care Excellence. (2011). *Generalised Anxiety Disorder and Panic Disorder in adults: management (CG113).* Retrieved from https://www.nice.org.uk/guidance/cg113

National Institute for Health and Care Excellence. (2013). *Social Anxiety Disorder: recognition, assessment, and treatment (CG159).* Retrieved from https://www.nice.org.uk/guidance/cg159

National Institute for Health and Care Excellence. (2018). *Post-traumatic Stress Disorder (NG116).* Retrieved from https://www.nice.org.uk/guidance/ng116

O'Connor, S. R., Tully, M. A., Ryan, B., Bradley, J. M., Baxter, G. D., & McDonough. (2015). Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. *BioMed Central Research Notes, 8*, 224-231. doi:10.1186/s13104-015-1181-1.

Olatunji, B. O., Deacon, B. J., & Abramowitz, J. S. (2009). The cruelest cure? Ethical issues in the implementation of exposure-based treatments. *Cognitive and Behavioral Practice, 16*, 172-180. doi: 10.1016/j.cbpra.2008.07.003.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157-159. doi: 10.2307/1164923.

Ougrin, D. (2011). Efficacy of exposure versus cognitive therapy in anxiety disorders: systematic review and meta-analysis. *BMC Psychiatry, 11,* 200-212. doi: 10.1186/1471-244X-11-200.

Pittig, A., Kotter, R., & Hoyer, J. (2019). The struggle of behavioral therapists with exposure: self-reported practicability, negative beliefs, and therapist distress about exposure-based interventions. *Behavior Therapy, 50,* 353-366. doi: 10.1016/j.beth.2018.07.003.

*Reid, A. M., Guzick, A. G., Balkhi, A. M., McBride, M., Geffken, G. R., & McNamara, J. P. H. (2017). The Progressive Cascading Model improves exposure delivery in trainee therapists learning exposure therapy for obsessive-compulsive disorder. *Training and Education in Professional Psychology, 11*, 260–265. doi: 10.1037/tep0000159.

Richard, D. C. S., & Gloster, A. T. (2007). Exposure therapy has a public relations problem: a dearth of litigation amid a wealth of concern. In D. C. S. Richard, & D. Lauterbach (Eds.), *Comprehensive handbook of the exposure therapies* (pp. 409–425). New York, NY: Academic Press.

Richard, D. C. S., & Lauterbach, D. (2007). *Handbook of exposure therapies*. Burlington, MA: Elsevier.

Rosqvist, J. (2005). *Exposure treatments for anxiety disorders: A practitioner's guide to concepts, methods, and evidence-based practice.* New York, NY: Routledge.

*Ruzek, J. I., Eftekhari, A., Rosen, C. S., Crowley, J. J., Kuhn, E., Foa, E. B., Hembree, E. A., & Karlin, B. E. (2016). Effects of a comprehensive training program on clinician beliefs about and intention to use prolonged exposure therapy for PTSD. *Psychological Trauma: Theory, Research, Practice, and Policy*, *8,* 348–355. doi: 10.1037/tra0000004.

Sampson, M., McGowan, J., Cogo, E., Grimshaw, J., Moher, D., & Lefebvre, C. (2009). An evidence-based practice guideline for the peer review of electronic search strategies. *Journal of Clinical Epidemiology*, *62*, 944-952. doi: 10.1016/j.jclinepi.2008.10.012.

Schare, M. L., & Wyatt, K. P. (2013). On the evolving nature of exposure therapy. *Behavior Modification, 37,* 243-256. doi: 10.1177/0145445513477421.

Scherr, S. R., Herbert, J. D., & Forman, E. M. (2015). The role of therapist experiential avoidance in predicting therapist preference for exposure treatment for OCD. *Journal of Contextual Behavioral Science, 4*, 21-29. doi: 10.1016/j.jcbs.2014.12.002.

Schumacher, S., Gaudlitz, K., Plag, J., Miller, R., Kirschbaum, C., Fehm, L., Fydrich, T., & Ströhle, A. (2014). Who is stressed? A pilot study of salivary cortisol and alpha amylase concentrations in agoraphobic patients and their novice therapists undergoing in vivo exposure. *Psychoneuroendocrinology, 49*, 280–289. doi: 10.1016/j.psyneuen.2014.07.016.

Schumacher, S., Miller, R., Fehm, L., Kirschbaum, C., Fydrich,T., & Ströhle, A. (2015). Therapists' and patients' stress responses during graduated versus flooding in vivo exposure in the treatment of specific phobia: a preliminary observational study. *Psychiatry Research, 230*, 668–675. doi: 10.1016/j.psychres.2015.10.020.

Schumacher, S., Schopka, N. M., Heinrich, M., & Knaevelsrud, C. (2019). The German adaptation of the Therapist Beliefs about Exposure Scale: a validation study among licensed Cognitive Behavioural Therapists in Germany. *Behavioural and Cognitive Psychotherapy, 47,* 164-180. doi: 10.1017/S1352465818000371.

Sheeran, P. (2002). Intention-behavior relations: a conceptual and empirical review. *European Review of Social Psychology, 12*, 1-36. doi: 10.1080/14792772143000003.

Sheeran, P., & Orbell, S. (1999). Augmenting the theory of planned behavior: roles for anticipated regret and descriptive norms. *Journal of Applied Social Psychology, 29*, 2107-2142. doi: 10.1111/j.1559-1816.1999.tb02298.x.

Smith, G. D. (2002). Data dredging, bias, or confounding: they can all get you into the BMJ and the Friday papers. *British Medical Journal, 21,* 1437-1438. doi: 10.1136/bmj.325.7378.1437.

Sniehotta, F. F., Scholz, U., & Schwarzer, R. (2005). Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. *Psychology & Health, 20,* 143-160. doi: 10.1080/08870440512331317670.

Stobie, B., Taylor, T., Quigley, A., Ewing, S., & Salkovskis, P. M. (2007). "Contents may vary": a pilot study of treatment histories of OCD patients. *Behavioural and Cognitive Psychotherapy, 35*, 273-282. doi: 10.1017/S135246580700358X.

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison and validation of Meta-Essentials: a free and simple tool for meta-analysis. *Research Synthesis Methods*, *8*, 537-553. doi: 10.1002/jrsm.1260.

Toli, A., Webb, T. L., & Hardy, G. E. (2016). Does forming implementation intentions help people with mental health problems to achieve goals? A meta-analysis of experimental studies with clinical and analogue samples. *British Journal of Clinical Psychology, 55,* 69-90. doi: 10.1111/bjc.12086.

van Minnen, A., Hendricks, L., & Olff, M. (2010). When do trauma experts choose exposure therapy for PTSD patients? A controlled study of therapist and patient factors. *Behaviour Research and Therapy, 48,* 312-320. doi: 10.1016/j.brat.2009.12.003.

Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour Research and Therapy, 47,* 119-127. doi: 10.1016/j.brat.2008.10.018.

*Waller, G., D'Souza Walsh, K., & Wright, C. (2016). Impact of education on clinicians' attitudes to exposure therapy for eating disorders. *Behaviour Research and Therapy, 76,* 76-80. doi: 10.1016/j.brat.2015.11.014.

Waller, G., & Turner, H. (2016). Therapist drift redux: why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behaviour Research and Therapy, 77,* 128-137. doi: 10.1016/j.brat.2015.12.005.

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin, 132,* 249-268. doi: 10.1037/0033-2909.132.2.249.

Whiteside, S. P. H., Deacon, B. J., Benito, K., & Stewart, E., (2016). Factors associated with practitioners' use of exposure therapy for childhood anxiety disorders. *Journal of Anxiety Disorders, 40,* 29-36. doi: 10.1016/j.janxdis.2016.04.001.

Wolitzky-Taylor, K., Fenwick, K., Lengnick-Hall, R., Grossman, J., Bearman, S. K., Arch, J., Miranda, J., & Chung, B. (2018). A preliminary exploration of the barriers to delivering (and receiving) exposure-based Cognitive Behavioral Therapy for anxiety disorders in adult community mental health settings. *Community Mental Health Journal, 54,* 899-911. doi: 10.1007/s10597-018-0252-x.

Appendix A

Adapted Downs and Black (1998) quality assessment checklist

Study	**Score**			**Notes**
Reporting	**0**	**1**	**2**	
1. Is the hypothesis/aim/objective of the study clearly described?				
2. *Are the main outcomes to be measured clearly described in the Introduction or Methods section?* If the main outcomes are first mentioned in the Results section, the question should be answered no				
3. *Are the characteristics of the participants included in the study clearly described*? In cohort studies and trials, inclusion and/or exclusion criteria should be given. In case-control studies, a case-definition and the source for controls should be given.				
4. *Are the interventions of interest clearly described?* Treatments and placebo (where relevant) that are to be compared should be clearly described.				
5. *Are the distributions of principal confounders in each group of subjects to be compared clearly described?* A list of principal confounders is provided.				
6. *Are the main findings of the study clearly described?* Simple outcome data (including denominators and numerators) should be reported for all major findings so that the reader can check the major analyses and conclusions. (This question does not cover statistical tests which are considered below)				
7. *Does the study provide estimates of the random variability in the data for the main outcomes?* In non-normally distributed data the inter-quartile range of results should be reported. In normally distributed data the standard error, standard deviation or confidence intervals should be reported. If the distribution of the data is not described, it must be assumed that the estimates used were appropriate and the question should be answered yes.				
8. *Have all important adverse events that may be a consequence of the intervention been reported?* This should be answered yes if the study demonstrates that there was a comprehensive attempt to measure adverse events. (A list of possible adverse events is provided).				

9. *Have the characteristics of participants lost to follow-up been described?* This should be answered yes where there were no losses to follow-up or where losses to follow up were so small that findings would be unaffected by their inclusion. This should be answered no, where a study does not report the number of participants lost to follow-up.				
10. *Have actual probability values been reported (e.g. 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?*				
External Validity All the following criteria attempt to address the representativeness of the findings of the study and whether they may be generalised to the population from which the study subjects were derived				
11. *Were the subjects asked to participate in the study representative of the entire population from which they were recruited?* The study must identify the source population for participants and describe how the participants were selected. Participants would be representative if they comprised the entire source population, an unselected sample of consecutive participants, or a random sample. Random sampling is only feasible where a list of all members of the relevant population exists. Where a study does not report the proportion of the source population from which the participants are derived, the question should be answered as unable to determine.				
12. *Were those subjects who were prepared to participate representative of the entire population from which they were recruited?* The proportion of those asked who agreed should be stated. Validation that the sample was representative would include demonstrating that the distribution of the main confounding factors was the same in the study sample and the source population.				
13. *Were the staff, places, and facilities where the participants were treated, representative of the treatment the majority of participants receive?* For the question to be answered yes the study should demonstrate that the intervention was representative of that in use in the source population. The question should be answered no if, for example, the intervention was undertaken in a specialist centre unrepresentative of the hospitals most of the source population would attend.				

Internal validity – bias				
14. *Was an attempt made to blind study subjects to the intervention they have received?* For studies where the participants would have no way of knowing which intervention they received, this should be answered yes				
15. *Was an attempt made to blind those measuring the main outcomes of the intervention?*				
16. *If any of the results of the study were based on "data dredging", was this made clear?* Any analyses that had not been planned at the outset of the study should be clearly indicated. If no retrospective unplanned subgroup analyses were reported, then answer yes.				
17. *In trials and cohort studies, do the analyses adjust for different lengths of follow up of participants, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?* Where follow up was the same for all study participants the answer should yes. If different lengths of follow up were adjusted for by, for example, survival analysis the answer should be yes. Studies where differences in follow-up are ignored should be answered no.				
18. *Were the statistical tests used to assess the main outcomes appropriate?* The statistical techniques used must be appropriate to the data. For example nonparametric methods should be used for small sample sizes. Where little statistical analysis has been undertaken but where there is no evidence of bias, the question should be answered yes. If the distribution of the data (normal or not) is not described it must be assumed that the estimates used were appropriate and the question should be answered yes.				
19. *Was compliance with the intervention/s reliable*? Where there was non compliance with the allocated treatment or where there was contamination of one group, the question should be answered no. For studies where the effect of any misclassification was likely to bias any association to the null, the question should be answered yes.				
20. *Were the main outcome measures used accurate (valid and reliable*)? For studies where the outcome measures are clearly described, the question should be answered yes. For studies which refer to other work or that demonstrates the outcomes				

measures are accurate, the question should be answered as yes.				
Internal validity - confounding (selection bias)				
21. *Were the participants in different intervention groups (trials and cohort studies) or were the cases and controls (case control studies) recruited from the same population?* For example, participants for all comparison groups should be selected from the same hospital. The question should be answered unable to determine for cohort and case control studies where there is no information concerning the source of participants included in the study.				
22. *Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?* For a study which does not specify the time period over which participants were recruited, the question should be answered as unable to determine.				
23. *Were study subjects randomised to intervention groups? Studies* which state that subjects were randomized should be answered yes except where method of randomisation would not ensure random allocation. For example, alternate allocation would score no because it is predictable.				
24. Was the randomised intervention assignment concealed from both participants and health care staff until recruitment was complete and irrevocable? All non-randomized studies should be answered no. If assignment was concealed from participants but not from staff, it should be answered no				
25. *Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?* This question should be answered no for trials if: the main conclusions of the study were based on analyses of treatment rather than intention to treat; the distribution of known confounders in the different treatment groups was not described; or the distribution of known confounders differed between the treatment groups but was not taken into account in the analyses. In non-randomised studies if the effect of the main confounders was not investigated or confounding				

was demonstrated but no adjustment was made in the final analyses the question should be answered as no.				
26. *Were losses of participants to follow-up taken into account?* If the numbers of participants lost to follow up are not reported, the question should be answered as unable to determine. If the proportion lost to followup was too small to affect the main findings, the question should be answered yes.				
Power				
27. Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%? Sample sizes have been calculated to detect a difference of x% and y%				
Total score = /28 = %				

Appendix B

Equation to convert standard error into standard deviation

Standard deviation was converted into standard error using the following equation:

$$SD = SE \times \sqrt{N}$$

SD = standard deviation

SE = standard error

N = number of cases

Appendix C

Quality assessment results summary table

	Item																												
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**Total**	**Quality**
Chin (2017)	1	1	1	1	0	1	1	0	0	1	0	0	1	0	0	0	1	1	0	1	1	0	0	0	1	0	0	13/28	Poor
Deacon (2013)	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0	1	1	1	0	1	1	0	0	0	0	1	0	15/28	Fair
Farrell (2016)	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0	0	1	1	0	1	1	0	0	0	1	1	0	15/28	Fair
Gega (2007)	1	1	1	1	0	0	1	0	1	1	0	0	1	0	1	1	1	1	0	1	1	1	1	0	0	1	1	18/28	Fair
Harned (2011) ET OLT	1	1	1	1	2	1	1	0	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	23/28	Good
Harned (2011) ET OLT+MI	1	1	1	1	2	1	1	0	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	23/28	Good
Harned (2014) OLT+ME	1	1	1	1	2	0	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	23/28	Good
Harned (2014) OLT+ME+LC	1	1	1	1	2	0	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	23/28	Good
Kaye (2018)	1	1	1	1	2	1	1	0	1	1	1	1	0	0	0	1	1	1	0	1	1	1	0	0	1	1	1	20/28	Good
Kobak (2017)	1	1	1	1	1	1	1	0	0	1	1	0	1	0	1	1	1	1	0	1	1	1	0	0	0	0	1	18/28	Fair
McDonough (2002)	1	1	0	1	0	1	1	0	1	1	1	0	1	0	0	1	1	1	0	1	1	1	1	0	0	1	0	17/28	Fair
Reid (2017)	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	0	1	1	1	1	1	0	0	0	0	1	0	16/28	Fair
Ruzek (2016)	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	1	1	1	0	1	1	0	0	0	1	1	0	17/28	Fair
Waller (2016)	1	1	1	1	1	1	1	0	1	0	0	1	1	0	0	1	1	1	0	1	1	1	0	0	1	1	1	19/28	Good

Note: Green = Yes; Orange = Unable to determine or partial (item 5); Red = No

Appendix D

Forest plots for moderator analyses

#	Study name / Subgroup name	Effect Size	CI LL	CI UL	Weight
1	Farrell (2016)	0.04	-0.54	0.62	24.79%
2	Gega (2007)	0.20	-0.12	0.52	25.76%
3	Kobak (2017)	2.65	2.15	3.15	25.14%
4	McDonough (2002)	0.67	-0.02	1.36	24.30%
5	**Fair**	**0.89**	**-1.03**	**2.81**	**47.75%**
6	Harned (2011) ET OLT	2.03	0.93	3.13	17.53%
7	Harned (2011) ET OLT+MI	3.22	1.87	4.57	15.52%
8	Harned (2014) OLT+ME	-0.14	-0.64	0.36	21.91%
9	Harned (2014) OLT+ME+LC	0.29	-0.21	0.79	21.91%
10	Kaye (2018)	1.29	1.03	1.55	23.12%
11	**Good**	**1.19**	**-0.41**	**2.79**	**52.25%**
12	**Combined effect size**	**1.05**	**0.70**	**1.39**	
13					

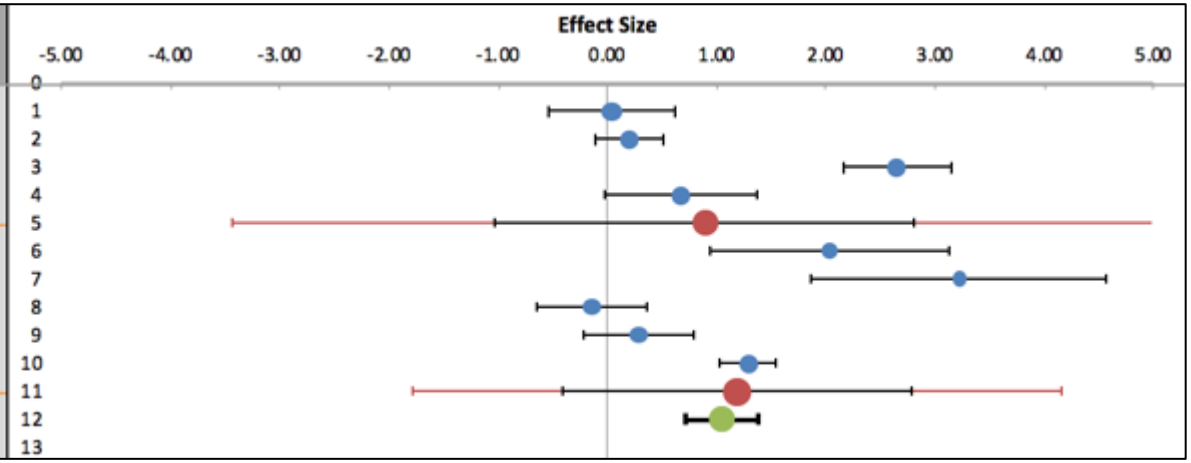


Figure 13: Forest plot for subgroup analysis of study quality within 'knowledge' meta-analysis

#	Study name / Subgroup name	Effect Size	CI LL	CI UL	Weight
1	Farrell (2016)	0.77	0.17	1.37	22.86%
2	Reid (2017)	0.68	0.34	1.02	25.09%
3	Ruzek (2016)	0.17	0.11	0.23	26.26%
4	Deacon (2013)	1.50	1.28	1.72	25.78%
5	**Fair**	**0.78**	**-0.11**	**1.67**	**31.41%**
6	Harned (2011) ET OLT	0.74	-0.19	1.67	12.21%
7	Harned (2011) ET OLT+MI	0.95	0.00	1.90	11.95%
8	Harned (2014) OLT+ME	0.41	-0.09	0.91	18.09%
9	Harned (2014) OLT+ME+LC	0.43	-0.09	0.95	17.77%
10	Kaye (2018)	1.52	1.22	1.82	20.97%
11	Waller (2016)	1.15	0.70	1.60	19.01%
12	**Good**	**0.89**	**0.41**	**1.38**	**68.59%**
13	**Combined effect size**	**0.86**	**0.74**	**0.98**	
14					

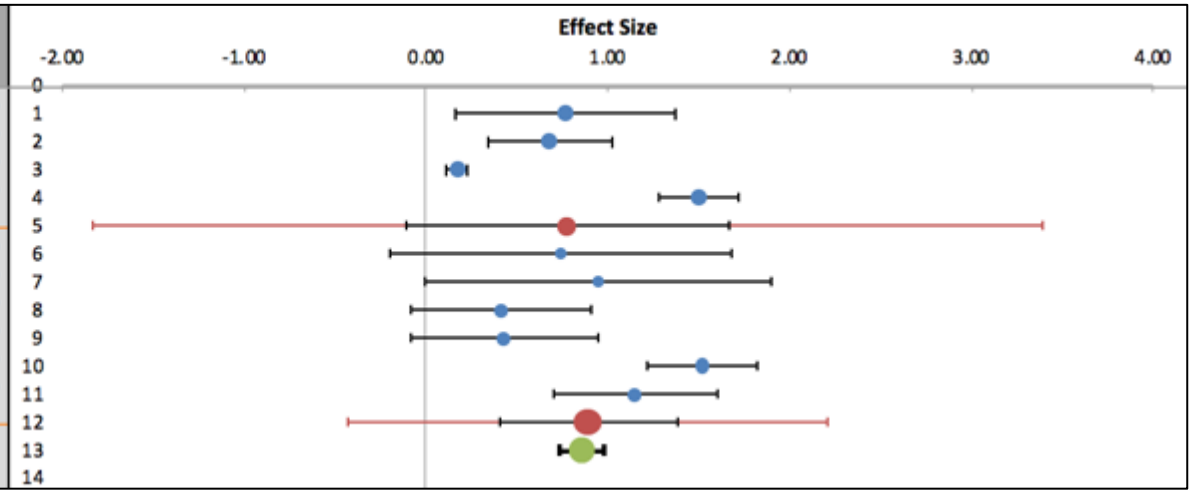


Figure 14: Forest plot for subgroup analysis of study quality within 'attitude' meta-analysis

#	Study name / Subgroup name	Effect Size	CI LL	CI UL	Weight
1	Farrell (2016)	0.69	0.11	1.27	26.75%
2	Gega (2007)	-0.23	-0.53	0.07	36.97%
3	Reid (2017)	-0.33	-0.65	-0.01	36.28%
4	**Fair**	**-0.02**	**-1.33**	**1.29**	**11.69%**
5	Harned (2011) ET OLT	0.43	-0.48	1.34	7.67%
6	Harned (2011) ET OLT+MI	0.07	-0.84	0.98	7.67%
7	Harned (2014) OLT+ME	0.04	-0.46	0.53	23.76%
8	Harned (2014) OLT+ME+LC	0.16	-0.34	0.65	23.76%
9	Kaye (2018)	-0.28	-0.68	0.12	37.13%
10	**Good**	**-0.02**	**-0.33**	**0.29**	**88.31%**
11	**Combined effect size**	**-0.02**	**-0.02**	**-0.02**	
12					

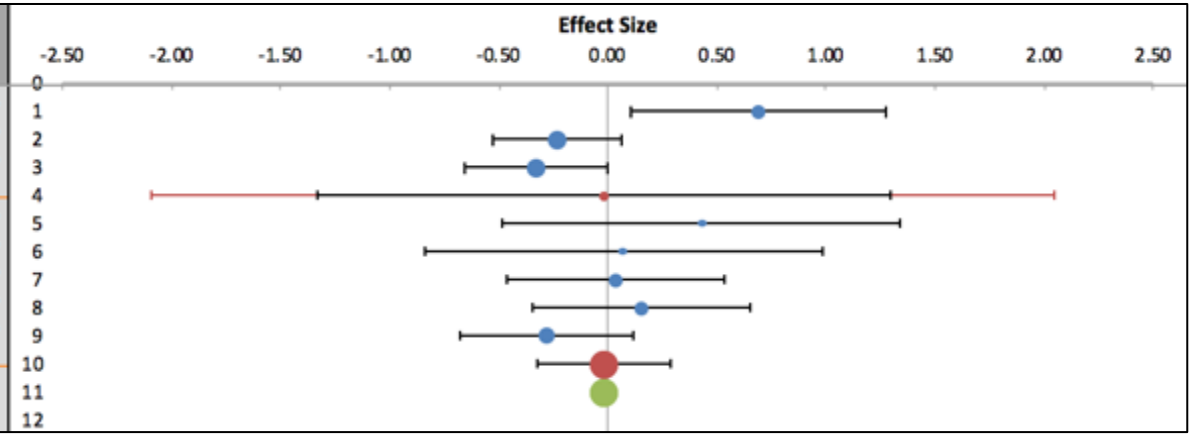


Figure 15: Forest plot for subgroup analysis of study quality within 'behaviour' meta-analysis

#	Study name / Subgroup name	Effect Size	CI LL	CI UL	Weight
1	Farrell (2016)	0.04	-0.54	0.62	15.45%
2	Gega (2007)	0.20	-0.12	0.52	17.53%
3	Harned (2011) ET OLT	2.03	0.93	3.13	11.03%
4	Harned (2011) ET OLT+MI	3.22	1.87	4.57	9.15%
5	Harned (2014) OLT+ME	-0.14	-0.64	0.36	16.16%
6	Harned (2014) OLT+ME+LC	0.29	-0.21	0.79	16.16%
7	McDonough (2002)	0.67	-0.02	1.36	14.52%
8	**Independent-groups**	**0.68**	**-0.33**	**1.69**	**58.99%**
9	Kaye (2018)	1.29	1.03	1.55	51.23%
10	Kobak (2017)	2.65	2.15	3.15	48.77%
11	**Repeated-measures**	**1.95**	**-6.68**	**10.59**	**41.01%**
12	**Combined effect size**	**1.20**	**-0.24**	**2.65**	
13					

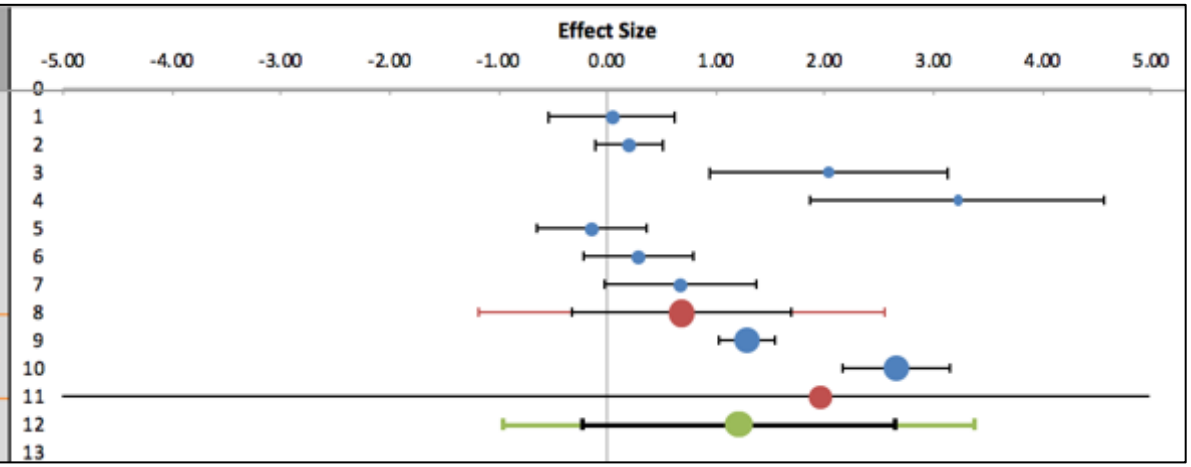


Figure 16: Forest plot for subgroup analysis of study design within the 'knowledge' meta-analysis

#	Study name / Subgroup name	Effect Size	CI LL	CI UL	Weight
1	Farrell (2016)	0.77	0.17	1.37	21.55%
2	Harned (2011) ET OLT	0.74	-0.19	1.67	9.58%
3	Harned (2011) ET OLT+MI	0.95	0.00	1.90	9.16%
4	Harned (2014) OLT+ME	0.41	-0.09	0.91	31.03%
5	Harned (2014) OLT+ME+LC	0.43	-0.09	0.95	28.69%
6	**Independent-groups**	**0.57**	**0.30**	**0.85**	**58.76%**
7	Chin (2017)	1.23	0.80	1.67	16.16%
8	Kaye (2018)	1.52	1.22	1.82	16.76%
9	Reid (2017)	0.68	0.34	1.02	16.60%
10	Ruzek (2016)	0.17	0.11	0.23	17.33%
11	Waller (2016)	1.15	0.70	1.60	16.13%
12	Deacon (2013)	1.50	1.28	1.72	17.03%
13	**Repeated-measures**	**1.04**	**0.48**	**1.59**	**41.24%**
14	**Combined effect size**	**0.77**	**0.26**	**1.27**	
15					

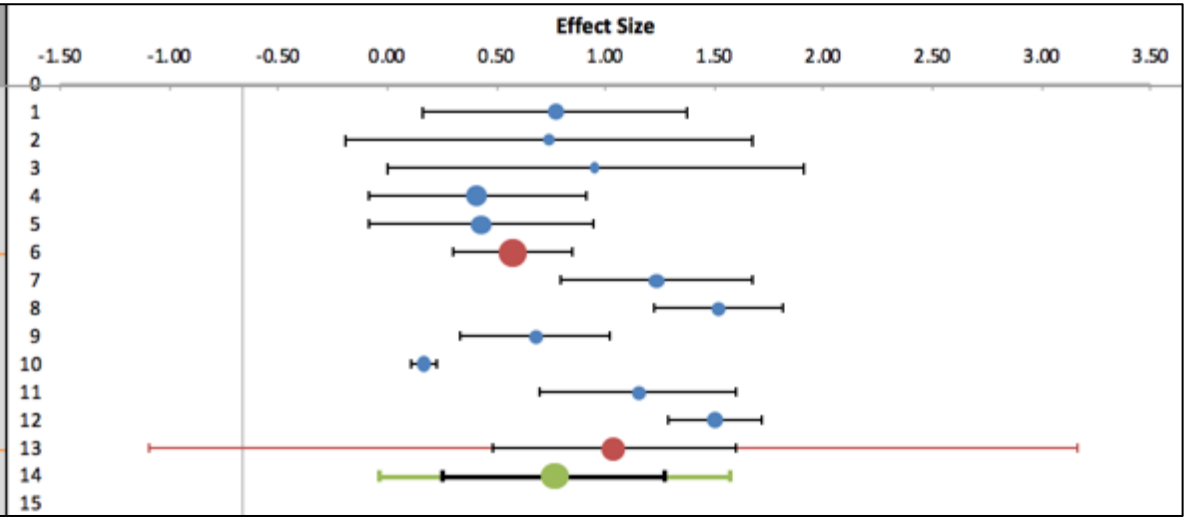


Figure 17: Forest plot for subgroup analysis of study design within the 'attitude' meta-analysis

#	Study name / Subgroup name	Effect Size	CI LL	CI UL	Weight
1	Farrell (2016)	0.04	-0.54	0.62	13.35%
2	Gega (2007)	0.20	-0.12	0.52	41.28%
3	Harned (2014) OLT+ME	-0.14	-0.64	0.36	17.79%
4	Harned (2014) OLT+ME+LC	0.29	-0.21	0.79	17.79%
5	McDonough (2002)	0.67	-0.02	1.36	9.78%
6	**Active**	**0.18**	**-0.12**	**0.48**	**52.98%**
7	Harned (2011) ET OLT	2.03	0.93	3.13	55.00%
8	Harned (2011) ET OLT+MI	3.22	1.87	4.57	45.00%
9	**Passive**	**2.57**	**-4.96**	**10.09**	**47.02%**
10	**Combined effect size**	**1.30**	**-1.61**	**4.22**	
11					

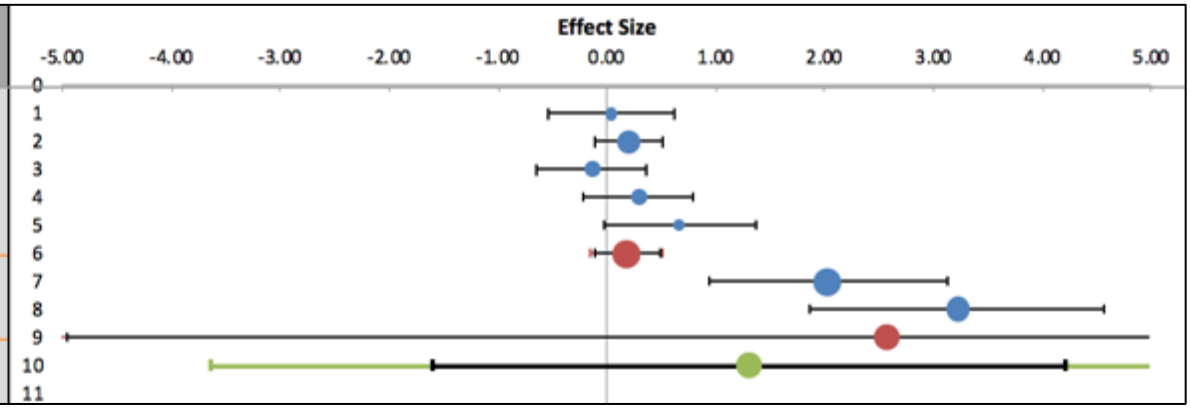


Figure 18: Forest plot for subgroup analysis of the nature of the comparison condition within the 'knowledge' meta-analysis

#	Study name / Subgroup name	Effect Size	CI LL	CI UL	Weight
1	Farrell (2016)	0.77	0.17	1.37	26.52%
2	Harned (2014) OLT+ME	0.41	-0.09	0.91	38.18%
3	Harned (2014) OLT+ME+LC	0.43	-0.09	0.95	35.30%
4	**Active**	**0.51**	**0.04**	**0.98**	**49.55%**
5	Harned (2011) ET OLT	0.74	-0.19	1.67	51.10%
6	Harned (2011) ET OLT+MI	0.95	0.00	1.90	48.90%
7	**Passive**	**0.84**	**-0.49**	**2.18**	**50.45%**
8	**Combined effect size**	**0.68**	**0.22**	**1.14**	
9					

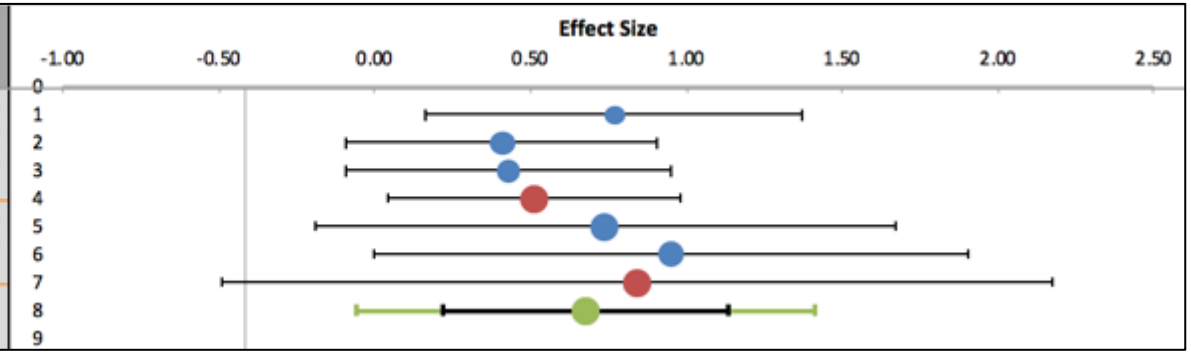


Figure 19: Forest plot for subgroup analysis of the nature of the comparison condition within the 'attitude' meta-analysis

#	Study name / Subgroup name	Effect Size	CI LL	CI UL	Weight
1	Harned (2014) OLT+ME	0.19	-0.31	0.69	50.00%
2	Harned (2014) OLT+ME+LC	0.26	-0.24	0.76	50.00%
3	**Active**	**0.23**	**-0.22**	**0.67**	**51.07%**
4	Harned (2011) ET OLT	1.59	0.55	2.63	52.91%
5	Harned (2011) ET OLT+MI	2.06	0.96	3.16	47.09%
6	**Passive**	**1.81**	**-1.17**	**4.79**	**48.93%**
7	**Combined effect size**	**1.00**	**-1.52**	**3.52**	
8					

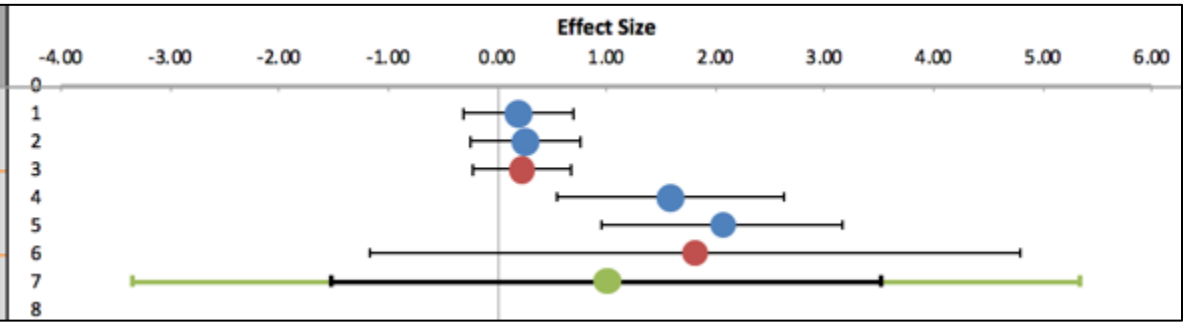


Figure 20: Forest plot for subgroup analysis of the nature of the comparison condition within the 'self-efficacy' meta-analysis

#	Study name / Subgroup name	Effect Size	CI LL	CI UL	Weight
1	Farrell (2016)	0.69	0.11	1.27	15.69%
2	Gega (2007)	-0.23	-0.53	0.07	25.97%
3	Harned (2014) OLT+ME	0.04	-0.46	0.53	18.22%
4	Harned (2014) OLT+ME+LC	0.16	-0.34	0.65	18.22%
5	Kaye (2018)	-0.28	-0.68	0.12	21.90%
6	**Active**	**0.02**	**-0.44**	**0.48**	**54.15%**
7	Harned (2011) ET OLT	0.43	-0.48	1.34	50.00%
8	Harned (2011) ET OLT+MI	0.07	-0.84	0.98	50.00%
9	**Passive**	**0.25**	**-2.04**	**2.54**	**45.85%**
10	**Combined effect size**	**0.13**	**-0.15**	**0.40**	
11					

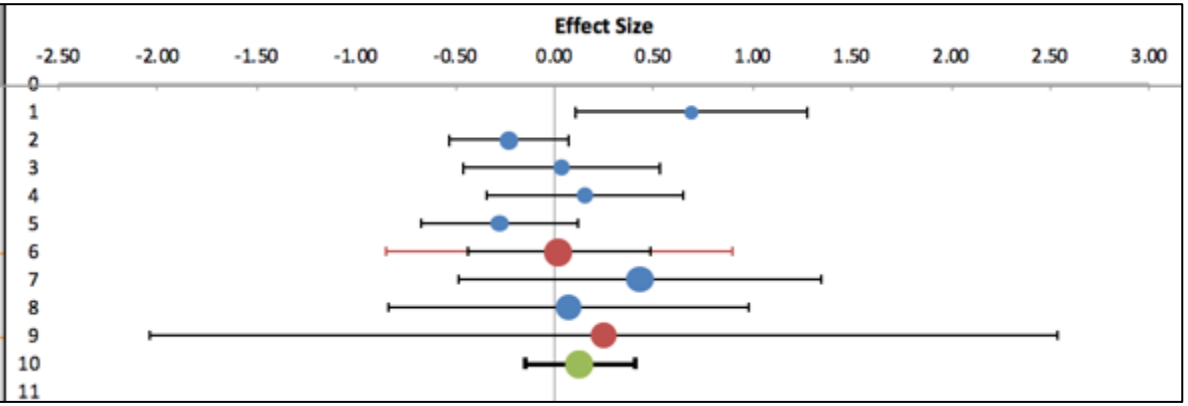


Figure 21: Forest plot for subgroup analysis of the nature of the comparison condition within the 'behaviour' meta-analysis

#	Study name / Subgroup name	Effect Size	CI LL	CI UL	Weight
1	Harned (2011) ET OLT	0.74	-0.19	1.67	12.21%
2	Harned (2011) ET OLT+MI	0.95	0.00	1.90	11.68%
3	Harned (2014) OLT+ME	0.41	-0.09	0.91	39.55%
4	Harned (2014) OLT+ME+LC	0.43	-0.09	0.95	36.56%
5	**ATETS**	**0.52**	**0.18**	**0.87**	**52.46%**
6	Farrell (2016)	0.77	0.17	1.37	15.17%
7	Kaye (2018)	1.52	1.22	1.82	21.82%
8	Reid (2017)	0.68	0.34	1.02	20.95%
9	Waller (2016)	1.15	0.70	1.60	18.67%
10	Deacon (2013)	1.50	1.28	1.72	23.40%
11	**TBES**	**1.16**	**0.66**	**1.65**	**47.54%**
12	**Combined effect size**	**0.82**	**0.09**	**1.56**	
13					

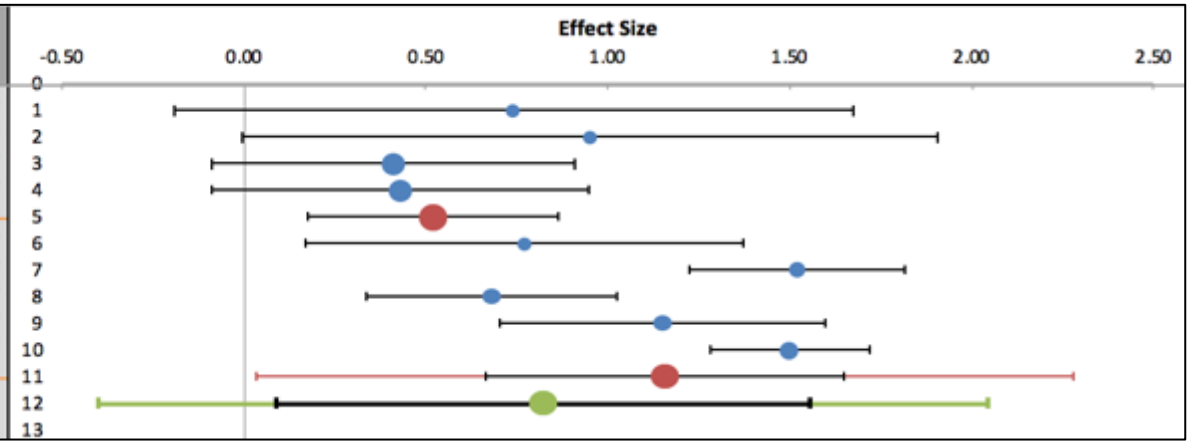


Figure 23: Forest plot for subgroup analysis of the nature of the attitudes outcome measure within the 'attitude' meta-analysis

#	Study name / Subgroup name	Effect Size	CI LL	CI UL	Weight
1	Gega (2007)	-0.23	-0.53	0.07	26.16%
2	Harned (2014) OLT+ME	0.04	-0.48	0.56	14.27%
3	Harned (2014) OLT+ME+LC	0.35	-0.15	0.85	15.05%
4	Kaye (2018)	-0.26	-0.66	0.14	19.77%
5	Reid (2017)	-0.40	-0.72	-0.08	24.75%
6	**Assessed**	**-0.15**	**-0.50**	**0.19**	**53.43%**
7	Farrell (2016)	0.69	0.11	1.27	12.99%
8	Harned (2011) ET OLT	0.43	-0.48	1.34	7.23%
9	Harned (2011) ET OLT+MI	0.07	-0.84	0.98	7.23%
10	Harned (2014) OLT+ME	0.03	-0.47	0.53	15.41%
11	Harned (2014) OLT+ME+LC	-0.04	-0.54	0.46	15.41%
12	Kaye (2018)	-0.29	-0.69	0.11	19.12%
13	Reid (2017)	-0.26	-0.58	0.06	22.60%
14	**Self-reported**	**0.01**	**-0.32**	**0.34**	**46.57%**
15	**Combined effect size**	**-0.08**	**-0.25**	**0.10**	
16					

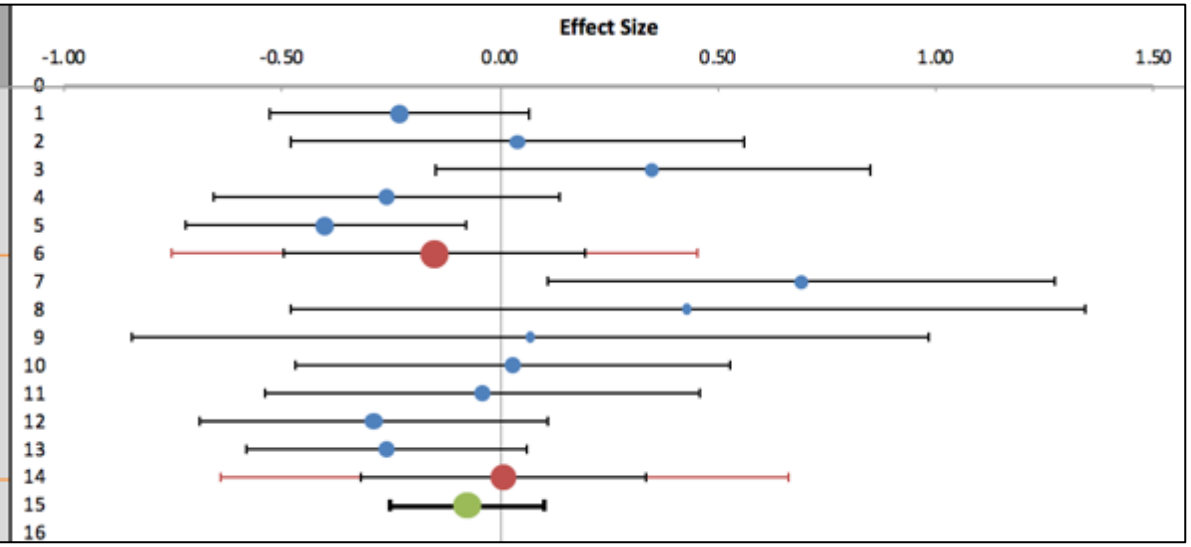


Figure 22: Forest plot for subgroup analysis of the nature of the behaviour outcome measure within the 'behaviour' meta-analysis

This page is intentionally blank

PART TWO: Research Report

Can forming implementation intentions help therapists to adhere to Cognitive Behavioural Therapy protocols regarding weighing adult patients with eating disorders? A randomised-controlled trial

Abstract

Objectives

Weighing patients is a key component of cognitive behavioural therapy (CBT) for eating disorders. However, many therapists do not weigh their patients, which can have a negative impact on patient outcomes. The current study investigated: (i) whether prompting therapists to form implementation intentions changed their weighing behaviours, and (ii) whether therapists' levels of anxiety and intentions to weigh moderated and/or mediated the effect of the intervention.

Design

Mixed between- and within-participants randomised-controlled trial.

Method

Eighty-four therapists actively using CBT with adults with eating disorders were randomised to an intervention or 'usual practice' control condition. Therapists in the intervention condition were prompted to form implementation intentions to support weighing. Therapists completed the following measures at baseline, post-intervention, and follow-up: Intolerance of Uncertainty Scale-12, specific anxiety about weighing, intentions to weigh, and weighing behaviour (percentage of patients weighed).

Results

No significant differences were found in weighing behaviour over time for either condition. However, therapists in the intervention group weighed a higher percentage of patients than those in the control group at post-intervention (U=552.50, p=.042, d=0.47); an effect that approached significance at follow-up (U=564.50, p=.061, d=0.41). Anxiety or intentions to weigh did not significantly impact the effect of the intervention.

Conclusions

Although there was no positive effect of forming implementation intentions on weighing behaviours, implementation intentions might help therapists to maintain effective behaviour. The lack of increase in weighing in the intervention group is in contrast to previous research and may be explained by ceiling effects or aspects of the study's methodology.

Practitioner points and limitations

- Implementation intentions may be helpful in maintaining effective therapist behaviour when using CBT for eating disorders.
- Forming implementation intentions may be beneficial across therapists regardless of their anxieties or motivation.
- Implementation intentions are a relatively quick, easy and cheap intervention, so even small effects might warrant the investment of time and effort.
- Therapists might have been more likely to participate if they already adhered more closely to CBT protocols, potentially leading to selection bias.
- Relatively high levels of attrition might have weakened the overall power of the study.

Keywords

Eating disorders, CBT, weighing, therapist drift, implementation intentions

Introduction

Eating disorders are serious mental health conditions, with significant health, social and economic costs (e.g., Arcelus, Mitchell, Wales & Nielsen, 2011; Becker, Grinspoon, Klinbanski & Herzog, 1999; Crow et al. 2009). They have the highest mortality rate amongst psychiatric conditions, and service costs are projected to rise to £23.8 million by 2026 (King's Fund, 2008). Therefore, it is essential that effective treatment is offered and delivered to patients with eating disorders.

The evidence that cognitive behavioural therapy (CBT) is an effective treatment for eating disorders is extensive. The National Institute for Health and Care Excellence (NICE) recommend CBT in the treatment of adults with anorexia nervosa, binge-eating disorder, bulimia nervosa, and other specified feeding and eating disorders (NICE, 2017). Compared with wait-list or active controls, including medication, CBT has demonstrated consistently equivalent or superior results in improvements in eating disorder symptoms (e.g., Hay, Bacaltchuk, Stefano, & Kashyap, 2009; Linardon, Wade, De la Piedad Garcia, & Brennan, 2017).

The role of weighing

Manuals for CBT for eating disorders instruct therapists to 'openly' weigh their patients, such that patients know their weight (Fairburn, 2008; Waller et al., 2007). Openly weighing patients with eating disorders is necessary for four main reasons (Waller & Mountford, 2015). First, it provides objective information about the patient's weight that enables the therapist to monitor patient safety (e.g., life-threatening weight-loss). Second, it provides a more valid measure of patterns of weight change, as compared to self-report. Third, open weighing can be used as an exposure task to help address anxiety around weighing and weight. Finally, open weighing can be used to test patients' predictions about weight gain through behavioural experiments.

Therapist drift

Unfortunately, many therapists do not adhere to evidence-based protocols (Tobin, Banker, Weisberg, & Bowers, 2007; von Ranson, Wallace, & Stevenson, 2013); despite research showing that manualised therapies are more effective than less-structured approaches (Addis & Waltz, 2002; Cukrowicz et al., 2011). For example, Waller, Stringer, and Meyer (2012) found that fewer than 40% of therapists using CBT with adults with eating disorders routinely openly weighed patients, and many therapists report not weighing patients (Mulkens, de Vos, de Graaff, & Waller, 2018).

Waller (2009) described therapists' failure to deliver key elements of treatment protocols as 'therapist drift'. The consequence of therapist drift is often poorer outcomes for patients (Waller & Turner, 2016; Wiborg, Knoop, Frank, & Bleijenberg, 2012), and may explain, in part, why CBT does not benefit all patients (Forbush, Richardson, & Bohrer, 2015; Kass, Kolko, & Wilfley, 2013). Therapist drift might result from a reliance on clinical judgement, even though clinical judgement has been found to be less effective than protocol-driven approaches (e.g., Dawes, Faust, & Meehl, 1989; Folke et al., 2017; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Meehl, 1954; Wilson, 1996). Therapist drift might also result from clinicians' fear that adhering to CBT protocols could damage the therapeutic alliance. However, there is support for the view that the therapeutic alliance is likely to be "necessary but not sufficient to produce an optimum therapeutic effect" (Beck, Rush, Shaw, & Emery, 1979, p. 45), indicating the need for therapists to (also) focus on their technique. Brown, Mountford, and Waller (2013) found that a focus on the therapeutic alliance, rather than elements of the therapeutic protocol or early behavioural change (e.g., structured eating, early focus on weight gain), resulted in less weight gain when using CBT for anorexia nervosa. Moreover, longitudinal research has demonstrated that early symptom change actually drives an improvement in the therapeutic alliance (Brown, et al., 2013; Crits-Cristoph, Connolly

Gibbons, & Hearon, 2006; Tang & DeRubeis, 1999; Turner, Bryant-Waugh, & Marshall, 2015). Taken together then, it is clear that therapists need to practice effective CBT techniques, such as open weighing.

Barriers to weighing patients

Many reasons have been advanced for why therapists do not weigh patients (Waller & Mountford, 2015). First, therapist drift when using CBT with adults with eating disorders might be a *motivational* problem, in that some therapists do not intend to weigh their patients in the first place. Although some therapists might simply not be aware of protocols and thus do not know that they should weigh their patients (Addis & Krasnow, 2000), Waller et al. (2013) found that 92% of therapists were aware of manuals but only half of these therapists 'often' used manuals. Therapists might choose not to use protocols due to negative attitudes towards manualised approaches or specific therapy tasks (Addis & Krasnow, 2000; Tobin, Banker, Weisberg, & Bowers, 2007), and/or because they place a higher value on clinical judgement over protocols (Waller & Turner, 2016).

Second, therapist drift might be a *volitional* problem, in that even if therapists intend to weigh their patients, they have difficulties translating their intentions into action. For example, practical issues might prevent weighing, such as patients being weighed by another professional, inadequate equipment (e.g., lack of weighing scales), an organisational lack of endorsement for weighing, time constraints in sessions, or patients refusing to be weighed. Furthermore, therapists' own anxieties regarding weighing patients might be a barrier, and evidence suggests that therapists with higher levels of anxiety are less likely to adhere to CBT techniques (Waller et al., 2012; Waller & Turner, 2016). One reason for this may be that weighing patients is an exposure-based task. Exposure requires therapists to purposively evoke distress in their patients. Consequently, therapists might become anxious about distressing patients, increasing eating disorder behaviours, damaging the therapeutic

relationship, or the negative impact on themselves through vicarious distress or increasing risk of litigation (Meyer, Farrell, Kemp, Blakey, & Deacon, 2014; Olatunji, Deacon, & Abramowitz, 2009; Turner, Tatham, Lant, & Waller, 2014; Waller, 2009).

How might these potential barriers to weighing patients be overcome?

Targeting motivational and volitional barriers to behaviour change. The Theory of Planned Behaviour (TPB; Ajzen, 1991) proposes that intentions – self-instructions to achieve certain outcomes (e.g., therapists might intend to weigh patients with eating disorders; Triandis, 1980) – are the proximal determinant of behaviour (Ajzen, 1991; Bandura, 1977). However, Webb and Sheeran's (2006) meta-analysis showed that a medium-to-large-sized change in intention leads only to a small-to-medium-sized change in behaviour, suggesting that a substantial proportion of variance in behaviour is not accounted for by goal intentions. This is commonly known as the 'intention-behaviour gap' (Sheeran, 2002; Sheeran & Webb, 2016) and suggests that overcoming barriers to weighing is unlikely to simply be a matter of increasing the strength of therapists' motivation or intentions to weigh.

The intention-behaviour gap has been documented in research investigating the effectiveness of training interventions for clinicians. For example, Tomasone, Ginis, Estabrookes, and Domenicucci (2014) investigated the effect of a training seminar on clinicians' rates of discussing physical activity with patients. Results showed that, although the intervention changed participants' beliefs as specified by the TPB (i.e., attitudes, subjective norms, perceived behavioural control, and intentions), it was not effective in changing or maintaining clinicians' behaviours. Similarly, a meta-analysis investigating the effectiveness of training clinicians in exposure therapy found that training interventions typically improved clinicians' knowledge, attitudes, self-efficacy and intentions, but not their behaviour (Trivasse, 2019). These findings support the idea that therapist drift in CBT for eating disorders might be a volitional, rather than motivational problem – even if therapists

have the intentions and confidence to weigh their patients, they struggle to translate their intentions into action.

Gollwitzer (1993) proposed that goal intentions are more likely to be translated into action if they are supplemented by specific behavioural plans. Planning may help individuals to overcome volitional problems, such as procrastination, competing goals, or struggling to seize opportunities to act (Gollwitzer, Bayer, & McCulloch, 2005; Sheeran, Milne, Webb, & Gollwitzer, 2005). Gollwitzer (1999) developed a specific form of planning that he referred to as 'implementation intentions' or 'if-then' planning. Whereas goal intentions outline what the individual aims to achieve (e.g., "I intend to weigh all my patients with eating disorders"), implementation intentions specify the behaviour that they will perform to achieve that goal and the context in which they will enact it (e.g., "If my patient becomes distressed, then I will remind them of the rationale for weighing!")

Supplementing goal intentions with 'if-then' plans has been found to promote goal realisation over and above strong intentions alone. For example, meta-analyses have found positive, medium-to-large effects of forming implementation intentions on goal attainment, which were maintained across different study designs, measures, and domains of goal attainment in clinical and non-clinical samples (Gollwitzer & Sheeran, 2006; Toli, Webb, & Hardy, 2016). Interventions based on implementation intentions have also been found to be beneficial in increasing clinicians' use of specific procedures (Casper, 2008), and in reducing stress and increasing work engagement (Gollwitzer, Mayer, Frick, & Oettingen, 2018).

Moderators of the effect of forming implementation intentions on outcomes. Although the positive effects of forming implementation intentions are well-established, several factors have been found to moderate those effects. For example, forming implementation intentions has been shown to have a greater impact on behaviour change in individuals with strong goal intentions compared to individuals with weak goal intentions,

suggesting that people need to be motivated to achieve the goal to benefit from forming implementation intentions (e.g., Elliott & Armitage, 2006; Sheeran, Webb, & Gollwitzer, 2005; van Osch, Reubsaet, Lechner, & de Vries, 2008; Wiedemann, Schuz, Sniehotta, Scholz, & Schwarzer, 2009). As such, therapists who have strong intentions to weigh their patients might benefit the most from forming implementation intentions.

Furthermore, forming implementation intentions appears to particularly support the attainment of difficult goals compared to easier goals. For example, forming implementation intentions has been found to facilitate goal attainment when goals are acted on at inconvenient times (Gollwitzer & Brandstätter, 1997), or when the goal is easily forgotten (Sheeran & Orbell, 1999). When participants find striving for goals more difficult, there might be more room for improvement, and thus a greater benefit from forming implementation intentions (Dewitte, Verguts, & Lens, 2003). Given that therapists' levels of anxiety have been shown to be barriers to weighing patients with eating disorders (Meyer et al., 2014; Turner et al., 2014; Waller, 2009), therapists with higher levels of anxiety may find weighing patients more difficult, and so forming implementation intentions might be more beneficial for therapists who are anxious about weighing. Alternatively, given that low levels of anxiety have been associated with lower therapist drift (Waller et al., 2012; Waller & Turner, 2016), therapists with low anxiety might benefit most from forming implementation intentions.

Current Study

The current research explored the use of implementation intentions as an intervention for therapists using CBT with adults with eating disorders. This study investigated the possibility that prompting therapists to set goal intentions (to weigh patients) and form implementation intentions (to support these intentions) might help them to adhere more

closely to CBT protocols, and more specifically, increase their weighing behaviour. Using a randomised-controlled trial (RCT) design, therapists in the intervention condition were reminded of the importance of weighing patients with eating disorders, and prompted to form 'if-then' plans specifying how they would take advantage of opportunities to weigh patients and deal with potential barriers (e.g., anxiety). A 'usual practice' control group was used to highlight any specific effects of the intervention by controlling for the effects of participating in a study. Additionally, the current study utilised a follow-up time-point to investigate whether any changes to weighing behaviour were sustained.

Moderating factors. In line with previous research showing the moderating effect of the strength of goal intentions (e.g., Sheeran, Webb, & Gollwitzer, 2005), the current study explored therapists' intentions to weigh their patients as a potential moderator of the effect of forming implementation intentions, hypothesising that therapists who are more motivated to weigh patients might benefit more from the intervention. Therapists' general level of anxiety was also investigated as a moderator of the effect of the intervention, with two competing hypotheses. First, it was hypothesised that therapists with low anxiety might benefit most from the intervention, as low levels of anxiety have been associated with lower therapist drift (Waller et al., 2012; Waller & Turner, 2016). Alternatively, therapists with higher levels of anxiety might find weighing patients more difficult and so might benefit more from forming implementation intentions.

Mediating factors. It is also important to understand why prompting therapists to set goal intentions and form implementation intentions may have a positive impact on their weighing behaviour. Therapist's anxiety specifically around weighing patients with eating disorders was assessed as a mediating factor given that anxiety is a likely barrier to weighing (Meyer et al., 2014; Turner et al., 2014; Waller, 2009). It was hypothesised that the

intervention will reduce specific weighing anxiety, and subsequently help therapists to weigh their patients.

Aims

1. To investigate whether prompting therapists to set goal intentions to weigh their patients and form implementation intentions designed to support these intentions changes therapists' weighing behaviours when using CBT for adults with eating disorders.
2. To investigate whether therapists' intentions to weigh their patients and therapists' general levels of anxiety moderate the effect of the intervention on weighing behaviour.
3. To investigate whether therapists' specific anxiety about weighing patients mediates the effect of the intervention on weighing behaviour.

Hypotheses

1. Therapists who are prompted to set goal intentions to weigh their patients and form implementation intentions designed to support these intentions will show a significantly greater increase in weighing behaviour compared to therapists who receive no intervention.
2. Therapists in the intervention condition will maintain the increase in weighing behaviour over time.
3. Therapists' general levels of anxiety will moderate the effect of the intervention on weighing behaviour, either because (i) therapists with low general anxiety will show a greater increase in weighing behaviour, or (ii) therapists with high general anxiety will show a greater increase in weighing behaviour.

4. The strength of therapists' intentions to weigh their patients will moderate the effect of the intervention on weighing behaviour, in that therapists who are more motivated will show a greater increase in weighing behaviour.
5. Therapist's anxiety about weighing patients will mediate the impact of the intervention on weighing behaviour.

Method

Design

A prospective RCT methodology was employed, based on a mixed between- and within-participants design. Participants were randomised to one of two conditions: (i) an intervention group, and (ii) a 'usual practice' control group. Changes in weighing behaviour were measured at three time-points – baseline, post-intervention, and follow-up.

Ethics

Ethical approval was received from the University of Sheffield's Department of Psychology Research Ethics Committee (Appendix A).

Confidentiality and data security. All data were kept confidential, used only for the purposes of the research, and kept secure within the password-protected Qualtrics database. Identifiable information was removed when data were downloaded for statistical analysis. Participants provided informed consent at the start of the study process and were informed that they had the right to withdraw their consent and have their data destroyed at any time prior to data analysis.

Minimal deception. Participants were not informed about the final follow-up measure prior to providing consent, in order to prevent any effects of participants knowing that they would be followed up (e.g., demand or measurement effects; Godin, Sheeran, Conner, & Germain, 2008; Orne, 1962). However, participants were provided with a full debrief at the

end of the study (Appendix B). An adverse event form was used to help monitor risk and complaints (Appendix C)[1].

Access to the intervention. At the end of the study, all participants, regardless of group allocation, received information about the importance of weighing along with the volitional help-sheet to create 'if-then' plans.

Power analysis.

A priori power analysis was conducted using Cohen's table (Cohen, 1992). A large-sized effect ($d = 0.80$) of the intervention on outcomes was estimated, based on the medium-to-large size of the effect of forming implementation intentions on outcomes reported by Gollwitzer and Sheeran (2006), alongside evidence of a large-sized effects in research exploring the effects of combined motivational and volitional interventions on behaviour change (e.g., Fritzsche, Schlier, Oettingen, & Lincoln, 2016; Milne, Orbell, & Sheeran, 2002). A large-sized effect, significance level of alpha = 0.05, two groups of participants, and three measurements, suggested that a total sample size of 52 participants (26 per group) would provide 80% power to detect an effect of similar magnitude. Eighty-percent power means that the study has an 80% chance of finding a statistically significant effect if the effect exists, or 80% chance of rejecting the null hypothesis if it is false. Eighty-percent power is considered acceptable for RCTs in clinical research (Wittes, 2002). The aim was to recruit an additional 15 therapists to allow for an estimated 30% attrition rate, resulting in a total planned sample size of 67 participants at baseline.

Participants

Therapists were required to be actively using CBT with adults with eating disorders and able to read and write in English. No other exclusion criteria were applied. Therapists were recruited between August 2018 and March 2019 through relevant professional contacts,

[1] No adverse events were reported.

conferences and training events in the UK, Netherlands, USA, Australia, and Sweden. Participants were encouraged to share the recruitment invitation with colleagues; meaning that the study adopted an opportunistic snowball sampling method. A total of 84 therapists participated – 40 in the control condition, and 44 in the intervention condition. The flow of participants through the study is presented in Figure 1.

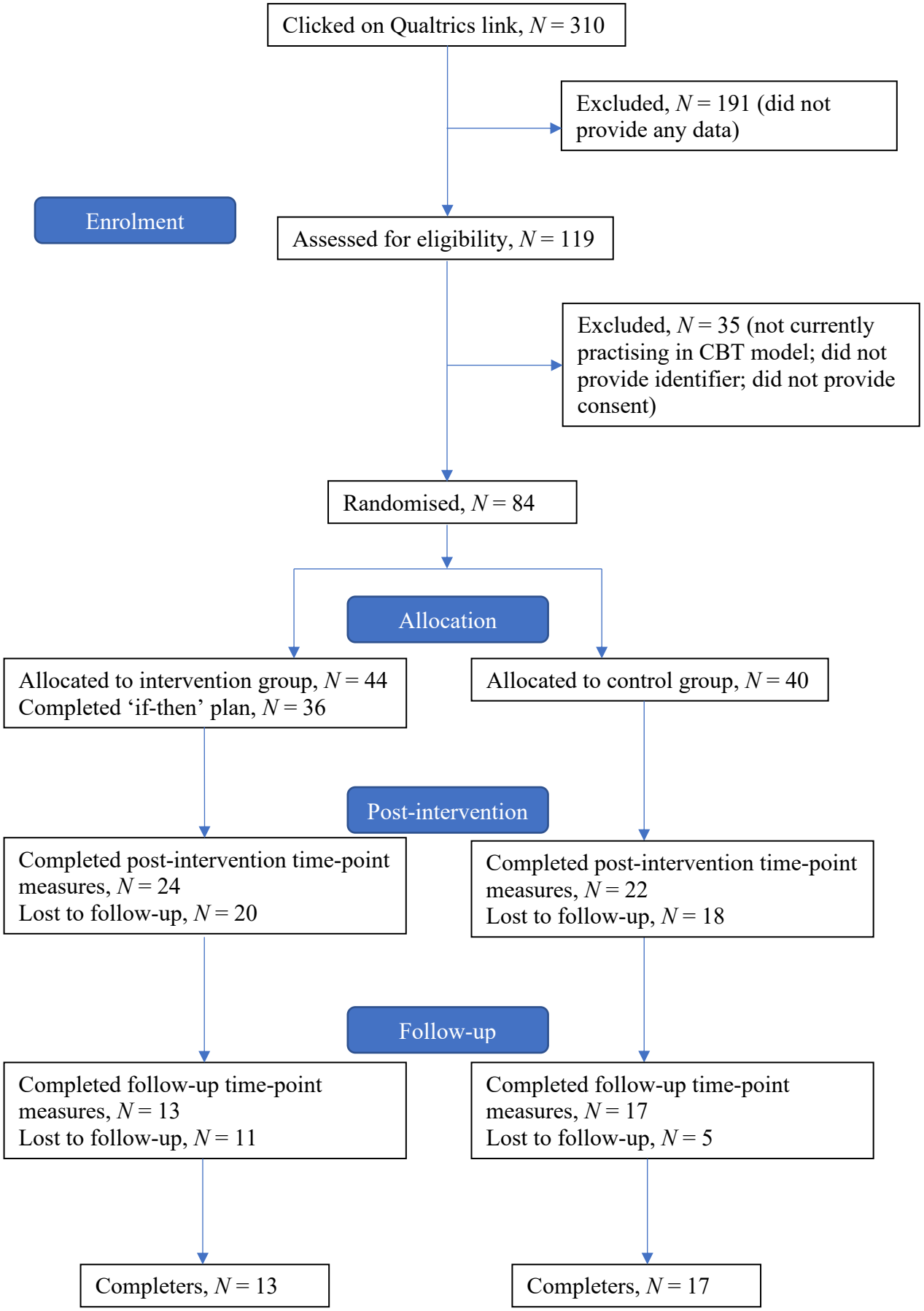


Figure 1: CONSORT diagram outlining the flow of participants through the study (Schulz, Altman, & Moher, 2010)

Measures

Weighing behaviour. Therapists were asked to report the total number of patients with eating disorders that they had seen in the past two weeks (a period that should allow for most regular patients to be seen), and the percentage of those patients who they openly weighed in session. Therapists were also asked additional questions regarding their weighing behaviour, such as the percentage of patients weighed by another professional. Asking about a range of weighing behaviours aimed to help therapists to view 'not weighing' as an option, thereby encouraging honesty and limiting the impact of demand characteristics.

General anxiety. The 12-item version of the Intolerance of Uncertainty Scale (IUS-12; Carleton, Norton, & Asmundson, 2007; Appendix E) was used to measure therapists' general levels of anxiety. The IUS-12 has been shown to be internally consistent ($\alpha = .91$) and strongly correlated ($r = .96$) with the original 27-item scale (Carleton, et al., 2007). Although the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg & Jacobs, 1983) is more commonly used to measure anxiety, it is likely to be very familiar to therapists and so was not considered suitable. The IUS-12 has been found to be moderately correlated and show similar concurrent validity to other anxiety scales (Khawaja & Yu, 2010).

Anxiety specific to weighing behaviour. Six questions were added to the IUS-12 to measure therapists' anxieties specifically about weighing patients (Appendix F). The questions took the same format as in the IUS-12, using a statement (e.g., 'I worry about weighing my clients because I am unsure how to respond if they refuse to be weighed') and a 5-point scale anchored by 'Not characteristic of me' and 'Entirely characteristic of me'. The questions were derived from existing literature and pilot work (outlined below), and Cronbach's alpha was calculated to assess reliability.

Intentions to weigh. Therapists' intentions to weigh their patients were measured by asking therapists to indicate the extent to which they agreed with the statement, "I intend to

weigh all the patients I am treating for an eating disorder" (on a 7-point scale, anchored by 'strongly disagree' and 'strongly agree'). Intentions to perform other therapy tasks were also measured (e.g., "I intend to set homework in sessions") to disguise the focus of the study.

Reliability of anxiety measures

The reliability of the IUS-12 and the measure of specific anxiety were assessed using Cronbach's alpha (Cronbach, 1951). A cut-off of alpha = 0.70 was taken to indicate an acceptable level of reliability (Tavakol & Dennick, 2011). The IUS-12 and the measure of specific anxiety demonstrated acceptable reliability at baseline and post-intervention time-points (Table 1).

Table 1: Reliability analysis results for anxiety measures

Time-point	Measure	Number of items	*N*	Cronbach's alpha
Baseline	IUS-12	12	79	.85
	Specific anxiety	6	79	.73
Post-intervention	IUS-12	12	48	.89
	Specific anxiety	6	48	.71

Pilot work to create intervention materials and measure of specific anxiety

Information about the importance of weighing. Participants in the intervention condition received an information sheet about the importance of weighing patients with eating disorders. This outlined the four key reasons for weighing patients (Waller & Mountford, 2015; Appendix G): (i) patient safety; (ii) accurately identifying changes in weight and eating patterns; (iii) reducing anxiety about weighing through exposure tasks; (iv) addressing the 'broken cognition' (i.e., any food intake is seen to have catastrophic effects on weight and thus intake is minimised to avoid this) through behavioural experiments.

Volitional help-sheet and measure of specific anxiety. Common potential barriers and solutions to weighing patients were identified from the existing literature and through interviews with clinicians working in the field of eating disorders (see Appendix I for interview schedule) to create the measure of specific anxiety and a volitional help-sheet. The

volitional help-sheet provided a structure for therapists to form goal intentions and implementation intentions to weigh their patients (Appendix H). Specifically, the author and another trainee clinical psychologist interviewed five clinicians[2] about the challenges that they encounter when weighing patients, and how they might tackle those challenges. For the volitional help-sheet, a list of potential barriers and solutions was created, and then refined to nine 'if' statements (Table 2) and nine 'then' statements (Table 3). The aim was for these to be clear, frequently endorsed, and general enough to apply to most therapists. This pilot work was conducted in parallel with another DClinPsy research project (see Appendix J for details).

[2] A record of these therapists' details was kept, to ensure that they were not included as participants in the main study.

Table 2: List of potential 'if' statements as part of the volitional help-sheet

'If' statements
If my clients becomes distressed …
If weighing my client makes me feel anxious …
If I feel uncomfortable sitting with my client's distress …
If I struggle to fit weighing into my session or I run out of time …
If I think that it is unlikely that I will weigh my client this session …
If I don't think that it's important to weigh this client …
If there are practical challenges to weighing my client …
If I think that I won't weigh because my client looks like they've gained weight or are a healthy weight …
If my client refuses to be weighed …

Table 3: List of potential 'then' statements as part of the volitional help-sheet

'Then' statements
Then I will remind myself that weighing my client is an opportunity to explore their thoughts and emotions!
Then I will take my feelings/experience to supervision and access support!
Then I will include 'weighing' on my session agenda, and weigh my client at the start of the session!
Then I will remind myself of the expectation of weighing in the treatment contract!
Then I will revisit the evidence-based protocol and remind myself of the rationale for weighing clients!
Then I will ensure that I have access to scales prior to the session!
Then I will remind myself of the importance of using objective measures to monitor my client's weight!
Then I will explain the rationale of weighing to them, and discuss their commitment to treatment!
Then I will not present being weighed as an option, but will ask when my client wants to be weighed during the session (e.g., "now or after 10 minutes?")!

Procedure

Baseline. Following recruitment, therapists provided demographic information and consent, and were randomly allocated to the intervention or control condition using Qualtrics' randomizer software. Both therapists and researchers were blind to allocation. Therapists were asked to not discuss the content of the study with other clinicians. Once randomised, therapists completed all of the measures. Following baseline data collection, therapists in the intervention condition received the information sheet and volitional help-sheet, and were

asked to form the intention to openly weigh patients every session and a personalised 'if-then' plan with the aim to increase their weighing behaviour. Measures were given prior to the information sheet or volitional help-sheet to minimise the impact of desirability bias (i.e., the concern that therapists may inflate their score to be in line with CBT protocols if they had just read information about the importance of weighing). The control group did not receive any further information.

Post-intervention. Two weeks after completing the first questionnaire, therapists received an email asking them to complete a second questionnaire (Appendix K), which contained the same measures as baseline. Therapist's data was not included if they did not complete the measures within two weeks of the post-intervention time-point, to ensure that all participants reflected on a similar period following recruitment.

Follow-up. Four weeks after the post-intervention time-point, therapists were asked to complete a third questionnaire containing the same measures as the post-intervention time-point (Appendix L). As before, therapist's data was not included if they did not complete the measures within two weeks of the follow-up time-point. After completing the measures at follow-up, therapists received an email explaining the full nature and purpose of the study (Appendix B). Therapists were informed that this was the final time that they would be contacted. Therapists in the control condition received the information sheet and volitional help-sheet.

Approach to data analysis

Data were analysed using the Statistical Package for Social Sciences (SPSS) Version 25. Descriptive statistics for baseline variables were explored, and attrition analyses were conducted to assess differences between completers and non-completers. Histograms, Q-Q plots and Kolmogorov-Smirnov statistics were used to assess the distribution of the data.

Two approaches to statistical analysis are outlined below, which aimed to manage missing data and non-normally distributed data.

Managing missing data. Multiple imputation was planned as the primary statistical method to manage missing data. Multiple imputation creates several plausible imputed data sets and combines the results to provide a 'pooled' data set (Rubin, 1987). Unlike single imputation, this process takes the uncertainty in imputations into account, thereby reducing error (Sterne et al., 2009). For accurate stable *p*-values following multiple imputation, recommendations have suggested that 40 imputations are needed where over 50% of the data is missing (Graham, Olchowski, & Gilreath, 2007), or that the number of imputations should be similar to the percentage of cases that are incomplete (Bodner, 2008; White, Royston and Wood (2011). Given the overall 64% attrition rate at follow-up (Figure 1), 50 imputations were used.

Primary analyses following multiple imputation. A two-by-three mixed-methods ANOVA was planned. However, SPSS does not support such ANOVAs as a statistical test following multiple imputation. Consequently, planned *t*-test comparisons were used. Paired-samples *t*-tests were used to assess differences between time-points for the control and intervention groups separately, and independent-samples *t*-tests were used to assess differences between groups at each time-point. Bonferroni corrections ($p = .017$ for three comparisons) were applied to minimise the risk of Type 1 error. Pearson product-moment correlations were conducted to investigate associations between changes in weighing behaviour and general anxiety, specific anxiety, and intentions to weigh. In the event of significant correlations, hierarchical regression was planned to explore moderators of the effect of the intervention, and mediation analysis was planned using Hayes Process Macro (Hayes, 2013).

Primary analyses in the case of non-normally distributed data. In the event of non-normally distributed data, non-parametric statistics were planned. As SPSS does not support the use of non-parametric tests following multiple imputation, an intent-to-treat (ITT) analysis using 'last observation carried forward' (LOCF) was used to manage missing data. Following LOCF, Friedman tests were used to investigate changes in weighing behaviour across time-points for the intervention and control groups separately. If Friedman tests were significant, post-hoc analyses were planned using Wilcoxon Signed Rank tests, with a Bonferroni correction applied (significance level $p = .017$) to control for Type 1 error. Mann-Whitney U tests were used to assess differences between groups at each time-point. Spearman Rank Order correlations were used to investigate associations between changes in weighing behaviour and general anxiety, specific anxiety, and intentions to weigh. In the event of significant correlations, hierarchical regression was planned to explore moderators of the effect of the intervention, and mediation analysis was planned using Hayes Process Macro (Hayes, 2013). However, in the context of non-normally distributed data, parametric regression analyses were conducted tentatively.

Multiple imputation followed by non-parametric statistics would have provided the most reliable results, as both missing data and non-normally distributed data would have been managed more effectively. Unfortunately, SPSS does not support this combination of analyses. Consequently, reporting both multiple imputation and LOCF statistical procedures aimed to offer a more comprehensive assessment of the data and account for the limitations of both statistical approaches.

Cohen's *d* effect sizes were calculated using the Psychometrica Effect Size Calculator (Lenhard & Lenhard, 2016), and interpreted in line with Cohen's (1992) recommendations[3].

[3] ($d = 0.20$ representing 'small', $d = 0.50$ representing 'medium', and $d = 0.80$ representing 'large')

Data analyses were repeated for the original data-set as recommended by CONSORT guidelines (Schulz et al., 2010; Appendix M).

Results

Demographic characteristics

Demographic characteristics for each condition are summarised in Table 4 for the original, LOCF, and multiple imputation data-sets. The majority of participants were female in both the control (87.5%) and intervention groups (77.3%). Most participants had postgraduate training, such as a doctorate in clinical psychology, and most had been working within a CBT model for over five years. Of those allocated to the intervention group, 86% (N = 38) completed an intention statement outlining their intentions to weigh every patient with an eating disorder, and 82% (N = 36) completed an 'if-then' plan, indicating that most participants complied with the intervention.

Normality testing

Significant Kolmogorov-Smirnov tests of normality and visual inspection of histograms and Q-Q plots indicated that most of the data violated the assumption of normality in both the control and intervention conditions (Appendix N). As such, LOCF and non-parametric statistical analyses were conducted in addition to multiple imputation and parametric analyses.

Table 4: Summary of demographic characteristics for original, LOCF, and multiple imputation data-sets

	Original		LOCF		Multiple imputation	
Variable	**Control (*N* = 40)**	**Intervention (*N* = 44)**	**Control (*N* = 40)**	**Intervention (*N* = 44)**	**Control (*N* = 40)**	**Intervention (*N* = 44)**
Age, *M (SD)*	35.80 (10.13)	39.83 (10.13)	35.80 (10.13)	39.83 (10.13)	35.80	40.20
Missing, *N*	-	2	-	2	-	-
Gender, *N (%)*						
Female	35 (87.5)	34 (77.3)	35 (87.5)	34 (77.3)	35 (87.5)	35 (79.5)
Male	5 (12.5)	8 (18.2)	5 (12.5)	8 (18.2)	5 (12.5)	9 (20.5)
Missing	-	2 (4.5)	-	2 (4.5)	-	-
Level of CBT training, *N (%)*						
Professional training	9 (22.5)	9 (20.5)	9 (22.5)	9 (20.5)	9 (22.5)	9 (20.5)
Postgraduate training	27 (67.5)	28 (63.6)	27 (67.5)	28 (63.6)	27 (67.5)	28 (63.6)
No training	0 (0.0)	1 (2.3)	0 (0.0)	1 (2.3)	0 (0.0)	1 (2.3)
Other	4 (10.0)	6 (13.6)	4 (10.0)	6 (13.6)	4 (10.0)	6 (13.6)
Length of CBT practice, *N (%)*						
< 1 year	8 (20.0)	7 (15.9)	8 (20.0)	7 (15.9)	8 (20.0)	7.6 (17.3)
1-2 years	8 (20.0)	7 (15.9)	8 (20.0)	7 (15.9)	8 (20.0)	7.5 (17.0)
2-5 years	11 (27.5)	12 (27.3)	11 (27.5)	12 (27.3)	11 (27.5)	12.4 (28.2)
> 5 years	13 (32.5)	16 (36.4)	13 (32.5)	16 (36.4)	13 (32.5)	16.5 (37.5)
Missing	-	2 (4.5)	-	2 (4.5)	-	-

Attrition analysis

From baseline to post-intervention, the rate of attrition in both conditions was 45%. From post-intervention to follow-up, attrition rates were 46% and 23% in the intervention and control conditions, respectively. From baseline to follow-up, there was an overall attrition rate of 64%, with 70% and 58% dropping out of the intervention and control conditions, respectively. Baseline data was compared between completers and non-completers. Mann-Whitney U tests were conducted to compare continuous variables, and chi-squared tests to compare categorical variables. Results showed that there were no significant differences at baseline between the completers and non-completers across all variables (see Table 5).

Table 5: Attrition analysis for completer and non-completer samples

Baseline variable	**Non-completers (*N* = 54)**	**Completers (*N* = 30)**	**Attrition analysis**		
			U	***z***	***p***
Age, *M (SD)*	38.12 (10.34)	37.43 (10.30)	802.50	0.22	.828
Missing, *N*	2	-			
Weighing behaviour, *M (SD)*	70.70 (38.82)	80.17 (30.82)	596.50	-0.82	.413
Missing, *N*	8	1			
General anxiety, *M (SD)*	1.73 (0.42)	1.85 (0.52)	663.50	-0.72	.469
Missing, *N*	5	-			
Specific anxiety, *M (SD)*	1.20 (0.29)	1.19 (0.27)	742.00	0.08	.940
Missing, *N*	5	-			
Intentions to weigh, *M (SD)*	6.18 (1.48)	6.59 (0.78)	575.50	-1.02	.309
Missing, *N*	9	1			
			X^2	**df**	***p***
Gender, *N (%)*					
Female	45 (83.3)	24 (80.0)			
Male	7 (13.0)	6 (20.0)			
Missing	2 (3.7)	-			
			0.22	1	.641
Level of CBT training, *N (%)*					
Professional training	13 (24.1)	5 (16.7)			
Postgraduate training	31 (57.4)	24 (80.0)			
No training	1 (1.9)	0 (0.0)			
Other	9 (16.7)	1 (3.3)			
			5.43	3	.143
Length of CBT practice, *N (%)*					
< 1 year	10 (18.5)	5 (16.7)			
1-2 years	13 (24.1)	2 (6.7)			
2-5 years	15 (27.8)	8 (26.7)			
> 5 years	14 (25.9)	15 (50.0)			
Missing	2 (3.7)	-			
			6.46	3	.091

Table 6: Summary of means and standard deviations for outcome variables in the original, LOCF, and multiple imputation data-sets

	Original		LOCF		Multiple imputation	
Variable	**Control**	**Intervention**	**Control**	**Intervention**	**Control (*N* = 40)**	**Intervention (*N* = 44)**
Baseline weighing behaviour, *M (SD)*	69.56 (37.21)	79.04 (34.67)	69.56 (37.21)	79.04 (34.67)	69.84 (38.49)	78.46 (32.36)
N	37	38	37	38	-	-
Post-intervention weighing behaviour, *M (SD)*	70.08 (36.42)	86.28 (29.90)	63.87 (40.08)	81.34 (33.76)	63.60 (40.48)	82.40 (33.57)
N	25	21	38	39	-	-
Follow-up weighing behaviour, *M (SD)*	77.41 (33.24)	86.03 (18.33)	67.08 (37.78)	81.43 (31.70)	68.98 (38.61)	83.77 (27.11)
N	17	13	38	39	-	-
Baseline general anxiety, *M (SD)*	1.83 (0.58)	1.72 (0.32)	1.83 (0.58)	1.72 (0.32)	1.84 (0.59)	1.72 (0.60)
N	38	41	38	41	-	-
Post-intervention general anxiety, *M (SD)*	1.81 (0.59)	1.62 (0.35)	1.77 (0.53)	1.66 (0.32)	1.78 (0.65)	1.65 (0.52)
N	26	22	38	41	-	-
Baseline specific anxiety, *M (SD)*	1.18 (0.24)	1.21 (0.31)	1.18 (0.24)	1.21 (0.32)	1.19 (0.25)	1.22 (0.40)
N	38	41	38	41	-	-
Post-intervention specific anxiety, *M (SD)*	1.22 (0.27)	1.19 (0.29)	1.21 (0.27)	1.20 (0.28)	1.26 (0.33)	1.22 (0.39)
N	26	22	38	41	-	-
Baseline intentions to weigh, *M (SD)*	6.17 (1.50)	6.50 (0.95)	6.17 (1.49)	6.50 (0.95)	6.22 (1.55)	6.45 (0.99)
N	36	38	36	38	-	-
Post-intervention intentions to weigh, *M (SD)*	6.63 (0.71)	6.20 (1.70)	6.19 (1.50)	6.38 (1.29)	6.38 (2.22)	6.26 (2.40)
N	24	20	37	39	-	-

Multiple imputation data analysis

Table 6 provides a summary of the pooled means and standard deviations of the primary outcome variables. For the control condition, paired-samples *t*-tests indicated that there were no significant differences in weighing behaviour between baseline and post-intervention, $t(803) = 1.28$, $p = .202$, $d = -0.16$ (95% CI [-0.60; 0.28]), baseline and follow-up, $t(360) = 0.20$, $p = .840$, $d = -0.02$ (95% CI [-0.46; 0.42]), or post-intervention and follow-up, $t(73258) = 1.19$, $p = .235$, $d = 0.14$ (95% CI [-0.31; 0.57]). For the intervention condition, paired-samples *t*-tests indicated that there were no significant differences in weighing behaviour between baseline and post-intervention, $t(1182) = 0.93$, $p = .352$, $d = 0.12$ (95% CI [-0.30; 0.54]), baseline and follow-up, $t(1417) = 1.47$, $p = .142$, $d = 0.18$ (95% CI [-0.25; 0.58]), or post-intervention and follow-up, $t(1541) = 0.43$, $p = .664$, $d = 0.05$ (95% CI [-0.38; 0.46]).

Independent-samples *t*-tests were used to assess differences between conditions at each time-point. There was no difference in weighing behaviour between the control and intervention conditions at baseline, $t(20317) = 1.11$, $p = .267$, $d = 0.24$ (95% CI [-0.19; 0.67]). A significant, medium-sized difference in weighing behaviour was found between the conditions at post-intervention, $t(8612) = 2.38$, $p = .017$, $d = 0.51$ (95% CI [0.07; 0.94]), indicating that participants in the intervention condition ($M = 82.40\%$) weighed a significantly greater proportion of their patients than those in the control condition ($M = 63.60\%$). No significant difference[4] was found at follow-up, $t(3784) = 2.08$, $p = .037$, $d = 0.45$ (95% CI [0.01; 0.88]).

Moderation analysis. Pearson product-moment correlations showed that general anxiety and intentions to weigh were not significantly correlated with changes in weighing

[4] With Bonferroni correction ($p = .017$) applied

behaviour across the time-points for the control condition (Table 7) and the intervention condition (Table 8). Therefore, no moderation analyses were performed.

Mediation analysis. Specific anxiety was not significantly correlated with the change in weighing behaviour across the time-points for the control condition (Table 7) or the intervention condition (Table 8), and so no mediation analyses were performed.

Table 7: Pearson product-moment correlations for the multiple imputation data-set control condition

		Change in weighing behaviour		
		Baseline to post-intervention (*N* = 40)	**Baseline to follow-up (*N* = 40)**	**Post-intervention to follow-up (*N* = 40)**
Time-point	**Variable**	***r (p)***	***r (p)***	***r (p)***
Baseline	**Age**	.06 (*.733*)	.08 (*.649*)	.004 (*.979*)
	Gender	.03 (*.864*)	.07 (*.699*)	.03 (*.881*)
	Level of training	-.04 (*.833*)	.13 (*.482*)	.13 (*.450*)
	Length of practice	.13 (*.486*)	-.04 (*.826*)	-.15 (*.353*)
	General anxiety	-.11 (*.511*)	-.01 (*.955*)	.10 (*.548*)
	Specific anxiety	-.17 (*.352*)	-.21 (*.267*)	.001 (*.997*)
	Intentions to weigh	-.11 (*.566*)	-.14 (*.502*)	.001 (*.995*)
	Weighing behaviour	-.32 (*.167*)	-.35 (*.178*)	.04 (*.833*)
Post-intervention	**General anxiety**	-.12 (*.513*)	-.05 (*.803*)	.08 (*.627*)
	Specific anxiety	-.06 (*.732*)	.07 (*.718*)	.11 (*.499*)
	Intentions to weigh	.01 (*.948*)	.01 (*.981*)	-.01 (*.964*)
	Weighing behaviour	.40* (*.023*)	-.12 (*.579*)	-.46* (*.005*)

*Note: * p < .05, ** p < .001*

Table 8: Pearson product-moment correlations for the multiple imputation data-set intervention condition

		Change in weighing behaviour		
		Baseline to post-intervention (*N* = 44)	**Baseline to follow-up (*N* = 44)**	**Post-intervention to follow-up (*N* = 44)**
Time-point	**Variable**	***r (p)***	***r (p)***	***r (p)***
Baseline	**Age**	.23 (*.159*)	.16 (*.333*)	-.22 (*.203*)
	Gender	-.22 (*.170*)	-.13 (*.436*)	.27 (*.120*)
	Level of training	-.08 (*.637*)	-.15 (*.385*)	-.07 (*.709*)
	Length of practice	.31 (*.065*)	.25 (*.165*)	-.28 (*.095*)
	General anxiety	-.20 (*.246*)	-.17 (*.333*)	.14 (*.434*)
	Specific anxiety	-.21 (*.238*)	-.09 (*.615*)	.29 (*.094*)
	Intentions to weigh	.05 (*.759*)	.08 (*.631*)	.03 (*.858*)
	Weighing behaviour	-.43* (*.017*)	-.47* (*.019*)	.17 (*.332*)
Post-intervention	**General anxiety**	-.15 (*.383*)	-.18 (*.319*)	.04 (*.824*)
	Specific anxiety	-.08 (*.676*)	.01 (*.968*)	.18 (*.294*)
	Intentions to weigh	.03 (*.848*)	.07 (*.713*)	.04 (*.836*)
	Weighing behaviour	.30 (*.070*)	.17 (*.343*)	-.34* (*.043*)

*Note: * $p < .05$, ** $p < .001$*

LOCF data analysis

Table 6 provides a summary of the means and standard deviations of primary outcome variables. Friedman Tests indicated that there were no significant differences in weighing behaviour across the three time-points, for the control condition, X^2 (2) = 1.24, p = .537, or the intervention condition, X^2 (2) = 1.88, p = .391.

Mann-Whitney U tests were used to assess differences between conditions at each time-point. The results indicated that there was no significant difference in weighing behaviour between conditions at baseline, U = 617.00, z = 0.97, p = .331, d = 0.26 (95% CI [-0.19; 0.72]). A small-to-medium-sized significant difference in weighing behaviour was found between conditions at post-intervention, U = 552.50, z = 2.03, p = .042, d = 0.47 (95% CI [0.02; 0.93]), suggesting that participants in the intervention condition (M = 81.34%) weighed a significantly greater proportion of their patients than those in the control condition (M = 63.87%). A small-to-medium-sized difference was also found to approach significance at follow-up, U = 564.50, z = 1.88, p = .061, d = 0.41 (95% CI [-0.04; 0.86]), with participants in the intervention condition (M = 81.43%) tending to weigh a greater proportion of their patients than those in the control condition (M = 67.08%).

Moderation analyses. Spearman Rank Order correlations showed that general anxiety and intentions to weigh were not significantly correlated with changes in weighing behaviour across the time-points for the control condition (Table 9) or the intervention condition (Table 10), and so no moderation analyses were performed.

Mediation analyses. Spearman Rank Order correlations showed that specific anxiety was not significantly correlated with changes in weighing behaviour across the time-points for the control condition (Table 9) or the intervention condition (Table 10), and so no mediation analyses were performed.

Table 9: Spearman Rank Order Correlations for the LOCF data-set control condition

		Change in weighing behaviour					
		Baseline to post-intervention		Baseline to follow-up		Post-intervention to follow-up	
Time-point	**Variable**	r_s ***(p)***	***N***	r_s ***(p)***	***N***	r_s ***(p)***	***N***
Baseline	**Age**	.28 (*.099*)	37	.27 (*.109*)	37	.03 (*.852*)	38
	Gender	-.08 (*.628*)	37	.06 (*.740)*	37	.15 (*.362*)	38
	Level of training	-.15 (*.385*)	37	.07 (*.694*)	37	.15 (*.377*)	38
	Length of practice	.22 (*.198*)	37	.08 (*.627*)	37	-.14 (*.409*)	38
	General anxiety	-.24 (*.151*)	37	-.12 (*.493*)	37	.23 (*.163*)	38
	Specific anxiety	-.20 (*.236*)	37	-.15 (*.388*)	37	.20 (*.224*)	38
	Intentions to weigh	-.17 (*.332*)	36	-.31 (*.064*)	36	-.17 (*317*)	36
	Weighing behaviour	-.29 (*.079*)	37	-.34* (*.042*)	37	-.06 (*.726*)	37
Post-intervention	**General anxiety**	-.23 (*.177*)	37	-.13 (*.435*)	37	.20 (*.220*)	38
	Specific anxiety	-.12 (*.469*)	37	-.01 (*.938*)	37	.28 (*.085*)	38
	Intentions to weigh	.01 (*.937*)	36	-.14 (*.431*)	36	-.17 (*.314*)	37
	Weighing behaviour	.29 (*.084*)	37	-.09 (*.600*)	37	-.44* (*.006*)	38

*Note: * $p < .05$, ** $p < .001$*

Table 10: Spearman Rank Order Correlations for the LOCF data-set intervention condition

		Change in weighing behaviour					
		Baseline to post-intervention		Baseline to follow-up		Post-intervention to follow-up	
Time-point	**Variable**	r_s **(p)**	***N***	r_s **(p)**	***N***	r_s **(p)**	***N***
Baseline	**Age**	.31 (*.057*)	38	.07 (*.692*)	38	-.41* (*.011*)	39
	Gender	-.04 (*.813*)	38	-.16 (*.333*)	38	.20 (*.215*)	39
	Level of training	-.03 (*.864*)	38	-.19 (*.253*)	38	-.01 (*.971*)	39
	Length of practice	.30 (*.068*)	38	.23 (*.157*)	38	-.15 (*.377*)	39
	General anxiety	-.14 (*.391*)	38	-.26 (*.120*)	38	.10 (*.551*)	39
	Specific anxiety	.04 (*.832*)	38	-.21 (*.208*)	38	.08 (*.611*)	39
	Intentions to weigh	-.15 (*.385*)	38	.24 (*.150*)	38	.12 (*.479*)	38
	Weighing behaviour	-.47* *(.003*)	38	-.28 (*.090*)	38	.19 (*.267*)	38
Post-intervention	**General anxiety**	-.02 (*.927*)	38	-.24 (*.153*)	38	-.01 (*.945*)	39
	Specific anxiety	-.02 (*.903*)	38	-.01 (*.936*)	38	.09 (*.568*)	39
	Intentions to weigh	-.03 (*.841*)	38	.33* (*.045*)	38	.22 (*.187*)	39
	Weighing behaviour	.05 (*.763*)	38	.12 (*.472*)	38	-.22 (*.186*)	39

*Note: * $p < .05$, ** $p < .001$*

Discussion

The current RCT aimed to evaluate the effectiveness of an intervention prompting therapists to form goal and implementation intentions designed to support weighing behaviour when using CBT with adults with eating disorders. The results of both a multiple imputation and an ITT-analysis using LOCF were reported to offer a more comprehensive assessment of the data and account for the limitations of both approaches. As the findings were similar, they will be discussed as one set of results.

In contrast to our predictions, the findings suggested that weighing behaviour did not change over time in either the control or intervention conditions. The finding that the intervention did not increase the proportion of patients who therapists weighed is in contrast to previous reviews which suggest that forming implementation intentions has a medium-to-large-sized effect on behaviour change (Gollwitzer & Sheeran, 2006; Toli, Webb, & Hardy, 2016).

However, differences were found between the conditions at the post-intervention time-point, with evidence that therapists in the intervention group weighed a higher percentage of patients than therapists in the control group. This difference appeared to be due to a decrease in weighing behaviour in the control condition, rather than an increase in weighing behaviour in the intervention condition, which might be interpreted as being in line with research identifying therapist drift when using CBT with adults with eating disorders (Waller, 2009). If that is the case, then these findings would provide further evidence for the need for research into interventions to support therapists to adhere to CBT protocols. The difference in weighing behaviour between the conditions remained marginally significant at the follow-up time-point, suggesting therapists in the control condition continued to weigh their patients less frequently at the end, compared to at the start of the study. Taken together, these results indicate that forming goal-intentions and implementation intentions might not help to

increase therapists' weighing behaviour, but might help to protect against therapist drift, as observed among therapists in the control condition.

Why was there no effect of forming implementation intentions over time?

Therapists in the current sample reported that they weighed a high percentage of patients at baseline (74%, across conditions), which might indicate that therapists found it easy to weigh their patients. There may have been no effect of forming implementation intentions because therapists were already performing the behaviour they intended to perform (i.e., weighing their patients), and so there was very little room for improvement (i.e., ceiling effects; Dewitte, Verguts, & Lens, 2003). Previous research has shown that when the task is easy, goal intentions alone have been shown to be sufficient for goal attainment (e.g., Gollwitzer & Brandstätter, 1997; Wieber, Odenthal, & Gollwitzer, 2010). Therapists in the current study reported strong baseline intentions to weigh their patients (6.34 out of 7, across conditions), a commitment that might mean that they already achieved this goal without forming implementation intentions. Furthermore, it is unlikely that the intervention would result in therapists weighing 100% of their patients as implementation intentions do not result in perfect execution of desired behaviour every time those behaviours are cued (Saddawi-Konefka, Schumacher, Baker, Charnin, & Gollwitzer, 2016). Alternatively, finding no effect of the intervention might be due to a lack of power in the study. A large effect was predicted based on previous research on the effects of forming if-then plans in a range of contexts (review by Gollwitzer & Sheeran, 2006) and evidence of a large-sized effect of combined motivational and volitional interventions on behaviour change (e.g., Fritzsche, Schlier, Oettingen, & Lincoln, 2016; Milne, Orbell, & Sheeran, 2002). However, little previous research has investigated the effects of a goal intention and implementation intention intervention for clinicians, and the current study recruited a sample of therapists supporting others to change their behaviour, rather than participants attempting to change their own

behaviour. Therefore, in retrospect, the present study might have lacked sufficient power to detect the effect of the intervention on clinicians, which might have been of small- or medium-magnitude, rather than large.

Another possibility is that therapists who had more negative views towards weighing patients might have chosen not to participate in the study, or negative views might have impacted on their engagement with the intervention. In the current study participants were self-selecting (i.e., they volunteered to take part in the research in response to a prompt) and there is evidence that therapists who have negative views towards CBT manuals are less likely to adhere to them (Waller & Turner, 2016; Wiborg et al., 2012). Although no differences in therapists' intentions to weigh patients were found between completers and non-completers, it might have been helpful to explore therapists' attitudes in greater detail to understand additional potential moderators of the intervention or barriers to using it. Furthermore, participants are more likely to respond to self-report questionnaires when they trust that the expected rewards of responding outweigh the anticipated costs (Dillman, 2007; Fan & Yan, 2010). If therapists did not view weighing more patients as rewarding, then they might have been less likely to engage.

Is it possible to explain when and why forming implementation intentions has an effect?

Levels of general anxiety or therapists' intentions to weigh were not associated with changes in weighing behaviour, suggesting that these variables did not moderate the effect of the intervention. Similarly, specific anxiety was not associated with changes in weighing behaviour. These results suggest that the effect of the intervention might have similar effects regardless of therapists' anxiety or motivation. However, it is important to note that therapists reported very low levels of general anxiety and specific anxiety, and very strong intentions to weigh patients. Therefore, floor and ceiling effects might have limited the potential to identify moderating and mediating factors.

Strengths, limitations, and recommendations for future research

A significant strength of the current research lies in the RCT design. The inclusion of a control group allowed for confounding factors (e.g., the impact of participating in a study) to be controlled for, thereby increasing the validity of the research. The randomisation of participants along with allocation concealment minimised the risk of selection-bias because groups should not differ in a systematic way and covariates can be controlled for (Suresh, 2011). Additionally, the current study did not apply stringent exclusion criteria, and thus recruited a range of different professionals with varying years of experience, which allowed for a more generalisable sample.

The ecological validity of the study was improved through participants implementing 'if-then' plans in their current place of work. Therefore, they faced real-world barriers to openly weighing their patients and potentially drew upon more realistic solutions. However, this design meant that participants might have struggled to fully engage in the intervention due to work pressures and time-constraints (Douville, Godin, Legare & Germain, 2014), which could have contributed to the high attrition rates seen in the current study (Rübsamen, Akmatov, Castell, Karch, & Mikolajczyk, 2017). Sixty-four percent of therapists who completed baseline measures did not complete the final follow-up measures, which may have weakened the overall power of the study. Future research could use qualitative methods to explore the reasons for attrition or lack of engagement. Furthermore, research could investigate the use of reminders to increase response rates, as recommended by Fan and Yan (2010). With these challenges in mind, it might have also been beneficial to conduct a feasibility study to investigate potential difficulties with recruitment and attrition, and/or conduct a pilot study to identify limitations of the methodology prior to conducting a more formal RCT, as "pilot and feasibility studies can provide sufficient methodological evidence about the design, planning and justification of a trial" (Blatch-Jones, Pek, Kirkpatrick, &

Ashton-Key, 2018, p.1). It is also worth noting that, participants were not asked to report their nationality, and so the potential for differential attrition between participants from different countries could not be investigated. Therapists in different countries might implement CBT for eating disorders differently and/or respond to the intervention differently. Therefore, future research could investigate whether CBT practice and the acceptability and impact of an implementation intentions intervention differs between therapists in different countries.

Attempts were made to limit the impact of attrition bias, including the use of multiple imputation and LOCF analysis, and no significant differences were found between completers and non-completers. However, the use of LOCF might have added bias to the findings, as single imputation techniques often underestimate the variability in the data (Jakobsen, Gluud, Wettersley, & Winkel, 2017; Salkind, 2010). However, when the goal of treatment is to increase the score on a scale (i.e., increase the percentage of patients openly weighed), LOCF carries forward a smaller improvement, and is often a more conservative estimate of the effect because it assumes no sustained change after loss to the research (Salkind, 2010). Multiple imputation is a more robust imputation method, and allowed for the variability in the data to be accounted for in the missing values (Sterne et al., 2009). Multiple imputation followed by non-parametric statistics would have provided the most reliable result, as both missing data and non-normally distributed data would have been managed more effectively. Unfortunately, SPSS does not support this combination of analyses. Consequently, reporting both statistical procedures aimed to offer a more comprehensive assessment of the data and account for the limitations of both approaches – both approaches produced similar findings.

An additional limitation is that recruiting participants using professional contacts and opportunistic snowballing may have introduced selection bias. For example, ceiling effects

might have occurred as participants might have been more likely to participate if they were confident in their CBT practise and adhered more closely to the protocol, or therapists using CBT with adults with eating disorders might be too similar in their interests and actions and thus have little room to improve or change. Thus, ceiling effects and existing protocol recommendations might mean that the opportunity to show improvements was too limited in this study. Future research could explore the effects of if-then planning for therapists using a more diverse range of therapy models (e.g., counselling, Family Based Treatment, Dialectical Behaviour Therapy). Furthermore, participants were in part recruited following CBT training events, and so may have been more likely to have recently been prompted to openly weigh patients every session. Consequently, selection bias may have contributed to the relatively high percentage of patients who therapists reported weighing at baseline, potentially leading to ceiling effects. However, opportunistic snowball sampling and the use of large professional contact lists meant reaching a wider population, which was intended to mitigate the effect of recruiting from training events.

Finally, the lack of an increase in weighing behaviour in the intervention group might be explained in part by self-report measures, which were used to assess all outcome variables. Self-report measures rely on the honesty of participants. Although measures were taken to reduce the risk of demand characteristics (e.g., including filler questions to disguise the true aim of the study), participants may have reported that their weighing behaviour and intentions were more in line with protocols than they actually were. Alternatively, high scores on weighing behaviour might be as a result of clinicians overestimating their level of competence (Brosan, Reynolds, & Moore, 2008; Walfish, McAlister, O'Donnell, & Lambert, 2012). Future research might benefit from using more objective measures of therapists' weighing behaviours, to minimise the risk of bias from self-report measures.

Clinical implications

Given the psychological, social, and financial impact of eating disorders, it is vital to identify effective ways to provide treatment. However, therapist drift in CBT for adults with eating disorders not only means that it is difficult to evaluate the effectiveness of treatments if therapists are not implementing the treatment appropriately, but can also have a negative impact on patient outcomes (Waller, 2009; Waller & Turner, 2016; Wiborg et al., 2012). The current findings indicate that, regardless of a therapist's levels of anxiety or motivation to weigh patients with eating disorders, forming implementation intentions may be beneficial in protecting against therapist drift. Therapists might therefore benefit from drawing upon this simple strategy to aid their adherence to CBT protocols for adults with eating disorders. Forming implementation intentions also represents a relatively quick and cheap intervention, which can be disseminated to large groups (Hagger & Luszczynska, 2014; Prestwich, Sheeran, Webb, & Gollwitzer, 2015). In such circumstances, even small effects might warrant the investment of time and effort.

Conclusions

The current study evaluated the effect of an intervention that prompted therapists to form implementation intentions, in order to support weighing behaviour when using CBT with adults with eating disorders. Participants in the intervention condition did not demonstrate a significant increase in their weighing behaviour, which is in contrast to previous research suggesting a medium-to-large-sized effect of implementation intentions on behaviour change. However, this might be due to ceiling effects and participant characteristics (e.g., finding it easy to weigh patients with eating disorders or negative attitudes towards weighing patients with eating disorders), or aspects of the methodology (e.g., the use of self-report measures). In line with the concept of therapist drift, the control group showed a decrease in weighing behaviour compared to the intervention group.

Taken together, these results suggest that although there was no positive effect of the intervention, forming implementation intentions may protect against therapist drift. There was no evidence that therapists' levels of anxiety or the strength of their intentions to weigh patients influenced the effect of the intervention, suggesting that the intervention may benefit therapists regardless of their anxieties or motivation. These findings offer further support for the use of implementation intentions to support therapeutic practice.

References

Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology, 68,* 1-9. doi: 10.1037//0022-006X.68.2.

Addis, M. E., & Waltz, J. (2002). Implicit and untested assumptions about the role of psychotherapy treatment manuals in evidence-based mental health practice. *Clinical Psychology: Science and Practice, 9,* 421-424. doi: 10.1093/clipsy.9.4.42.

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.

Arcelus, J., Mitchell, A. J., Wales, J., & Nielsen, S. (2011). Mortality rates in patients with anorexia nervosa and other eating disorders: a meta-analysis of 36 studies. *Archives of General Psychiatry, 68,* 724-731. doi: 2011.74.archgenpsychiatry/1001.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-215. doi: 10.1037/0033-295X.84.2.191

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *The cognitive therapy of depression.* New York, NY: Guilford Press.

Becker, A. E., Grinspoon, S. K., Klibanski, A., & Herzog, D. B. (1999). Eating disorders. *New England Journal of Medicine, 340,* 1092-1098. doi: 10.1056/NEJM199904083401407.

Blatch-Jones, A. J., Pek, W., Kirkpatrick, E., & Ashton-Key, M. (2018). Role of feasibility and pilot studies in randomised controlled trials: a cross-sectional study. *BMJ Open, 8*, 1-9. doi: 10.1136/bmjopen-2018-022233.

Bodner, T. E. (2008) "What improves with increased missing data imputations?" *Structural Equation Modeling*: *A Multidisciplinary Journal* 15: 651-675. doi: 10.1080/10705510802339072.

Brosan, L., Reynolds, S., & Moore, R. G. (2008). Self-evaluation of cognitive therapy performance: do therapists know how competent they are? *Behavioural and Cognitive Psychotherapy, 36*, 581. doi: 10.1017/S1352465808004438.

Brown, A., Mountford, V., & Waller, G. (2013). Therapeutic alliance and weight gain during cognitive behavioural therapy for anorexia nervosa. *Behaviour Research and Therapy, 51*, 216-220. doi: 10.1016/j.brat.2013.01.008.

Carleton, R. N., Norton, P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: a short version of the intolerance of uncertainty scale. *Journal of Anxiety Disorders, 21,* 105-117. doi: 10.1016/j.janxdis.2006.03.014.

Casper, E. S. (2008). Using implementation intentions to teach practitioners: changing practice behaviors via continuing education. *Psychiatric Services, 59,* 747-752. doi: 10.1176/appi.ps.59.7.747.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159. doi: 10.1037/0033-2909.112.1.155.

Cowdrey, N. D., & Waller, G. (2015). Are we really delivering evidence-based treatments for eating disorders? How eating-disordered patients describe their experience of cognitive therapy. *Behaviour Research and Therapy, 75,* 72-77. doi: 10.1016/j.brat.2015.10.009.

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: perceptions of burden. *Social Science Computer Review, 19*, 146–162. doi: 10.1177/089443930101900202.

Crits-Cristoph, P., Connolly Gibbons, M. B., & Hearon, B. (2006). Does the alliance cause good outcome? Recommendations for future research on the alliance. *Psychotherapy: Theory, Research, Practice, Training, 43,* 280-285. doi: 10.1037/0033-3204.43.3.280.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334. doi: 10.1007/bf02310555.

Crow, S. J., Peterson, C. B., Swanson, S. A., Raymond, N. C., Specker, S., Eckert, E. D., & Mitchell, J. E. (2009). Increased mortality in bulimia nervosa and other eating disorders. *American Journal of Psychiatry, 166,* 1342-1346. doi: 10.1176/appi.ajp.2009.09020247.

Cukrowicz, K. C., Timmons, K. A., Sawyer, K., Caron, K. M., Gummelt, H. D., & Joiner, T. E. (2011). Improved treatment outcome associated with the shift to empirically supported treatments in an outpatient clinic is maintained over a ten-year period. *Professional Psychology: Research and Practice, 42,* 145-152. doi: 10.1037/a0021937.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgement. *Science, 243,* 1668-1674. doi: 10.1126/science.2648573.

Dewitte, S., Verguts, T., & Lens, W. (2003). Implementation intentions do not enhance all types of goals: the moderating role of goal difficulty. *Current Psychology, 22,* 73-89. doi: 10.1007/s12144-003-1014-6.

Dillman, D. A. (2007). *Mail and internet surveys: The tailored design method. 2007 update with new internet, visual, and mixed-mode guide* (2nd ed.). New York, NY: John Wiley & Sons.

Douville F., Godin G., Légaré F., & Germain, M. (2014). Implementation intentions as a strategy to increase the notification rate of potential ocular tissue donors by nurses: a clustered randomized trial in hospital settings. *Nursing Research and Practice, 2014,* 1-7. doi: 10.1155/2014/921263.

Elliot, M. A., & Armitage, C. J. (2006). Effects of implementation intentions on the self-reported frequency of drivers' compliance with speed limits. *Journal of Experimental Psychology Applied, 12,* 108-117. doi: 10.1037/1076-898X.12.2.108.

Fairburn, C, G. (2008). *Cognitive behavior therapy and eating disorders.* New York, NY: The Guildford Press.

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: a systematic review. *Computers in Human Behavior, 26,* 132-139. doi: 10.1016/j.chb.2009.10.015.

Folke, S., Daniel, S. I. F., Gondan, M., Lunn, S., Taekker, L., & Poulsen, S. (2017). Therapist adherence is associated with outcome in cognitive-behavioral therapy for bulimia nervosa. *Psychotherapy, 54,* 195-200. doi: 10.1037/pst0000107.

Forbush, K. T., Richardson, J. H., & Bohrer, B. K. (2015). Clinicians' practices regarding blind versus open weighing among patients with eating disorders. *International Journal of Eating Disorders, 48,* 905-911. doi: 10.1002/eat.22369.

Fritzsche, A., Schlier, B., Oettingen, G., & Lincoln, T. M. (2016). Mental contrasting with implementation intentions increases goal-attainment in individuals with mild to moderate depression. *Cognitive Therapy and Research*, *40,* 1-8. doi: 10.1007/s10608-015-9749-6.

Godin, G., Bélanger-Gravel, A., Amireault, S., Gallani, M. C., Vohl, M. C., & Pérusse, L. (2010). Effect of implementation intentions to change behaviour: moderation by intention stability. *Psychology Reports, 106*, 147-159. doi: 10.2466/PR0.106.1.147-159.

Godin, G., Sheeran, P., Conner, M., & Germain, M. (2008). Asking questions changes behavior: mere measurement effects on frequency of blood donation. *Health Psychology, 27*, 179-184. doi: 10.1037/0278-6133.27.2.179.

Gollwitzer, P. M. (1993). Goal achievement: the role of intentions. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (pp. 141-185). Chichester, UK: Wiley.

Gollwitzer, P. M. (1999). Implementation intentions: strong effects of simple plans. *American Psychologist, 54*, 493-503. doi: 10.1037/0003-066X.54.7.493.

Gollwitzer, P. M., Bayer, U. C., & McCulloch, K. C. (2005). The control of the unwanted. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 485-515). New York, NY: Oxford University Press.

Gollwitzer, P. M., & Brandstätter, V. (1997). Implementation intentions and effective goal pursuit. *Journal of Personality and Social Psychology, 73*, 186-199. doi: 10.1037/0022-3514.73.1.186.

Gollwitzer, P. M., Mayer, D., Frick, C., & Oettingen, G. (2018). Promoting the self-regulation of stress in health care providers: an internet-based intervention. *Frontiers in Psychology, 9,* 1-11. doi: 10.3389/fpsyg.2018.00838.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: a meta-analysis of effects and processes. *Advances in Experimental Social Psychology, 38*, 69-119. doi:10.1016/S0065-2601 (06)38002-1.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8,* 206-213. doi: 10.1007/s11121-007-0070-9.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological Assessment, 12,* 19-30. doi: 10.1037//1040-3590.12.1.19.

Hagger, M. S., & Luszczynska, A. (2014). Implementation intention and action planning interventions in health contexts: state of the research and proposals for the way

forward. *Applied Psychology: Health and Well-being, 6,* 1-47. doi: 10.1111/aphw.12017.

Hay, P. P., Bacaltchuk, J., Stefano, S., & Kashyap, P. (2009). Psychological treatments for bulimia nervosa and binging. *Cochrane Database of Systematic Reviews, 4,* 1-170. doi: 10.1002/14651858.CD000562.pub3.

Hayes, A. F. (2013). *Methodology in the social sciences. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* New York, NY: Guilford Press.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials: a practical guide with flowcharts. *BMC Medical Research Methodology, 6,* 162-172. doi: 10.1186/s12874-017-0442-1.

Kass, A. E., Kolko, R. P., & Wilfley, D. E. (2013). Psychological treatments for eating disorders. *Current Opinion in Psychiatry, 26,* 549-555. doi: 10.1097/YCO.0b013e328365a30e.

Khawaja, N. G., & Yu, L. N. H. (2010). A comparison of the 27-item and 12-item intolerance of uncertainty scales. *Clinical Psychologist, 14,* 97-106. doi: 10.1080/13284207.2010.502542.

King's Fund (2008). *Paying the price: the cost of mental health care in England to 2026.* Retrieved from https://www.kingsfund.org.uk/sites/files/kf/Paying-the-Price-the-cost-of-mental-health-care-England-2026-McCrone-Dhanasiri-Patel-Knapp-Lawton-Smith-Kings-Fund-May-2008_0.pdf

Lenhard, W., & Lenhard, A. (2016). *Calculation of effect sizes.* Dettelbach, Germany: Psychometrica. Available at https://www.psychometrica.de/effect_size.html. doi: 10.13140/RG.2.1.3478.4245.

Linardon, J., Wade, T. D., De la Piedad Garcia, X., & Brennan, L. (2017). The efficacy of cognitive-behavioral therapy for eating disorders: a systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology, 85,* 1080-1094. doi: 10.1037/ccp0000245.

Lippke, S., Wiedemann, A. U., Ziegelmann, J. P., Reuter, T., & Schwarzer, R. (2009). Self-efficacy moderates the mediation of intentions into behavior via plans. *American Journal of Health Behavior, 33*, 521-529. doi: 10.5993/AJHB.33.5.5.

Loucas, C. E., Fairburn, C. G., Whittington, C., Pennant, M. E., Stockton, S., & Kendall, T. (2014). E-therapy in the treatment and prevention of eating disorders: A systematic review and meta-analysis. *Behaviour Research and Therapy, 63,* 122-131. doi: 10.1016/j.brat.2014.09.011.

Luszczynska, A., Cao, D. S., Mallach, N., Pietron, K., Mazurkiewicz, M., & Schwarzer, R. (2010). Intentions, planning, and self-efficacy predict physical activity in Chinese and Polish adolescents: two moderated mediation analyses. *International Journal of Clinical and Health Psychology, 10*, 265-278.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press.

Meyer, J. M., Farrell, N. R., Kemp, J. J., Blakey, S. M., & Deacon, B. J. (2014). Why do clinicians exclude anxious clients from exposure therapy? *Behaviour Research and Therapy, 54*, 49-53. doi:10.1016/j.brat.2014.01.004.

Milne, S., Orbell, S., & Sheeran, P. (2002). Combining motivational and volitional interventions to promote exercise participation: protection motivation theory and implementation intentions. *British Journal of Health Psychology, 7,* 163-184. doi: 10.1348/135910702169420.

Mulkens, S., de Vos, C., de Graaff, A., & Waller, G. (2018). To deliver or not to deliver Cognitive Behavioral Therapy for eating disorders: replication and extension of our understanding of why therapists fail to do what they should do. *Behaviour Research and Therapy, 106,* 57-63. doi: 10.1016/j.brat.2018.05.004.

National Institute of Health and Care Excellence. (2017). *Eating disorders: recognition and treatment [NG69].* Retrieved from https://www.nice.org.uk/guidance/ng69

Olatunji, B. O., Deacon, B. J., & Abramowitz, J. S. (2009). The cruelest cure? Ethical issues in the implementation of exposure-based treatments. *Cognitive and Behavioral Practice, 16*, 172-180. doi: 10.1016/j.cbpra.2008.07.003.

Orne, M. T. (1962). On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776-783. doi: 10.1037/h0043424.

Presseau, J., Johnston, M., Heponiemi, T., Elovainio, M., Francis, J. J., Eccles, M. P., Steen, N., Hrisos, S., Stamp, E., Grimshaw, J. M., Hawthorne, G., & Sniehotta, F. F. (2014). Reflective and automatic processes in health care professional behaviour: a dual process model tested across multiple behaviours. *Annals of Behavioral Medicine, 48,* 347-358. doi: 10.1007/s12160-014-9609-8.

Prestwich, A., Sheeran, P., Webb, T. L., & Gollwitzer, P. (2015). Implementation intentions. In M. Conner, & P. Norman (Eds.), *Predicting health behavior* (3rd ed., pp. 321-357). New York, NY: McGraw Hill.

Rubin, D. B. (1987). *Multiple imputation for non-response in surveys.* New York, NY: John Wiley & Sons. doi: 10.1002/9780470316696.

Rübsamen, N., Akmatov, M. K., Castell, S., Karch, A., & Mikolajczyk, R. T. (2017). Factors associated with attrition in a longitudinal online study: results from the HaBIDS

panel. *BMC Medical Research Methodology, 17*. 132-143. doi: 10.1186/s12874-017-0408-3.

Saddawi-Konefka, D., Schumacher, D. J., Baker, K. H., Charnin, J. E., & Gollwitzer, P. M. (2016). Changing physician behavior with implementation intentions: closing the gap between intentions and actions. *Academic Medicine, 20,* 1-6. doi: 10.1097/ACM.0000000000001172.

Salkind, N. J. (2010). *Encyclopedia of research design.* Thousand Oaks, CA: SAGE Publications, Inc.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. British Medical Journal*, 340*, c332. doi: 10.1136/bmj.c332.

Sheeran, P. (2002). Intention-behavior relations: a conceptual and empirical review. *European Review of Social Psychology, 12*, 1-36. doi: 10.1080/14792772143000003.

Sheeran, P., Milne, S., Webb, T. L., & Gollwitzer, P. (2005). Implementation intentions and health behaviors. In M. Conner, & P. Norman (Eds.), *Predicting health behavior* (2nd ed., pp. 276-323). Buckingham, UK: Open University Press/McGraw-Hill.

Sheeran, P., & Orbell, S. (1999). Implementation intentions and repeated behaviour: augmenting the predictive validity of the theory of planned behaviour. *European Journal of Social Psychology, 29,* 349-369. doi: 10.1002/(SICI)1099-0992(199903/05)29:2/33.0.CO;2-Y.

Sheeran, P., & Webb, T. L. (2016). The intention-behaviour gap. *Social and Personality Psychology Compass, 10*, 503–518. doi: 10.1111/spc3.12265.

Sheeran, P., Webb, T. L., Gollwitzer, P. M. (2005). The interplay between goal intentions and implementation intentions. *Personality and Social Psychology Bulletin, 31,* 87-98. doi: 10.1177/0146167204271308.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Research Methods & Reporting, 338,* b2393. doi: 10.1136/bmj.b2393.

Suresh, K. P. (2011). An overview of randomization techniques: an unbiased assessment of outcome in clinical research. *Journal of Human Reproductive Sciences, 4,* 8-11. doi: 10.4103/0974-1208.82352.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*, 894-904. doi: 10.1037/0022-006X.73.1.168.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2,* 53-55. doi: 10.5116/ijme.4dfb.8dfd.

Tobin, D. L., Banker, J. D., Weisberg, L., & Bowers, W. (2007). I know what you did last summer (and it was not CBT): a factor analytic model of international psychotherapeutic practice in the eating disorders. *International Journal of Eating Disorders, 40,* 754-757. doi: 10.1002/eat.20426.

Toli, A., Webb, T. L., & Hardy, G. E. (2016). Does forming implementation intentions help people with mental health problems to achieve goals? A meta-analysis of experimental studies with clinical and analogue samples. *British Journal of Clinical Psychology, 55*, 69-90. doi:10.1111/bjc.12086.

Tomasone, J. R., Ginis, K. A. M., Estabrookes, P. A., & Domenicucci, L. (2014). 'Changing Minds': determining the effectiveness and key ingredients of an educational intervention to enhance healthcare professionals' intentions to prescribe physical

activity to patients with physical disabilities. *Implementation Science, 9,* 1-11. doi: 10.1186/1748-5908-9-30.

Triandis, H. C. (1980). Values, attitudes, and interpersonal behaviour. In H. E. Howe, Jr. & M. Page (Eds.), *Nebraska symposium of motivation* (Vol. 27, pp. 195-259). Lincoln, NE: University of Nebraska Press.

Trivasse, H. (2019). *Does training clinicians in exposure therapy for anxiety disorders change their knowledge, attitudes, or practice? A meta-analysis* (Unpublished doctoral dissertation). University of Sheffield, UK.

Turner, H., Bryant-Waugh, R., & Marshall, E. (2015). The impact of early symptom change and therapeutic alliance on treatment outcome in cognitive-behavioural therapy for eating disorders. *Behaviour Research and Therapy, 73,* 165-169. doi: 10.1016/j.brat.2015.08.006.

Turner, H., Tatham, M., Lant, M., Mountford, V. A., & Waller, G. (2014). Clinicians' concerns about delivering cognitive-behavioural therapy for eating disorders. *Behaviour Research and Therapy, 57,* 38-42. doi: 10.1016/j.brat.2014.04.003.

van Osch, L., Lechner, L., Reubsaet, A., Wigger, S., & de Vries H. (2008). Relapse prevention in a national smoking cessation contest: effects of coping planning. *British Journal of Health Psychology, 13*, 525-535. doi: 10.1348/135910707X224504.

von Ranson, K. M., Wallace, L. M., & Stevenson, A. (2013). Psychotherapies provided for eating disorders by community clinicians: infrequent use of evidence-based treatment. *Psychotherapy Research, 23,* 333-343. doi: 10.1080/10503307.2012.735377.

Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports, 110,* 639–644. doi: 10.2466/02.07.17.PR0.110.2.639-644.

Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour Research and Therapy, 47,* 119-127. doi: 10.1016/j.brat.2008.10.018.

Waller, G., Cordery, H., Corstorphine, E., Hinrichsen, H., Lawson, R., Mountford, V., & Russell, K. (2007). *Cognitive behavioural therapy for eating disorders: A comprehensive treatment guide*. Cambridge, UK: Cambridge University Press.

Waller, G., & Mountford, V. (2015). Weighing patients within cognitive-behavioural therapy for eating disorders: how, when and why. *Behaviour Research and Therapy, 70,* 1-10. doi: 10.1016/j.brat.2015.04.004.

Waller, G., Mountford, A., Tatham, M., Turner, H. Gabriel, C., & Webber, R. (2013). Attitudes towards psychotherapy manuals among clinicians treating eating disorders. *Behaviour Research and Therapy, 51*, 840-844. doi:10.1016/j.brat.2013.10.004.

Waller, G., Stringer, H., & Meyer, C. (2012). What cognitive behavioural therapy techniques do therapists report using when delivering cognitive behavioural therapy for the eating disorders? *Journal of Consulting and Clinical Psychology, 80,* 171-175. doi: 10.1037/a0026559.

Waller, G., & Turner, H. (2016). Therapist drift redux: why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behaviour Research and Therapy, 77,* 129-137. doi: 10.1016/j.brat.2015.12.005.

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin, 132,* 249-268. doi: 10.1037/0033-2909.132.2.249.

White, I. R., Royston, P., & Wood, A. M. (2011) Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine*, *30*, 377-399. doi: 10.1002/sim.4067.

Wiborg, J. F., Knoop, H., Frank, L. E., & Bleijenberg, G. (2012). Towards an evidence-based treatment model for cognitive behavioral interventions focusing on chronic fatigue syndrome. *Journal of Psychosomatic Research, 72,* 399-404. doi: 10.1016/j.jpsychores.2012.01.018.

Wieber, F., Odenthal, G., & Gollwitzer, P. (2010). Self-efficacy feelings moderate implementation intention effects. *Self and Identity, 9,* 177-194. doi: 10.1080/15298860902860333.

Wiedemann, A. U., Schuz, B., Sniehotta, F., Scholz, U., & Schwarzer, R. (2009). Disentangling the relation between intentions, planning, and behaviour: a moderated mediation analysis. *Psychology & Health, 24*, 67-79. doi: 10.1080/08870440801958214.

Wilson, G. T. (1996). Manual-based treatments: the clinical application of research findings. *Behaviour Research and Therapy, 34,* 295-314. doi: 10.1016/0005-7967(95)00084-4.

Wittes, J. (2002). Sample size calculations for randomized controlled trials. *Epidemiologic Reviews, 24,* 39-53. doi: doi.org/10.1093/epirev/24.1.39.

Appendix A

Ethical approval letter from the University of Sheffield's Department of Psychology

Research Ethics Committee

Downloaded: 08/06/2018 Approved: 31/01/2018

Heidi Trivasse
Registration number: 160124488 Psychology
Programme: Clinical Psychology

Dear Heidi

PROJECT TITLE: Can forming implementation intentions help therapists to adhere to Cognitive Behavioural Therapy protocols regarding weighing adult patients with eating disorders?
APPLICATION: Reference Number 017422

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 31/01/2018 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

University research ethics application form 017422 (dated 05/01/2018). Participant information sheet 1038627 version 1 (05/01/2018).
Participant consent form 1038630 version 1 (05/01/2018).
Participant consent form 1038631 version 1 (05/01/2018).

The following optional amendments were suggested:

The information sheets (consent forms) should include information about why the participant has been contacted, how long the interview will take and the university complaints procedure

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely

Thomas Webb Ethics Administrator Psychology

Appendix B

Debrief email

Thank you for taking the time to complete this study.

Sometimes in psychological research it is necessary to not tell people about the true purpose of a study at the beginning, as in doing so it may affect how a person responds to the questions asked. We told you that the research was about the use of Cognitive Behaviour Therapy (CBT) as a treatment approach when working with adults with eating disorders. This is true, but the specific purpose of the current study was to understand whether forming specific 'if-then' plans can increase therapists' adherence to the CBT evidence-based protocol – weighing clients. To investigate this idea, we asked some therapists to form 'if-then' plans and then compared their responses to a group of therapists who continued practice as usual (without goal setting or forming implementation intentions). The research was also interested in whether therapists' levels of anxiety (e.g., about a patients reaction to being weighed) explain why therapists may not weigh clients – this is why we asked you questions about how you feel about weighing your patients with eating disorders.

If you have any further questions about this research, then please contact: htrivasse1@sheffield.ac.uk.

As this is an ongoing study, other professionals working with people with eating disorders may be asked to get involved. It is therefore important for the integrity of the study that you do not talk about the study's true purpose, until all data is collected. Data is expected to be collected by September 2018.

Your participation in this research is very important, and we hope that you understand why we did not tell you the full purpose of the study at the outset. It is not uncommon, however, to feel dissatisfied in having participated in research where the intentions were not

fully stated. If you no longer want your data to be used for this research, you can request that your responses be withdrawn by emailing htrivasse1@sheffield.ac.uk.

If you would like to know more about the findings of this research, then please email htrivasse1@sheffield.ac.uk.

Thank you once again for taking part.

The project is supervised by:

Professor Glenn Waller, Clinical Psychology Unit, University of Sheffield

Dr Thomas Webb, Department of Psychology, University of Sheffield

Appendix C

Adverse event and complaints form

Adverse Incident/Complaint Form (Psychology Version) for health care research projects that the University of Sheffield is the research governance sponsor of

This report form is for use if and when an adverse event incident occurs or a complaint is made relating to a health care research project where the University is the research governance sponsor. It should be completed by the Principal (or Chief) Investigator of the project and agreed with the Chair of the Ethics Committee or if a Clinical Unit project with the Director of Research Training. It will then be discussed with the Head of Department.

Guidance notes are included at the end of the report form (boxes on the form can be expanded).

1. Research Project Title:	
2. 6 digit URMS number (if applicable):	
3. Principal/Chief Investigator:	
4. Supervisor/s:	
5. Who initially discovered the adverse event/Complaint?	
6. When was the adverse event/complaint reported to the Principal/Chief Investigator?	
7. When was the adverse event/complaint reported to the Head of Department/School?	
8. When did the adverse event/complaint actually occur?	
9. Where did it happen?	

10. What actually happened and what was the impact of the adverse event/complaint?
11. Why did the adverse event/complaint occur?

12. Describe what action(s) have been taken to address the impact of this specific adverse event/complaint:	
13. Describe what action(s) have been taken or are planned to limit the risk of a similar event/complaint re-occurring (add any general notes here to qualify the information given elsewhere in the report):	
Agreed and authorised by:	
Name of Principal/Chief Investigator: *Insert name here* Signature:	Date: *insert date here*
Name of Head of Chair of Ethics Committee/Director of Research Training: *Insert name here* Signature:	Date: *insert date here*

Guidance Notes:

1.
Adverse events/complaints should be reported to the Head of Department/School as soon as possible and normally within **5 working days**. If the time exceeds this, this should be a consideration in 13.

2.
Once complete, this report should be kept in the project's site file for reference and a copy sent to Research and Innovation Services, New Spring House, 231 Glossop Road marked for the attention of the Head of the Planning and Business Support Section (Mrs Deborah McClean).

3.
Advice and guidance on completion of the report, analysis of the event and potential actions can be obtained from Research and Innovation Services (Richard Hudson: ext. 21448).

4.
An 'adverse event' is an unexpected event that includes, but is broader than, unintended errors and mistakes which arise as a result of research activity and result in one or more research participants having symptoms or being caused physical or psychological harm or serious distress.

Examples of this include:

- A human participant has an adverse reaction to a drug treatment, the use of which had been approved by a Research Ethics Committee.
- An invasive instrument is used incorrectly, the use of which had been approved by a Research Ethics Committee, and the human participant suffers harm or has an extended stay in hospital.

- A human participant is asked a series of questions regarding his/her sex life, a line of questioning that a Research Ethics Committee approved. However, for the interviewee, the questions revive painful memories of being abused as a child and the interviewee suffers serious distress such as to warrant therapy.

5.

A 'complaint' is any approach made by a research participant to the researcher, their supervisor or collaborator with respect to the conduct of the study

Appendix D

Invitation to participate email

You are invited to take part in an online study, which is interested in the use of Cognitive Behaviour Therapy (CBT) as a treatment approach when working with adults with eating disorders. Our aim is to improve service provision for patients with eating disorders and improve patient outcomes.

The online survey will take up to 30 minutes to complete, and will ask questions about you, your beliefs, and different aspects of your therapeutic approach. You will be contacted again over the forthcoming weeks, and asked to complete another shorter survey, which should take no longer than 15 minutes to complete.

Participation in this study is voluntary and confidential, and you may choose to withdraw from the study at any time. Your contributions to this research area are greatly valued, but participation will involve your commitment to monitoring a small number of therapy behaviours over a short period of time. A full debrief will be available following participation. If you would like to take part, please click on the following link:

www.qualtricslink....

This research has received ethical approval from the University of Sheffield's Department of Psychology Ethics Committee.

Thank you in advance for your time, it is much appreciated.

If you would like any further information about the study, then please contact: htrivasse1@sheffield.ac.uk (Heidi Trivasse, Trainee Clinical Psychologist).

The project is supervised by:

Professor Glenn Waller, Clinical Psychology Unit, University of Sheffield

Dr Thomas Webb, Department of Psychology, University of Sheffield

Appendix E

Intolerance of Uncertainty Scale – Short form

ASDP IUS-12 1 of 1

Initials/ID #:_____

Date:_____

Intolerance of Uncertainty Scale – Short Form

(Carleton, Norton, & Asmundson, 2007)

Please circle the number that best corresponds to how much you agree with each item.

	Not at all characteristic of me	A little characteristic of me	Somewhat characteristic of me	Very characteristic of me	Entirely characteristic of me
1. Unforeseen events upset me greatly.	1	2	3	4	5
2. It frustrates me not having all the information I need.	1	2	3	4	5
3. Uncertainty keeps me from living a full life.	1	2	3	4	5
4. One should always look ahead so as to avoid surprises.	1	2	3	4	5
5. A small unforeseen event can spoil everything, even with the best of planning.	1	2	3	4	5
6. When it's time to act, uncertainty paralyses me.	1	2	3	4	5
7. When I am uncertain I can't function very well.	1	2	3	4	5
8. I always want to know what the future has in store for me.	1	2	3	4	5
9. I can't stand being taken by surprise.	1	2	3	4	5
10. The smallest doubt can stop me from acting.	1	2	3	4	5
11. I should be able to organize everything in advance.	1	2	3	4	5
12. I must get away from all uncertain situations.	1	2	3	4	5

Score:_____

Appendix F

Measure of specific anxiety

1. I feel anxious about weighing my clients because I think that they might be distressed by the experience, and I find that hard to tolerate.
2. I avoid weighing my clients because I anticipate that weighing will lead to a 'fight' or drama
3. I feel uncomfortable weighing my clients because it feels intrusive and not collaborative to ask them to do something they may not want to do.
4. I am concerned about weighing my clients in case it causes deterioration (e.g., my clients stop eating or the therapy 'goes backwards' as a result)
5. I worry about weighing my clients because I am unsure how to respond if they refuse to be weighed
6. I am unlikely to weigh my clients because I do not want to increase their anxiety.

Appendix G

'Why we should weigh patients with eating disorders?' information sheet

According to NICE guidelines, CBT is the recommended treatment for working with adults with eating disorders. In most cases, CBT protocols for eating disorders recommend that patients are weighed in every session and made aware of their weight. There are four key reasons for weighing patients in therapy sessions:

1. Patient safety – common to all psychotherapies is the need to ensure that patients are physically safe. Weighing allows therapists to monitor a patient's weight to identify and minimise risk.
2. Identifying changes in eating patterns – weighing provides a more accurate indication of changes in weight and eating patterns (e.g. changes in weight due to undisclosed binge-eating or laxative abuse) than does patient's (likely post-hoc) self-reports of their weight or food intake.
3. Reducing anxiety about weighing – many patients are highly anxious about being weighed. Weighing 'in session' provides opportunity to address this anxiety through exposure or behavioural experimentation. Specifically, weighing may provide an opportunity for you to discuss and target the avoidance and safety behaviours around being weighed (e.g. "I will have to starve myself if I know my weight").
4. Addressing the 'broken cognition' – people with eating disorders often experience a cognitive disconnection between eating and weight gain. That is, any food intake is seen to have catastrophic effects on weight and thus intake is minimised to avoid this. Weighing patients provides the data to challenge these beliefs and test predictions about weight gain based on changes to eating.

The latter two reasons can be addressed if weighing is completed in the therapy session and is thus a key task within CBT for eating disorders.

Appendix H

Volitional help-sheet

Many therapists can find it challenging to weigh their clients at every session. Evidence suggests it can be helpful to form a plan to address difficult situations you might experience when weighing clients. Please identify one situation and a related response that you will use, should that challenging situation arise. Please see below for some suggested examples. The situation and responses are not linked, and may be paired in a way that fits for you.

Please identify an if- statement that you think represents a challenge you face when weighing clients, and then identify a response that feels relevant to you.

If …	Then …
If my clients becomes distressed …	Then I will remind myself that weighing my client is an opportunity to explore their thoughts and emotions!
If weighing my client makes me feel anxious …	Then I will take my feelings/experience to supervision and access support!
If I feel uncomfortable sitting with my client's distress …	Then I will include 'weighing' on my session agenda, and weigh my client at the start of the session!
If I struggle to fit weighing into my session or I run out of time …	Then I will remind myself of the expectation of weighing in the treatment contract!
If I think that it is unlikely that I will weigh my client this session …	Then I will revisit the evidence-based protocol and remind myself of the rationale for weighing clients!
If I don't think that it's important to weigh this client …	Then I will ensure that I have access to scales prior to the session!
If there are practical challenges to weighing my client …	Then I will remind myself of the importance of using objective measures to monitor my client's weight!
If I think that I won't weigh because my client looks like they've gained weight or are a healthy weight …	Then I will explain the rationale of weighing to them, and discuss their commitment to treatment!
If my client refuses to be weighed …	Then I will not present being weighed as an option, but will ask when my client wants to be weighed during the session (e.g., "now or after 10 minutes?")!

Please now write out your plan below using the format 'if [situation], then I will [response]', and commit yourself to carrying it out. You can use one of the examples, or create an original plan relevant to you.

Appendix I

Pilot project interview schedule

Participants to be provided information about the purpose of the interview

- I am conducting a study with clinicians who work with adults with an eating disorder. The project is interested in why some clinicians might not weigh their clients with an eating disorder.
- This interview will help inform the development of a volitional help-sheet. This is a tool that that will be used in the experimental condition of the research project to support clinicians to develop implementation intention plans.
- The study will explore whether the volitional help-sheet might be useful in supporting clinicians to weigh people with an eating disorder.
- By weighing, I am referring to you as the clinician using scales to take the weight of your client, measured in either stones, pounds or kg, within a therapy session.

Consent and right to withdraw

- You are in no obligation to engage in this interview
- You have the right to withdraw at any point, or refuse to answer any questions
- Any information you provide will be kept confidential. However, as participation in this interview excludes you from the study, you will be given an identifying code.
- Do you have any questions? Are you still happy to take part in the interview?

Identification of challenges/barriers to weighing

- I am interested in your experience of weighing clients with eating disorders.
- Typically, how often would you say that you weigh your clients within a therapy session?
- Are there any times that you would choose not to weigh your client?

  PROMPT: you say you don't find weighing difficult… could you say what you think *other clinicians* may find difficult about weighing?

  PROMPT: you say you don't think it is important to weigh your client at each contact…. could you tell me more about your reasoning for this?

- What is it that makes weighing clients difficult, awkward or impossible? And why?

  PROMPT: what impact does your client's reaction to being weighed have on your decision to weigh? (E.g. anxiety, distress, anger etc)
  PROMPT: do your emotions impact on your ability to weigh? (Anxiety, uncertainty) Could you say more about that?
  PROMPT: do you/other clinicians think that weighing a client could impact on your therapeutic relationship? How?
- In which circumstances are you most likely to weigh your client?

Identification of potential strategies for overcoming challenges

The research will look at ways that clinicians might be helped to weigh their clients with eating disorders. This is thought to be important because existing literature suggests that

despite routine weighing being an integral part of the evidence based therapeutic treatment of people with an eating disorder, many clinicians don't regularly weigh their clients.

- Were you aware of this aspect of the protocol/evidence base underlying the treatment model you use (CBT)?

Now that we have talked about the importance of weighing, I'd like to move on to think about ways that we might support clinicians to be able to weigh in line with treatment protocol.

- Do you have any ideas or solutions that would help increase the likelihood that you would weigh your client?

PROMPT: You mentioned 'x' as a barrier to weighing, is there anything that would help you feel more inclined to weigh in this scenario?
PROMPT: Could any practical arrangements be made to help you weigh your client?
PROMPT: (where appropriate) it sounds like the emotional experience of weighing a client has an impact, is there anything that would help you feel more inclined to weigh when this situation arises?

- Can you think of any solutions for other clinicians who may find it difficult to weigh their clients?
- If you had to arrive at the 'top 5 solutions' for regular and routine weighing of clients, what would these be?

Ending

- I've completed all the questions I wanted to ask you today, is there anything you'd like to ask me?
- Thank-you for taking the time to talk with me today

Appendix J

Parallel DClinPsy research project information

A second DClin Psy project was completed in parallel to the current study. The parallel project implemented the same intervention and planned to analyse the findings in a similar way, but focused on therapists working with children and adolescents with eating disorders using Family-based Treatment. Aspects of these projects were designed together to manage time and resources more effectively, but were conducted, analysed and reported independently.

Appendix K

Invitation email to participate part two (post-intervention)

Two weeks ago, you were invited to take part in an online research study interested in the use of Cognitive Behaviour Therapy (CBT) as a treatment approach when working with adults with eating disorders.

Thank you for completing the first part of this study. Please find below the link to the second part of the online survey. This should take no longer than 15 minutes to complete.

www.qualtricslink....

Participation in this study is voluntary. Your answers will remain anonymous, and you may choose to withdraw from the study at any time. Your contributions to this research area are greatly valued. A full debrief will be available following participation.

This research has received ethical approval from the University of Sheffield's Department of Psychology Ethics Committee.

If you would like any further information about the study, then please contact htrivasse1@sheffield.ac.uk (Heidi Trivasse, Trainee Clinical Psychologist).

The project is supervised by:

Professor Glenn Waller, Clinical Psychology Unit, University of Sheffield

Dr Thomas Webb, Department of Psychology, University of Sheffield

Appendix L

Invitation email to participate part three (follow-up)

Four weeks ago, you took part in an online research study interested in the use of Cognitive Behaviour Therapy (CBT) adults with eating disorders. Thank you for taking the time to complete the first two parts of this study. This research is interested in whether elements of clinical practice change over time. We did not tell you at the time, but we would therefore like to ask you to complete one more short online survey. Please find below the link to the final part of the online survey. This should take no longer than 15 minutes to complete.

www.qualtricslink....

Participation in this study is voluntary, and you may choose to withdraw from the study at any time. Your contributions to this research area are greatly valued. This is the final time you will be asked to complete any measure and a full debrief will be available following participation.

This research has received ethical approval from the University of Sheffield's Department of Psychology Ethics Committee.

If you would like any further information about the study please contact: htrivasse1@sheffield.ac.uk (Heidi Trivasse, Trainee Clinical Psychologist).

The project is supervised by:

Professor Glenn Waller, Clinical Psychology Unit, University of Sheffield

Dr Thomas Webb, Department of Psychology, University of Sheffield

Appendix M

Data analyses of original data-set

Data analysis of original data-set

Table 11 provides a summary of the demographic characteristics of the original data-set. Table 12 provides a summary of the means and standard deviation scores of primary outcome variables. The results of the Friedman Tests indicated that there were no significant differences in the percentage of patients that therapists weighed across the three time-points, for the control condition ($N = 16$), X^2 (2) = 1.70, $p = .428$, or the intervention condition ($N = 13$), X^2 (2) = 3.27, $p = .195$. Consequently, no post-hoc analyses were performed.

Mann-Whitney U tests were used to assess differences between conditions at each time-point. The results indicated that there was no significant difference between the control and intervention condition at baseline, $U = 617.00$, $z = -0.97$, $p = .331$. A significant medium-sized difference was found between the control and intervention conditions at the post-intervention time-point, $U = 180.00$, $z = -1.95$, $p = .051$, $d = 0.56$, suggesting that participants in the intervention condition ($M = 86.28\%$) weighed significantly more of their patients than those in the control condition ($M = 70.08\%$). No significant difference was found between the control and intervention conditions at the follow-up time-point, $U = 98.00$, $z = -0.54$, $p = .588$.

M**oderation analyses**. Spearman Rank Order correlations showed that baseline general anxiety and baseline intentions to weigh were not significantly correlated with changes in weighing behaviour between baseline and post-intervention for the control condition (Table 13) or the intervention condition (Table 14), and so no moderation analyses were performed.

Mediation analyses. Spearman Rank Order correlations showed that specific anxiety was not significantly correlated with changes in weighing behaviour across the time-points

for the control condition (Table 13) or the intervention condition (Table 14), and so no mediation analyses were performed.

Table 11: Summary of demographic characteristics for original data-set

Variable	Control group (*N* = 40)	Intervention group (*N* = 44)
Age, *M (SD)*	35.80 (10.1)	39.83 (10.1)
Missing, *N*	-	2
Gender, *N (%)*		
Female	35 (87.5)	34 (77.3)
Male	5 (12.5)	8 (18.2)
Missing	-	2 (4.5)
Level of CBT training, *N (%)*		
Professional training	9 (22.5)	9 (20.5)
Postgraduate training	27 (67.5)	28 (63.6)
No training	0 (0.0)	1 (2.3)
Other	4 (10.0)	6 (13.6)
Length of CBT practice, *N (%)*		
< 1 year	8 (20.0)	7 (15.9)
years	8 (20.0)	7 (15.9)
2-5 years	11 (27.5)	12 (27.3)
> 5 years	13 (32.5)	16 (36.4)
Missing	-	2 (4.5)

Table 12: Summary of means and standard deviations for outcome variables in original data-set

Variable	Control	Intervention
Baseline weighing behaviour, *M (SD)*	69.56 (37.21)	79.04 (34.67)
N	37	38
Post-intervention weighing behaviour, *M (SD)*	70.08 (36.42)	86.28 (29.90)
N	25	21
Follow-up weighing behaviour, *M (SD)*	77.41 (33.24)	86.03 (18.33)
N	17	13
Baseline general anxiety, *M (SD)*	1.83 (0.58)	1.72 (0.32)
N	38	41
Post-intervention general anxiety, *M (SD)*	1.81 (0.59)	1.62 (0.35)
N	26	22
Baseline specific anxiety, *M (SD)*	1.18 (0.24)	1.21 (0.31)
N	38	41
Post-intervention specific anxiety, *M (SD)*	1.22 (0.27)	1.19 (0.29)
N	26	22
Baseline intent to weigh, *M (SD)*	6.17 (1.50)	6.50 (0.95)
N	36	38
Post-intervention intent to weigh, *M (SD)*	6.63 (0.71)	6.20 (1.70)
N	24	20

Table 13: Spearman Rank Order Correlations for the original data-set control condition

		Weighing behaviour change					
		Baseline to post-intervention		**Baseline to follow-up**		**Post-intervention to follow-up**	
Time-point	**Variable**	***r_s (p)***	***N***	***r_s (p)***	***N***	***r_s (p)***	***N***
Baseline	**Age**	.34 (*.106*)	24	.12 (*.657*)	16	.04 (*.894*)	17
	Gender	-.11 (*.612*)	24	.39 (*.140*)	16	.23 (*.376*)	17
	Level of training	-.20 (*.339*)	24	.26 (*.337*)	16	.32 (*.215*)	17
	Length of practice	.31 (*.146*)	24	-.22 (*.412*)	16	-.19 (*.458*)	17
	General anxiety	-.23 (*.280*)	24	.04 (*.878*)	16	.27 (*.287*)	17
	Specific anxiety	-.25 (*.234*)	24	.16 (*.568*)	16	.26 (*.311*)	17
	Intentions to weigh	-.20 (*.347*)	24	-.53* (*.034*)	16	-.33 (*.208*)	16
	Weighing behaviour	-.39 (*.058*)	24	-.43 (*.097*)	16	-.01 (*.968*)	16
Post-intervention	**General anxiety**	-.30 (*.169*)	23	-.10 (*.706*)	16	.21 (*.432*)	16
	Specific anxiety	-.05 (*.821*)	23	.21 (*.433*)	16	.36 (*.175*)	16
	Intentions to weigh	-.04 (*.847*)	23	-.57* (*.021*)	16	-.38 (*.135*)	17
	Weighing behaviour	.43* (*.036*)	24	-.33 (*.214*)	16	-.64* (*.005*)	17

*Note: * $p < .05$, ** $p < .001$*

Table 14: Spearman Rank Order Correlations for the original data-set intervention condition

		Weighing behaviour change					
		Baseline to post-intervention		**Baseline to follow-up**		**Post-intervention to follow-up**	
Time-point	**Variable**	***r_s (p)***	***N***	***r_s (p)***	***N***	***r_s (p)***	***N***
Baseline	**Age**	.48* (*.032*)	20	.01 (.966)	13	-.83** (*<.001*)	13
	Gender	-.04 (*.854*)	20	-.26 (*.395*)	13	.35 (*.239*)	13
	Level of training	-.12 (*.610*)	20	-.28 (*.354*)	13	-.05 (*.879*)	13
	Length of practice	.29 (*.214*)	20	.40 (*.175*)	13	-.13 (*.681*)	13
	General anxiety	.21 (*.368*)	20	-.26 (*.387*)	13	.25 (*.417*)	13
	Specific anxiety	.05 (*.844*)	20	-.32 (*.282*)	13	.21 (*.485*)	13
	Intentions to weigh	-.21 (*.376*)	20	.51 (*.074*)	13	.22 (*.467*)	13
	Weighing behaviour	-.79** (*<.001*)	20	-.54 (*.056*)	13	.35 (*.239*)	13
Post-intervention	**General anxiety**	.01 (*.963*)	20	-.31 (*.311*)	13	.05 (*.880*)	13
	Specific anxiety	.08 (*.728*)	20	-.03 (*.926*)	13	.16 (*.613*)	13
	Intentions to weigh	-.02 (*.922*)	19	.60* (*.030*)	13	.31 (*.308*)	13
	Weighing behaviour	-.05 (*.833*)	20	.18 (*.554*)	13	-.41 (*.170*)	13

*Note: * $p < .05$, ** $p < .001$*

Appendix N

Results of normality testing

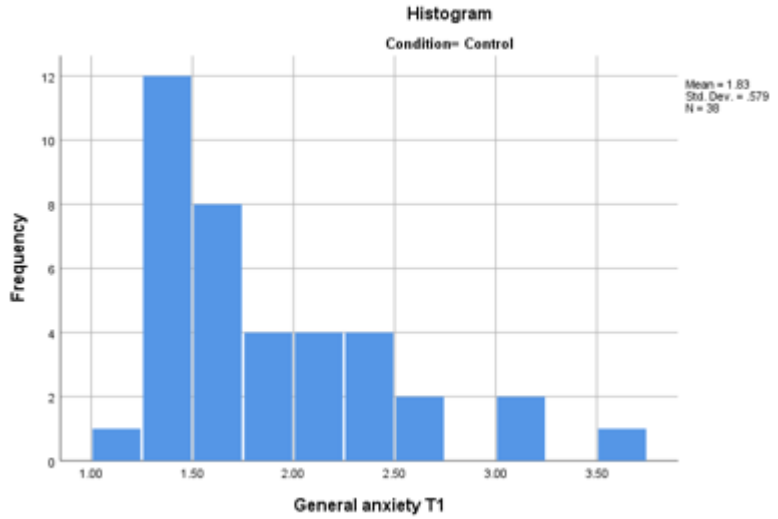
Original data-set

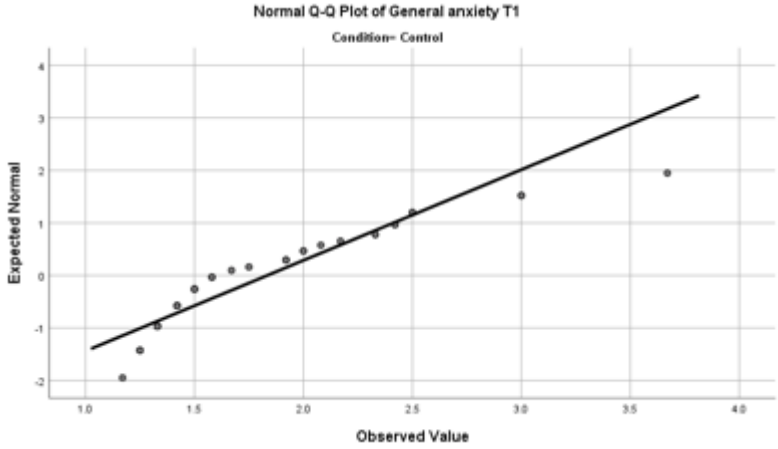
Kolmogorov-Smirnov statistics.

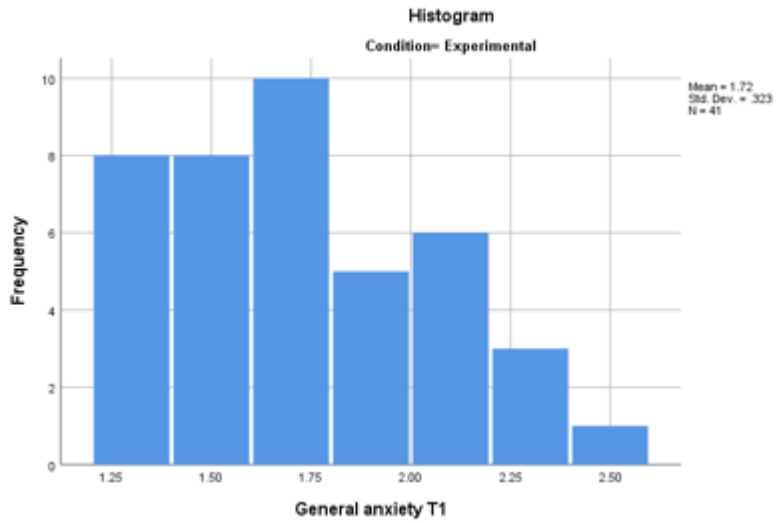
Table 15: Kolmogorov-Smirnov statistics for control and intervention groups in original data-set

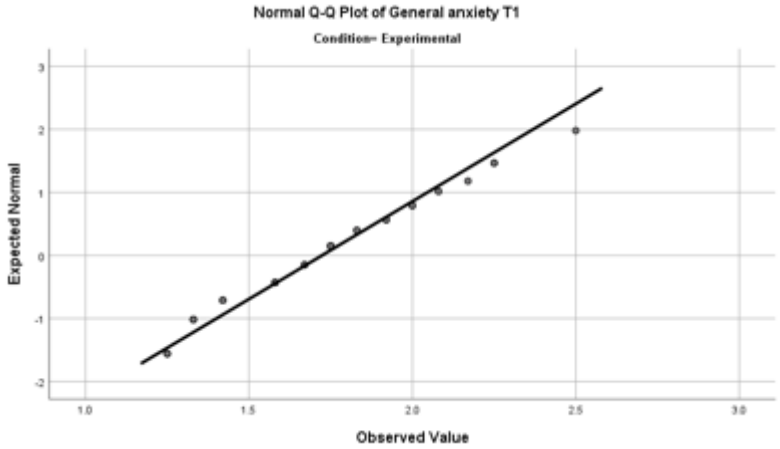
Condition	Variable	Kolmogorov-Smirnov	df	*p*
Control	Baseline general anxiety	.20	38	.001
	Post-intervention general anxiety	.14	26	.200
	Baseline specific anxiety	.30	38	<.001
	Post-intervention specific anxiety	.29	26	<.001
	Baseline intentions to weigh	.38	36	<.001
	Post-intervention intentions to weigh	.45	24	<.001
	Baseline weighing behaviour	.25	37	<.001
	Post-intervention weighing behaviour	.25	25	<.001
	Follow-up weighing behaviour	.30	17	<.001
Intervention	Baseline general anxiety	.10	41	.200
	Post-intervention general anxiety	.17	22	.097
	Baseline specific anxiety	.29	41	<.001
	Post-intervention specific anxiety	.29	22	<.001
	Baseline intentions to weigh	.36	38	<.001
	Post-intervention intentions to weigh	.38	20	<.001
	Baseline weighing behaviour	.30	38	<.001
	Post-intervention weighing behaviour	.35	21	<.001
	Follow-up weighing behaviour	.24	13	.042

Histograms and Q-Q plots for baseline general anxiety.

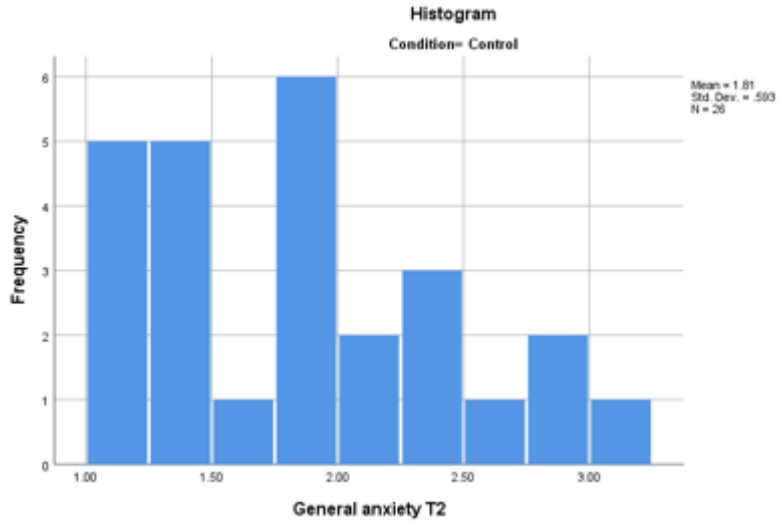

Histogram
Condition= Control
Mean = 1.83
Std. Dev. = .579
N = 38
Frequency
General anxiety T1

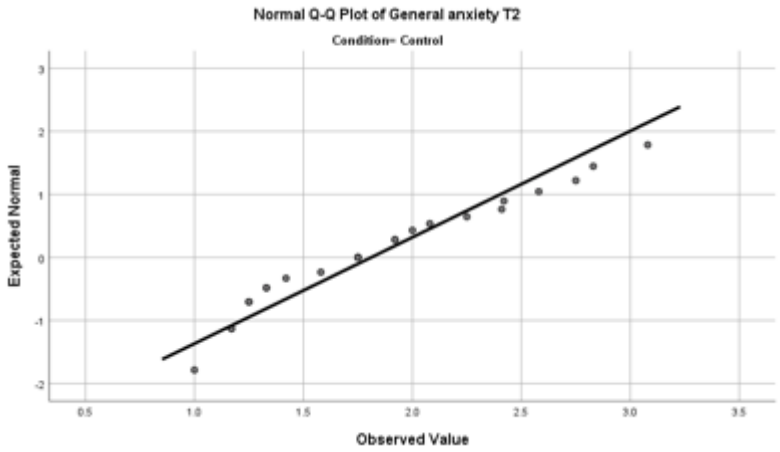

Normal Q-Q Plot of General anxiety T1
Condition= Control
Expected Normal
Observed Value

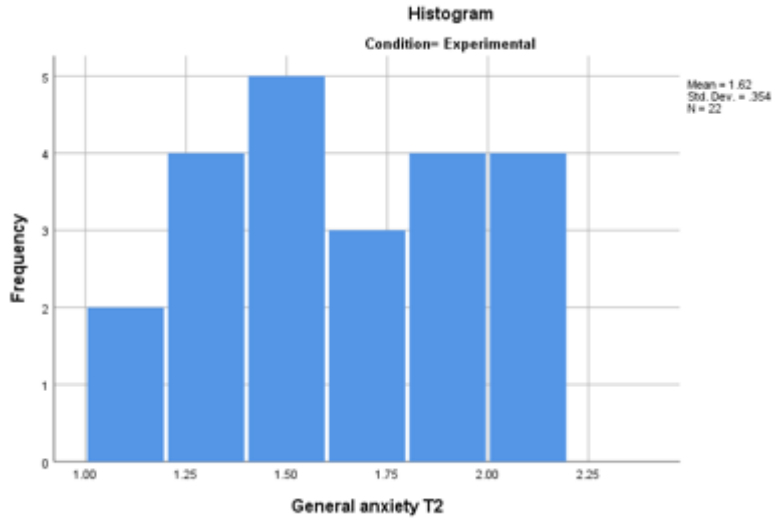

Histogram
Condition= Experimental
Mean = 1.72
Std. Dev. = .323
N = 41
Frequency
General anxiety T1

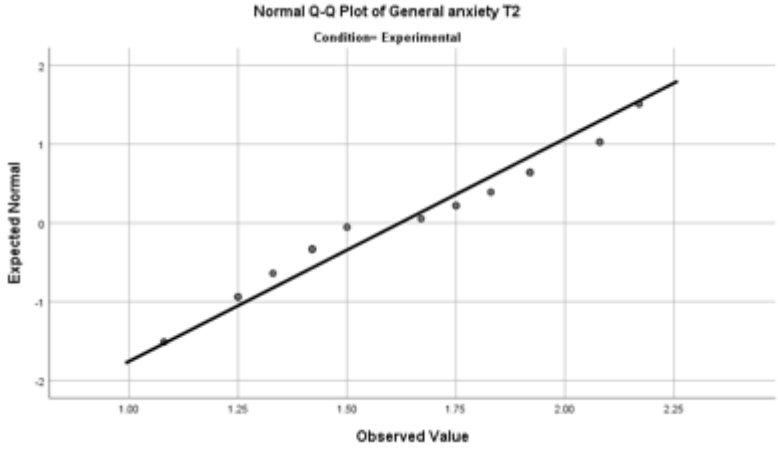


Histograms and Q-Q plots for post-intervention general anxiety.

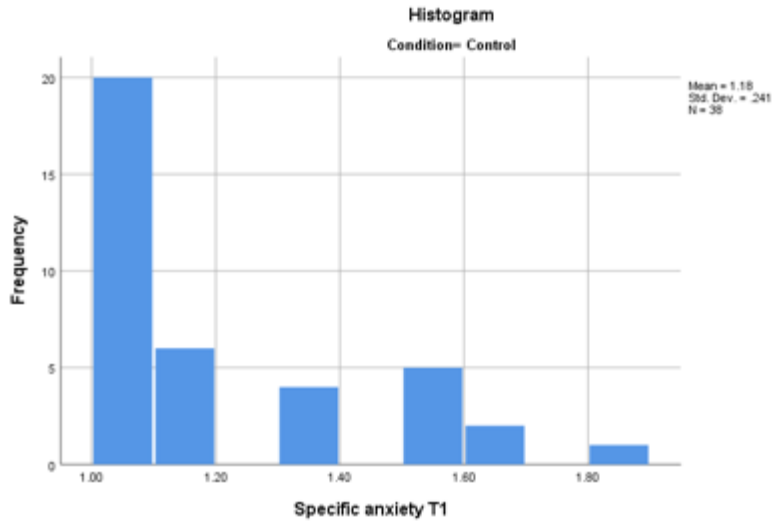


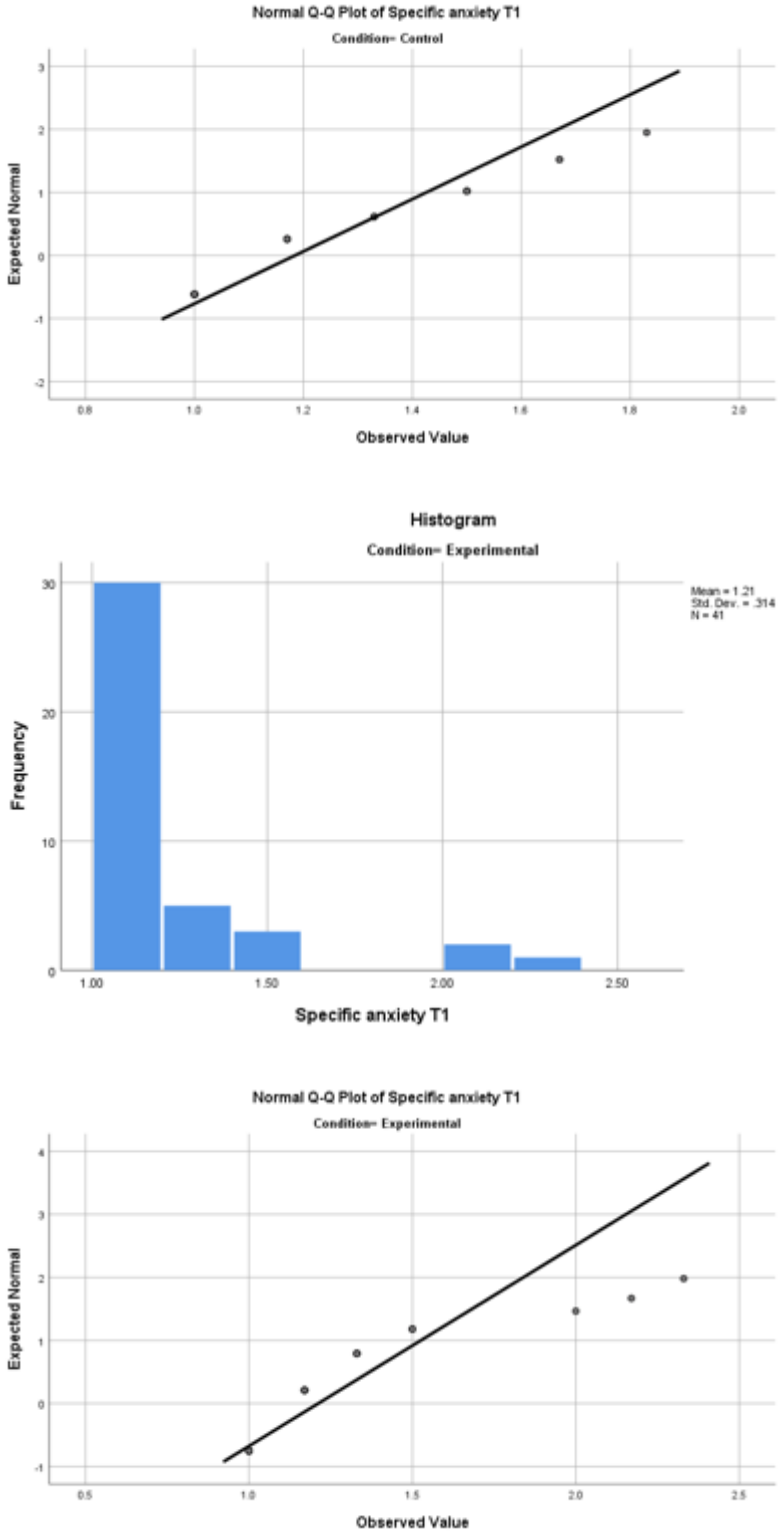




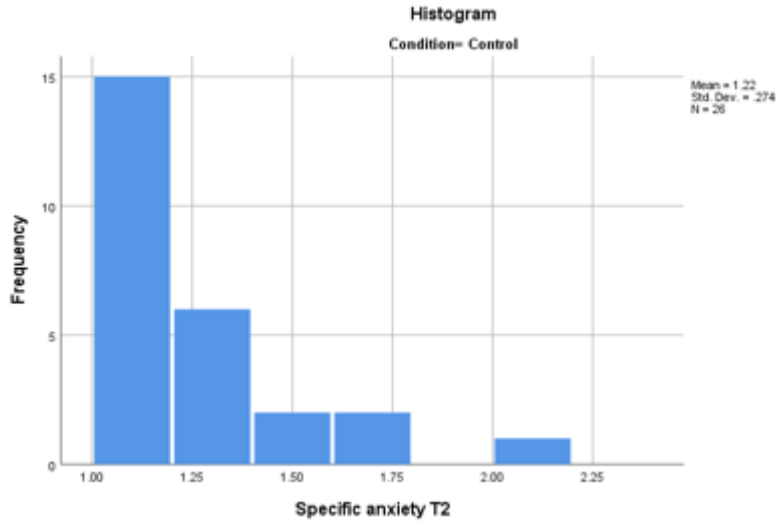


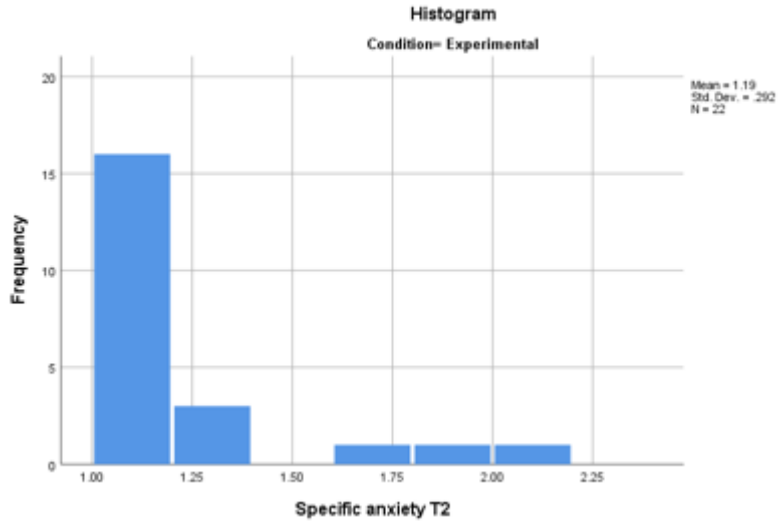
Histograms and Q-Q plots for baseline specific anxiety.

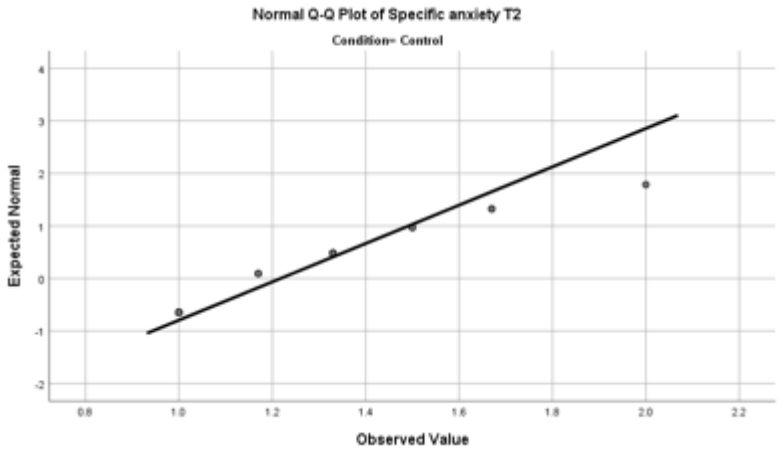


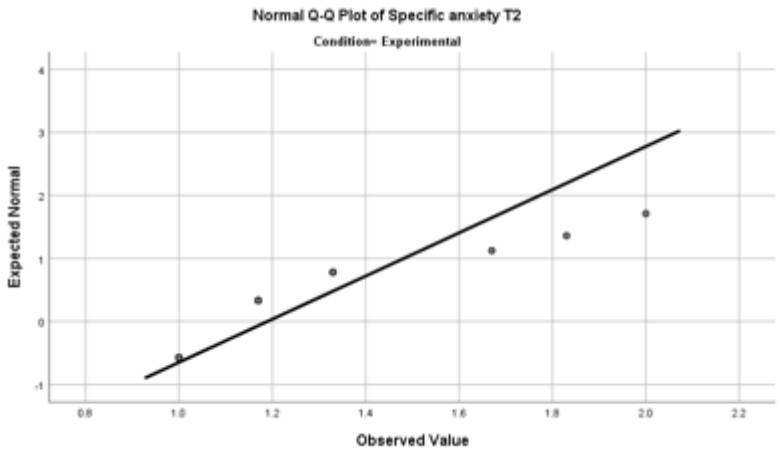


Histograms and Q-Q plots for post-intervention specific anxiety.

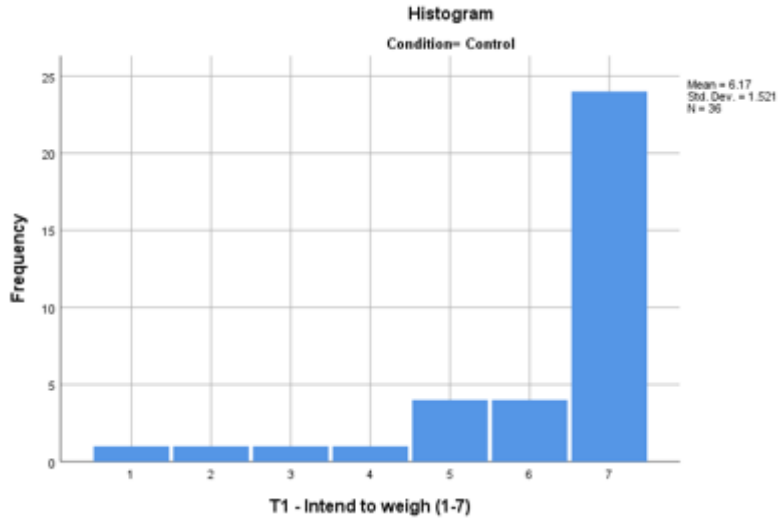

Histogram
Condition= Control
Mean = 1.22
Std. Dev. = .274
N = 26
Frequency
Specific anxiety T2

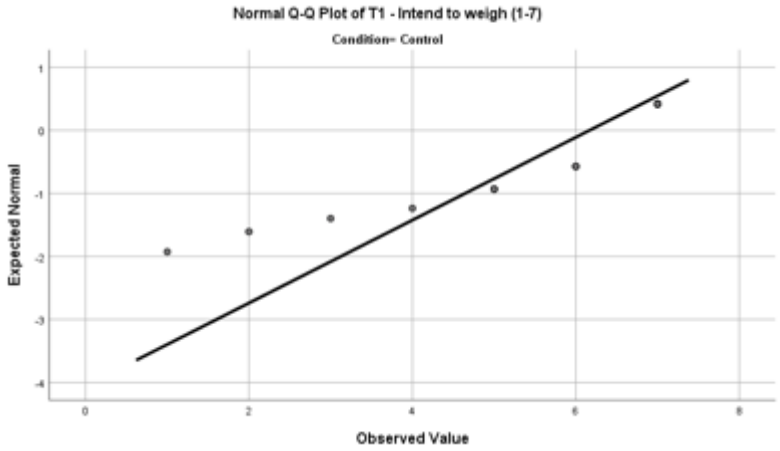

Histogram
Condition= Experimental
Mean = 1.19
Std. Dev. = .292
N = 22
Frequency
Specific anxiety T2

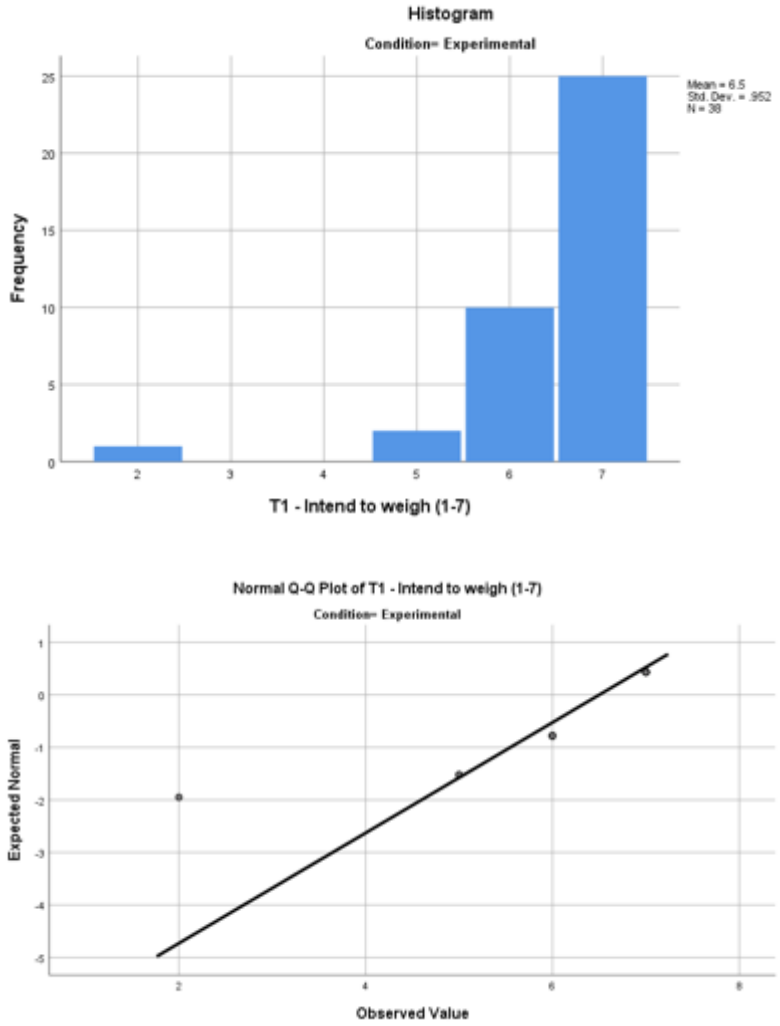

Normal Q-Q Plot of Specific anxiety T2
Condition= Control
Expected Normal
Observed Value



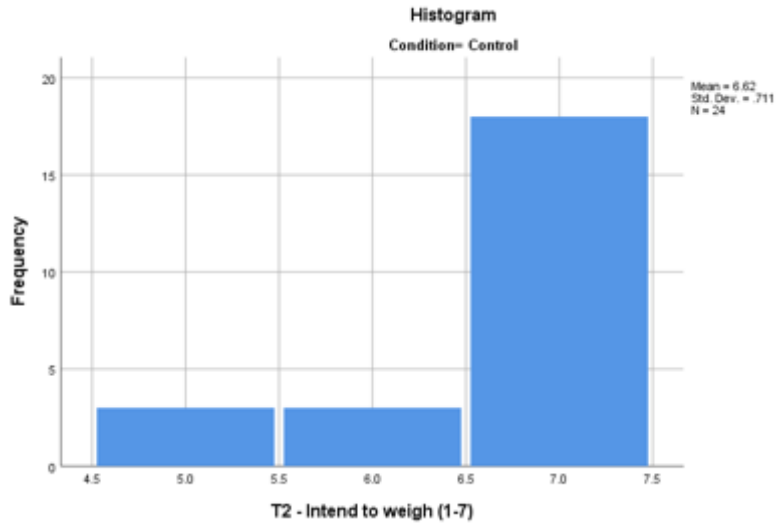
Histograms and Q-Q plots for baseline intentions to weigh.

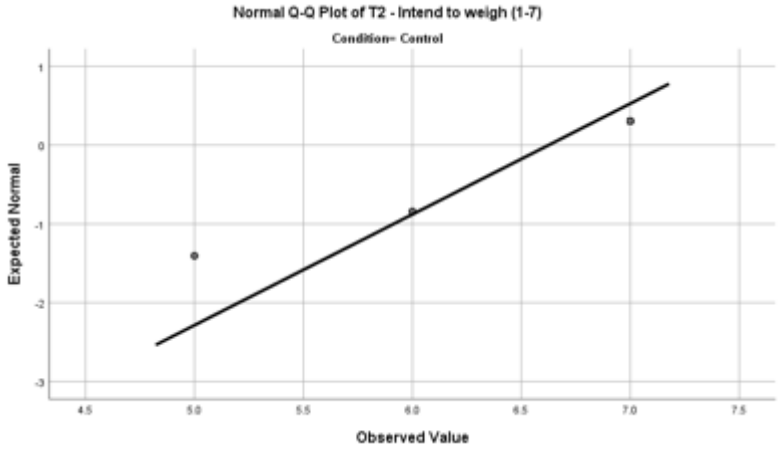


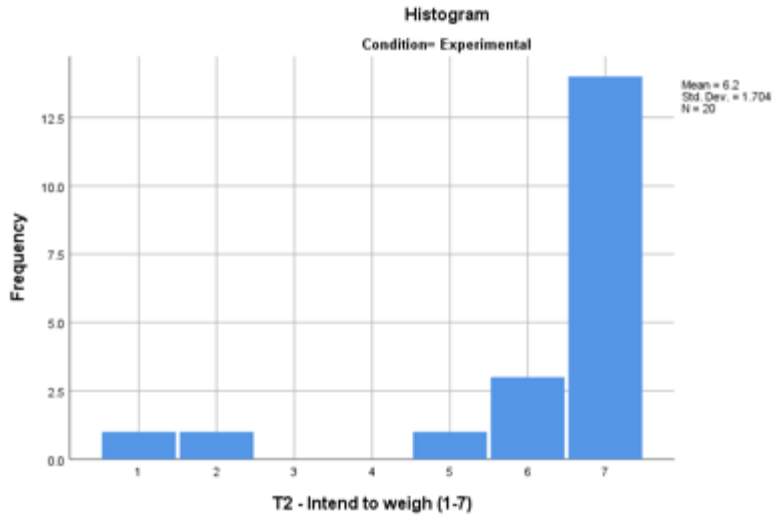


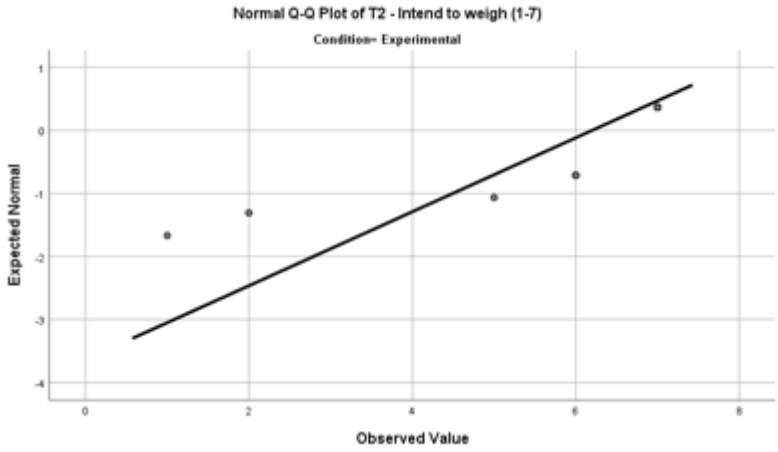


Histograms and Q-Q plots for post-intervention intentions to weigh.

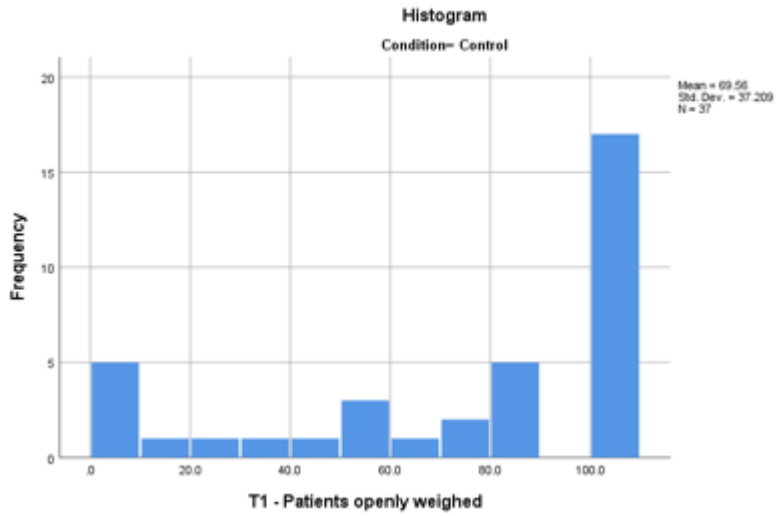


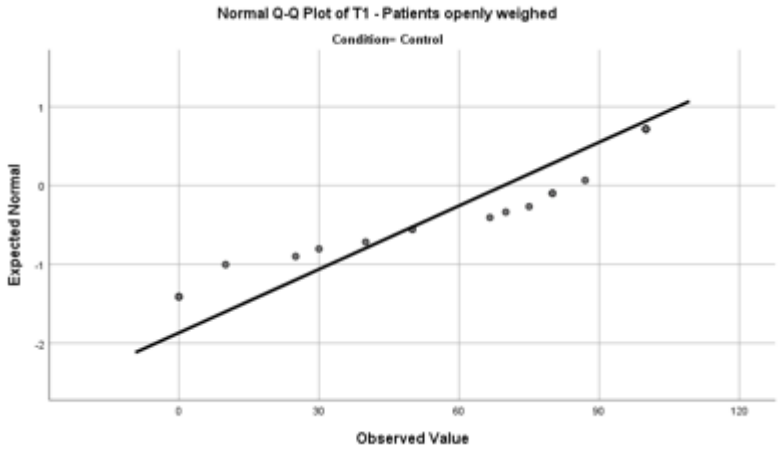


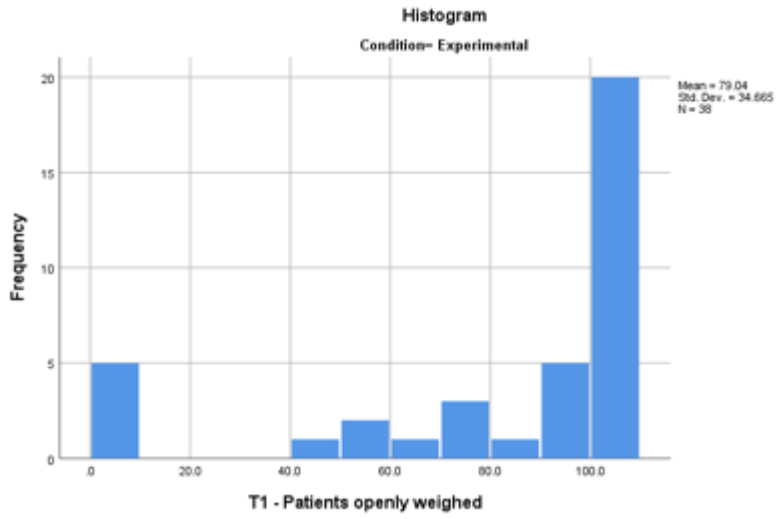


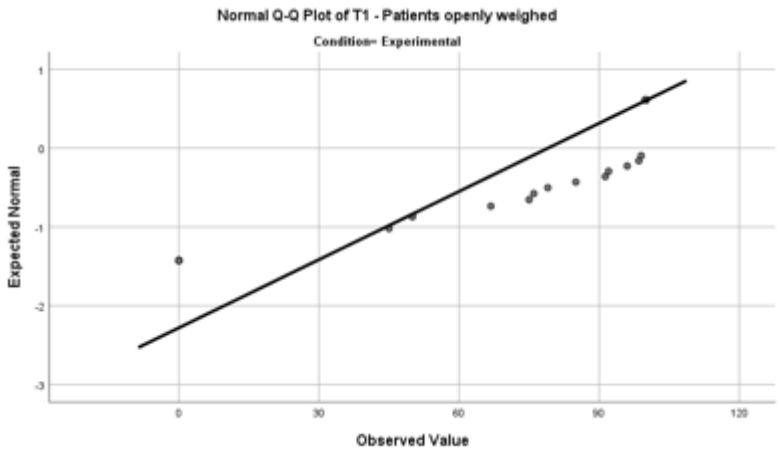


Histograms and Q-Q plots for baseline weighing behaviour.

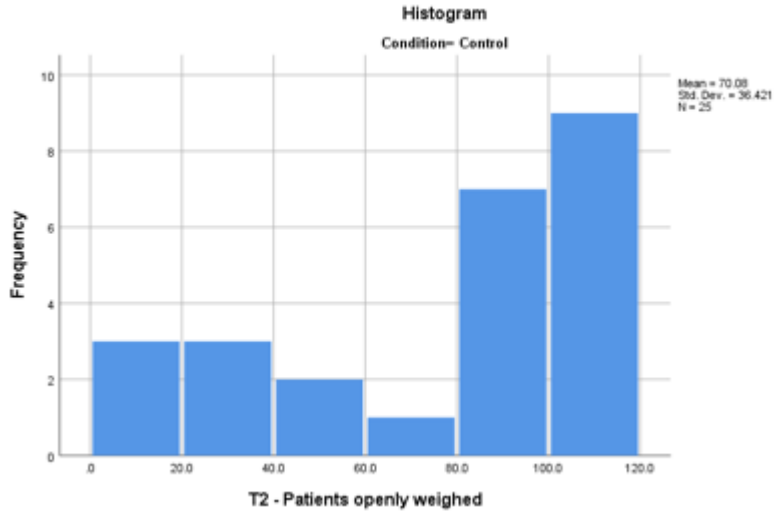

Histogram
Condition= Control
Mean = 69.56
Std. Dev. = 37.209
N = 37
Frequency
T1 - Patients openly weighed

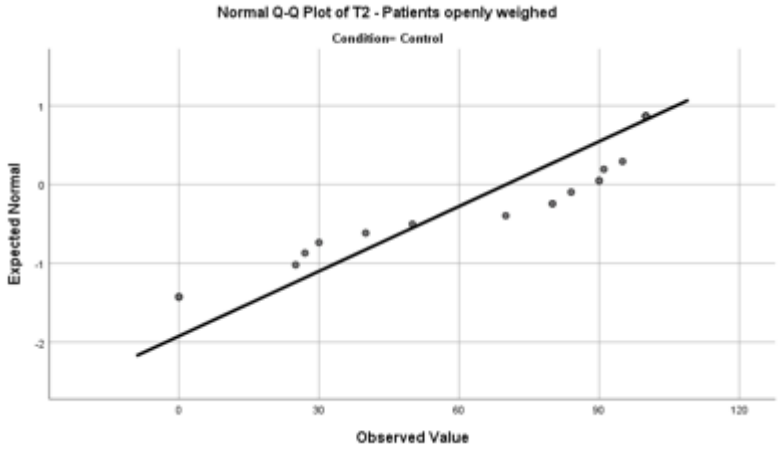

Normal Q-Q Plot of T1 - Patients openly weighed
Condition= Control
Expected Normal
Observed Value

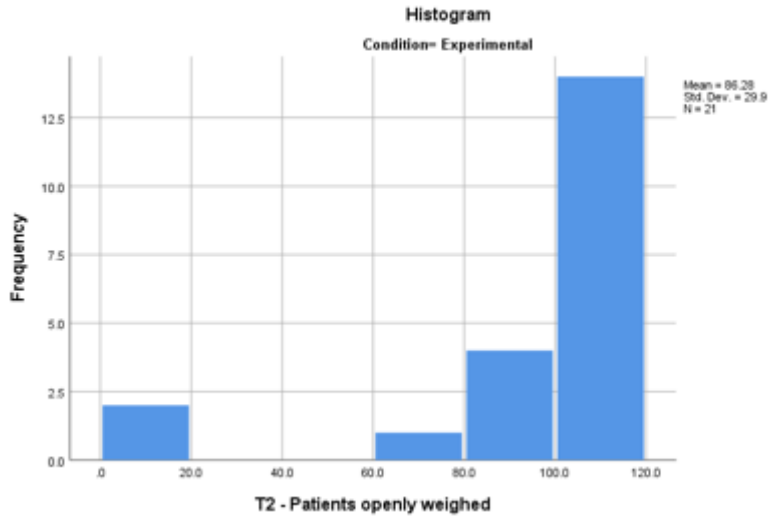

Histogram
Condition= Experimental
Mean = 79.04
Std. Dev. = 34.665
N = 38
Frequency
T1 - Patients openly weighed

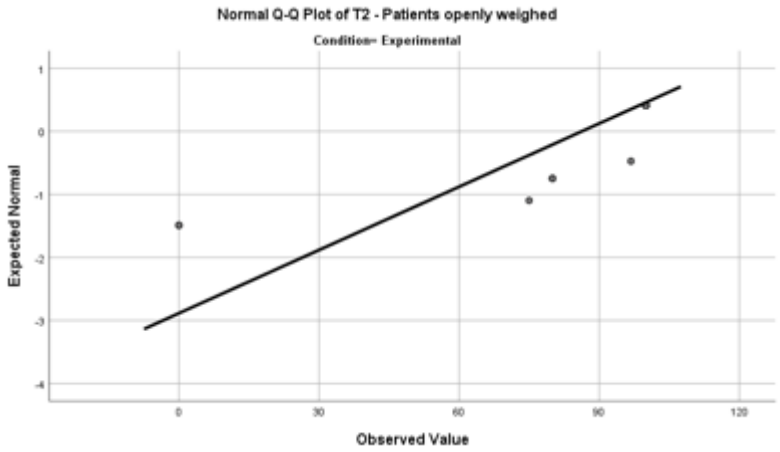


Histograms and Q-Q plots for post-intervention weighing behaviour.

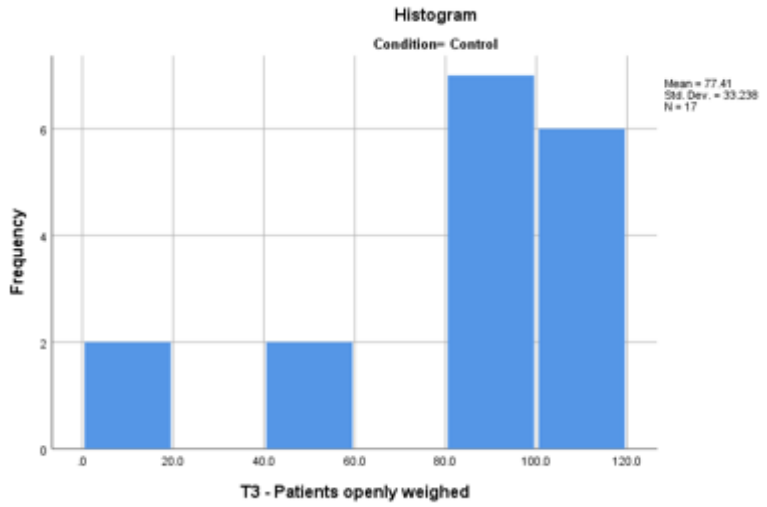


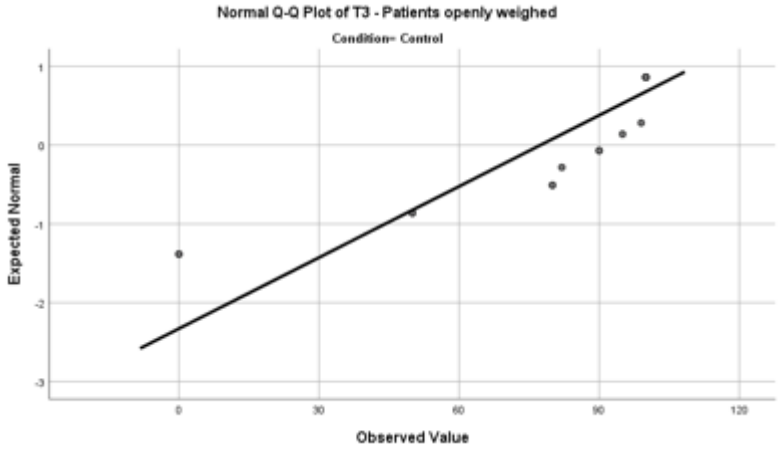


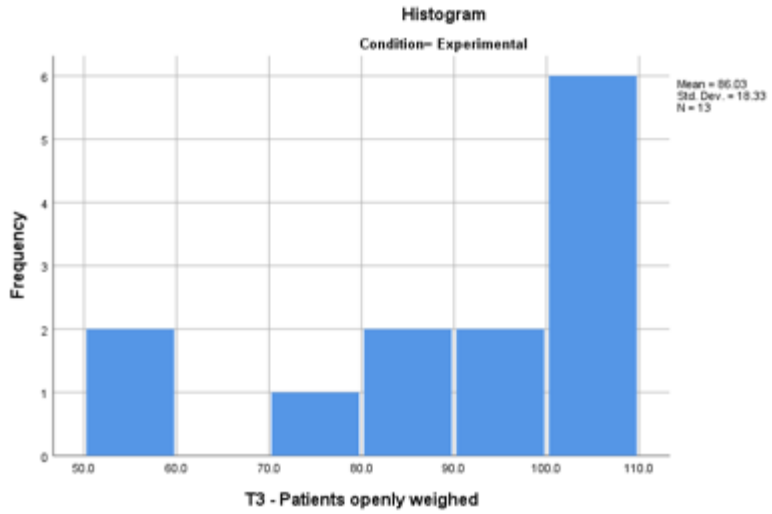


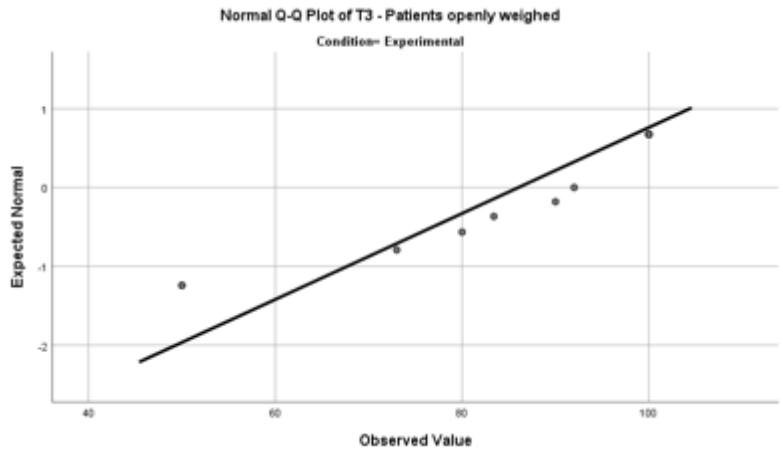


Histograms and Q-Q plots for follow-up weighing behaviour.


Histogram
Condition= Control
Mean = 77.41
Std. Dev. = 33.238
N = 17
Frequency
0
2
4
6
.0
20.0
40.0
60.0
80.0
100.0
120.0
T3 - Patients openly weighed


Normal Q-Q Plot of T3 - Patients openly weighed
Condition= Control
Expected Normal
1
0
-1
-2
-3
0
30
60
90
120
Observed Value


Histogram
Condition= Experimental
Mean = 86.03
Std. Dev. = 18.33
N = 13
Frequency
0
1
2
3
4
5
6
50.0
60.0
70.0
80.0
90.0
100.0
110.0
T3 - Patients openly weighed



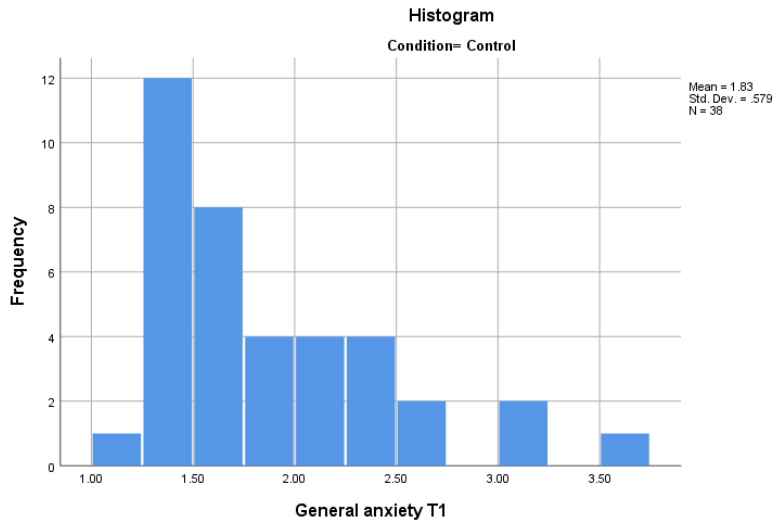
LOCF data-set

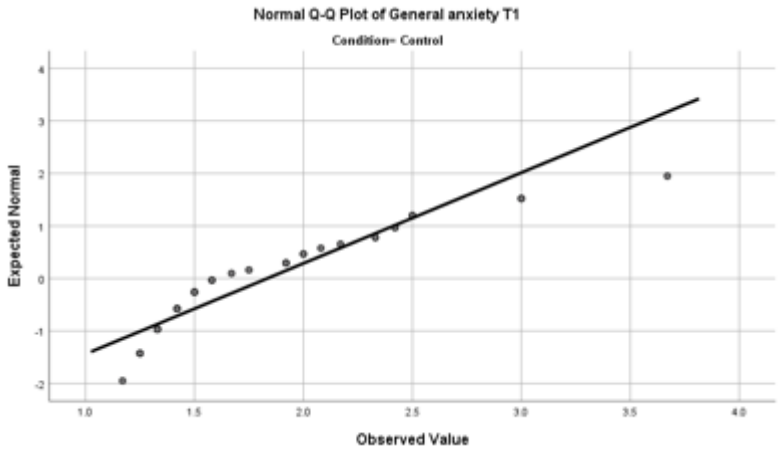
Kolmogorov-Smirnov statistics.

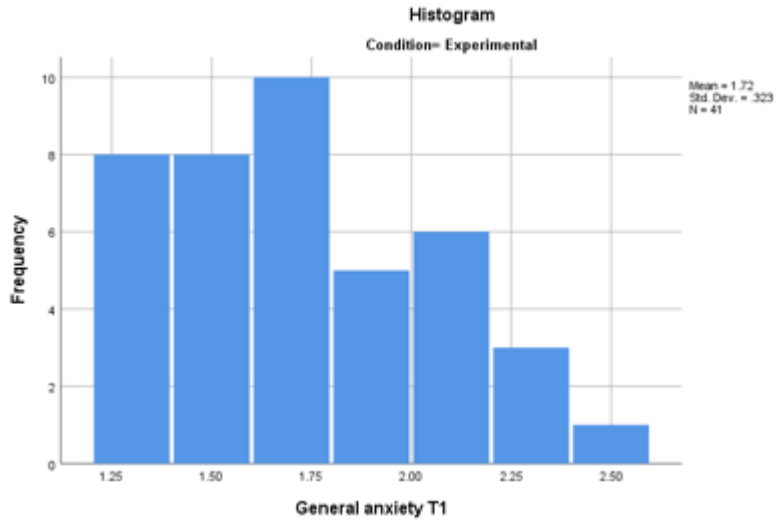
Table 16: Kolmogorov-Smirnov statistics for control and intervention groups in LOCF data-set

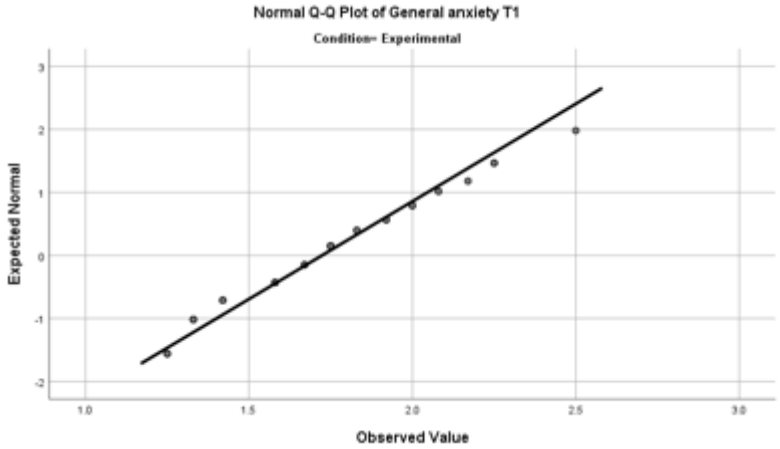
Condition	Variable	Kolmogorov-Smirnov	df	*p*
Control	Baseline general anxiety	.20	38	.001
	Post-intervention general anxiety	.14	38	.053
	Baseline specific anxiety	.30	38	<.001
	Post-intervention specific anxiety	.31	38	<.001
	Baseline intentions to weigh	.38	36	<.001
	Post-intervention intentions to weigh	.38	37	<.001
	Baseline weighing behaviour	.25	37	<.001
	Post-intervention weighing behaviour	.22	38	<.001
	Follow-up weighing behaviour	.21	38	<.001
Intervention	Baseline general anxiety	.10	41	.200
	Post-intervention general anxiety	.12	41	.196
	Baseline specific anxiety	.29	41	<.001
	Post-intervention specific anxiety	.25	41	<.001
	Baseline intentions to weigh	.36	38	<.001
	Post-intervention intentions to weigh	.35	39	<.001
	Baseline weighing behaviour	.30	38	<.001
	Post-intervention weighing behaviour	.32	39	<.001
	Follow-up weighing behaviour	.28	39	<.001

Histograms and Q-Q plots for baseline general anxiety.

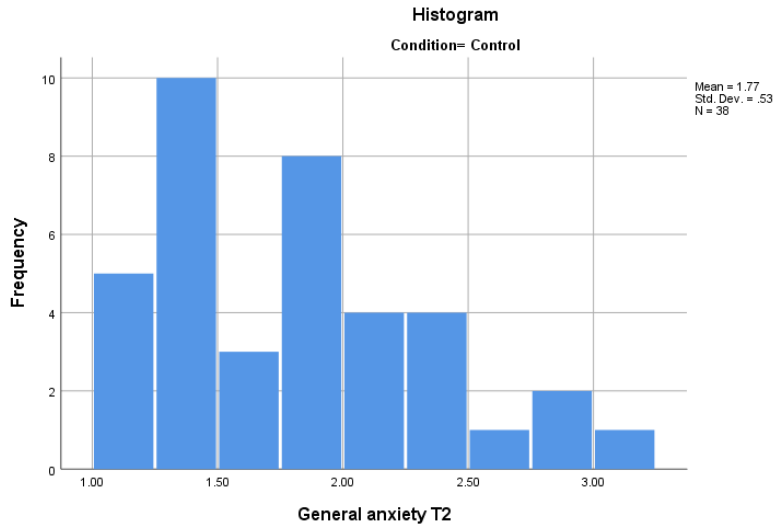

Histogram
Condition= Control
Mean = 1.83
Std. Dev. = .579
N = 38
Frequency
General anxiety T1

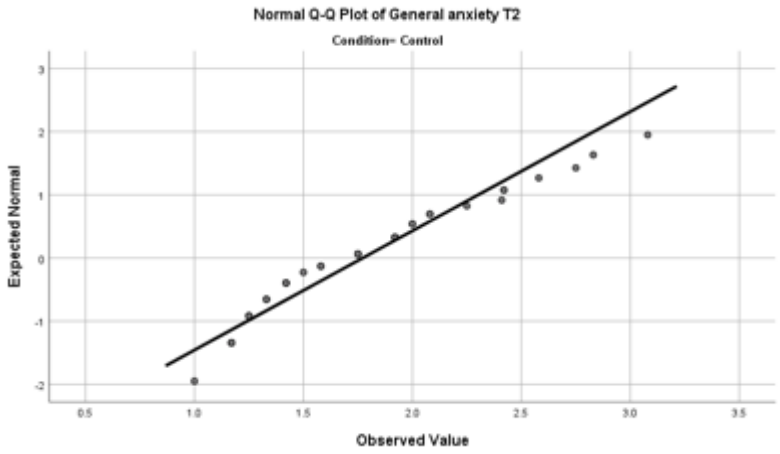

Normal Q-Q Plot of General anxiety T1
Condition= Control
Expected Normal
Observed Value

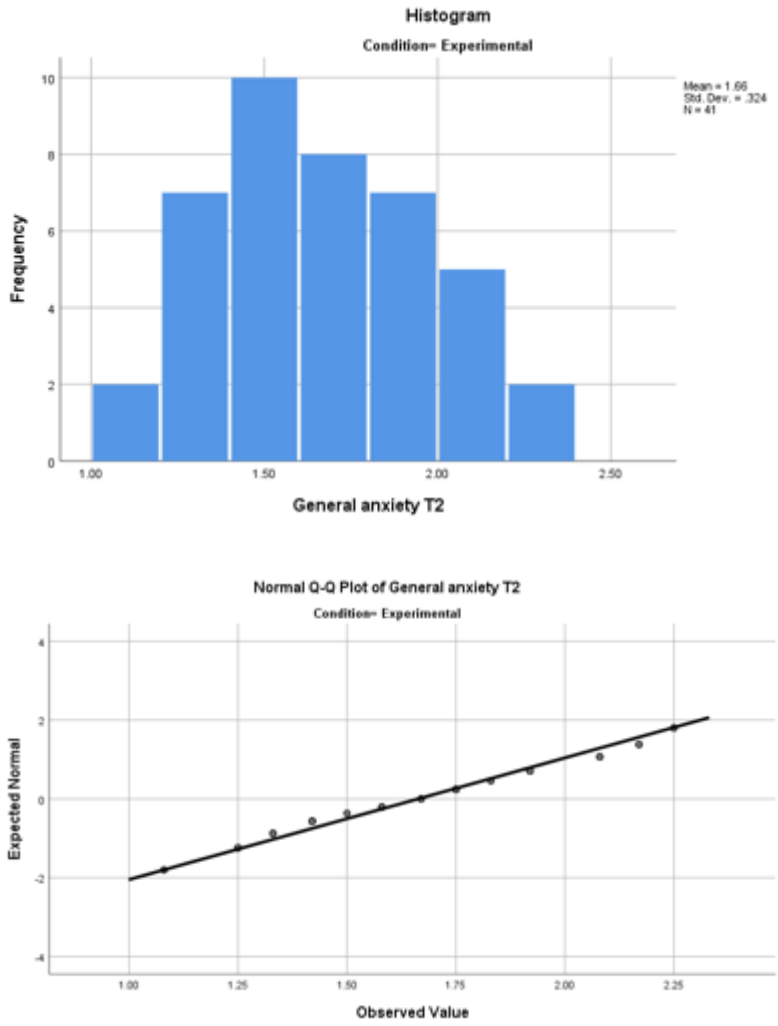

Histogram
Condition= Experimental
Mean = 1.72
Std. Dev. = .323
N = 41
Frequency
General anxiety T1



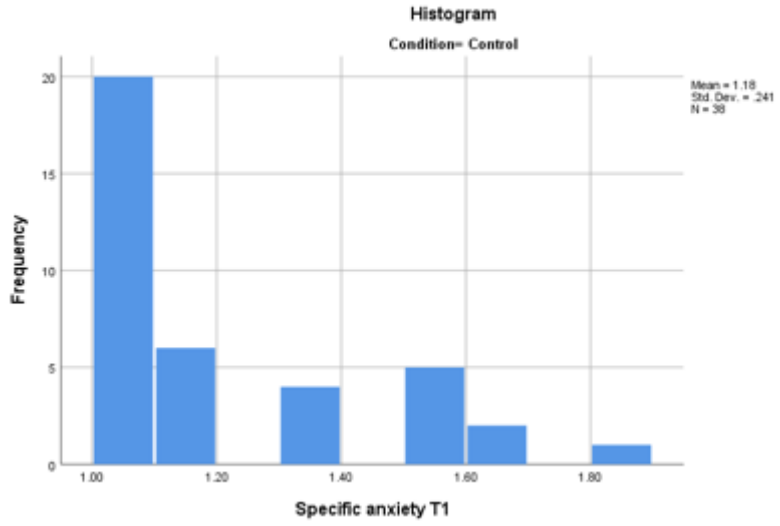
Histograms and Q-Q plots for post-intervention general anxiety.

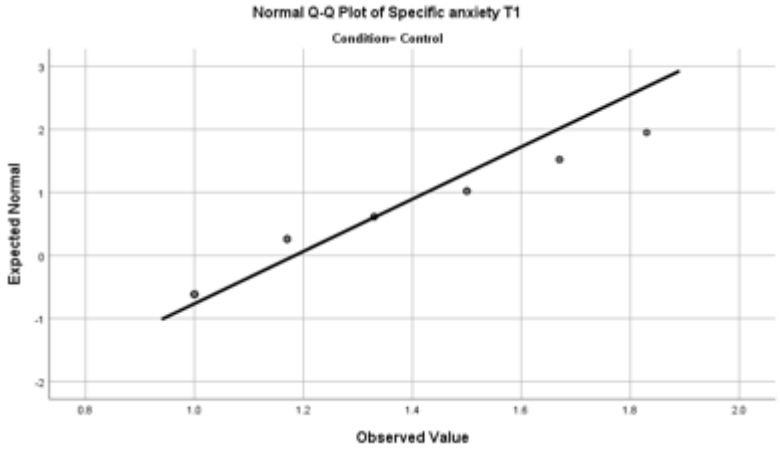


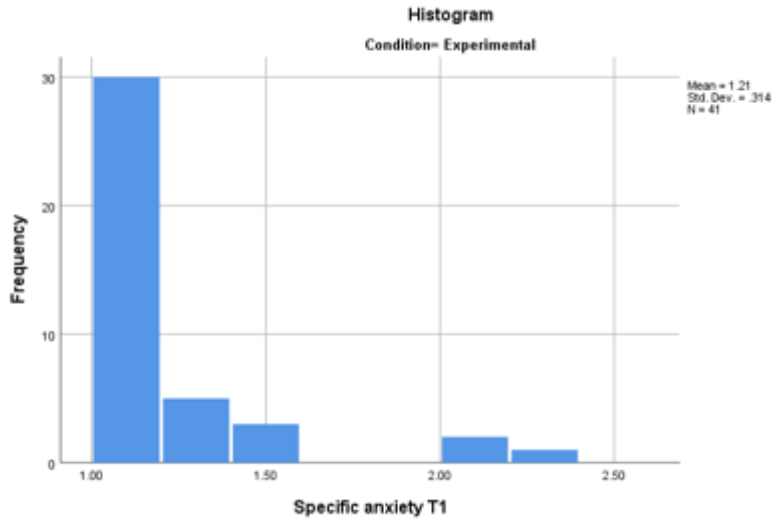


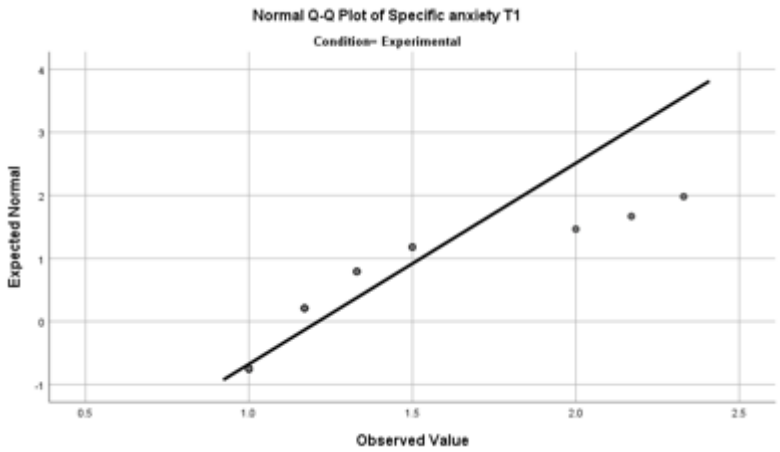


Histograms and Q-Q plots for baseline specific anxiety.

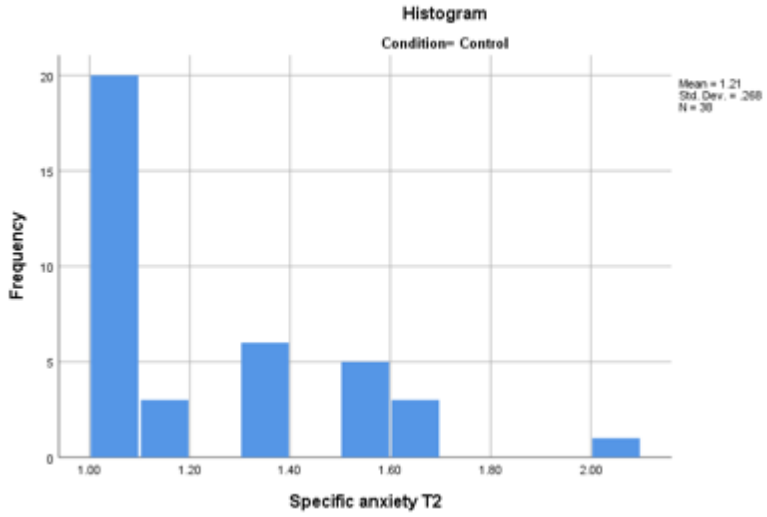

Histogram
Condition= Control
Mean = 1.18
Std. Dev. = .241
N = 38
Frequency
Specific anxiety T1

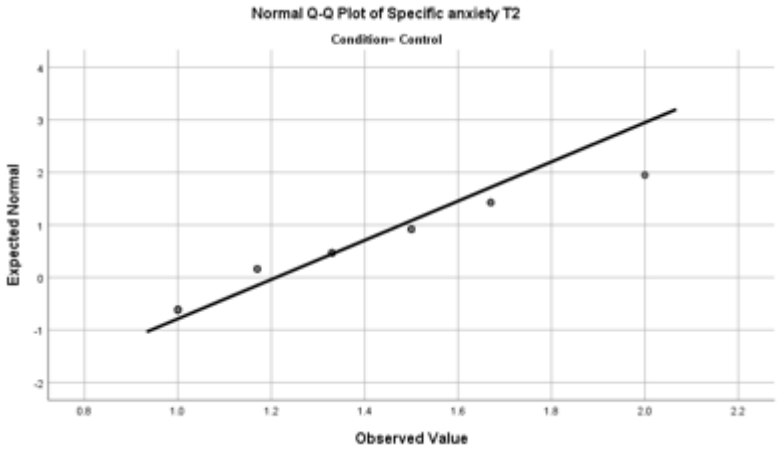

Normal Q-Q Plot of Specific anxiety T1
Condition= Control
Expected Normal
Observed Value

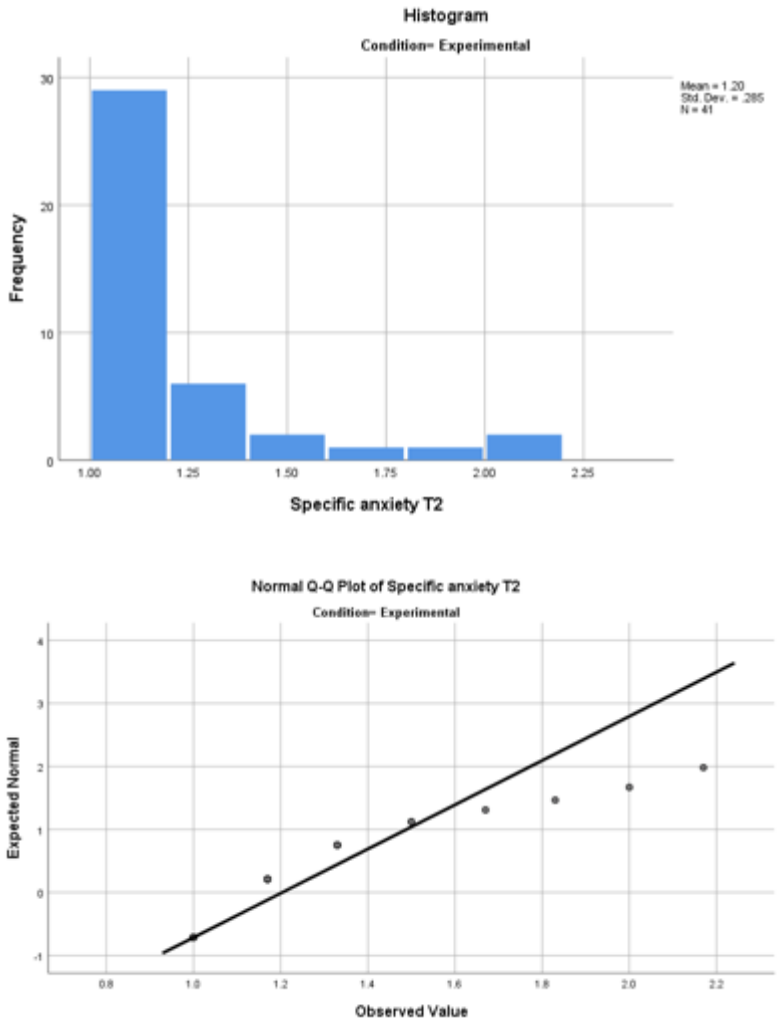

Histogram
Condition= Experimental
Mean = 1.21
Std. Dev. = .314
N = 41
Frequency
Specific anxiety T1



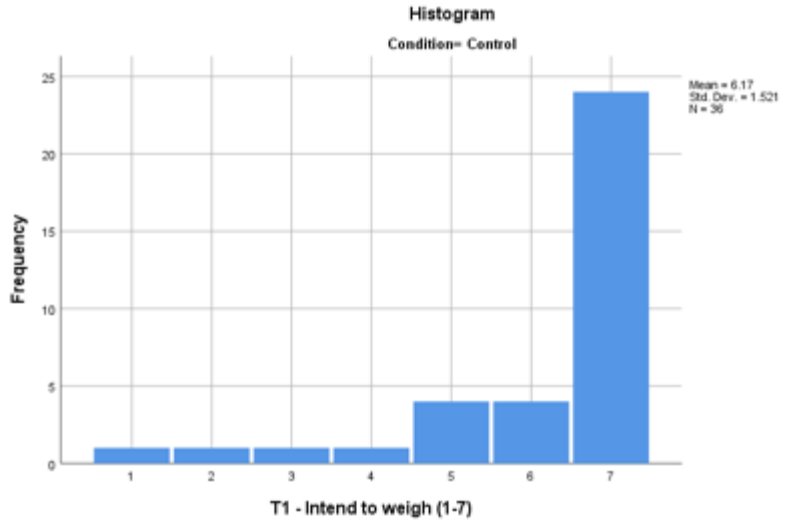
Histograms and Q-Q plots for post-intervention specific anxiety.

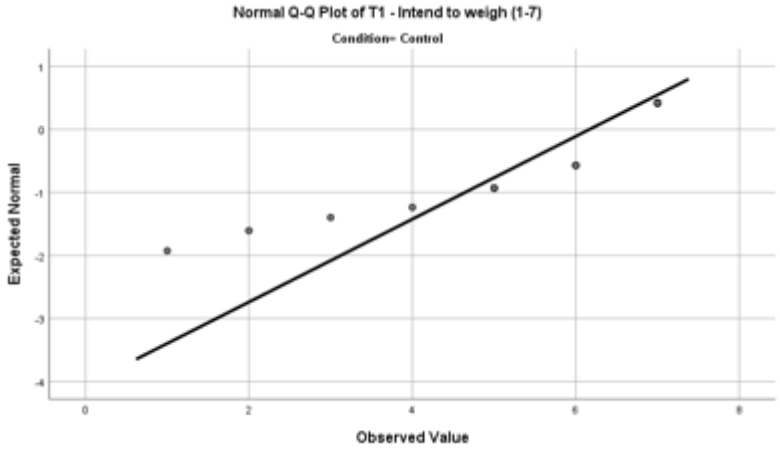


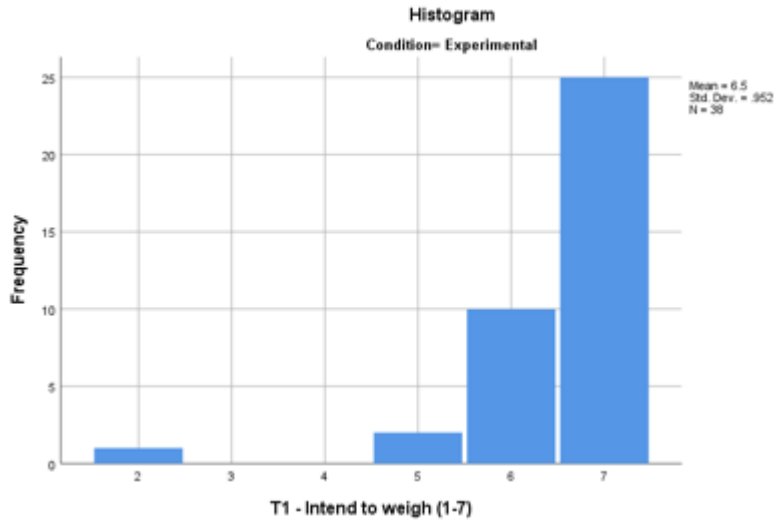


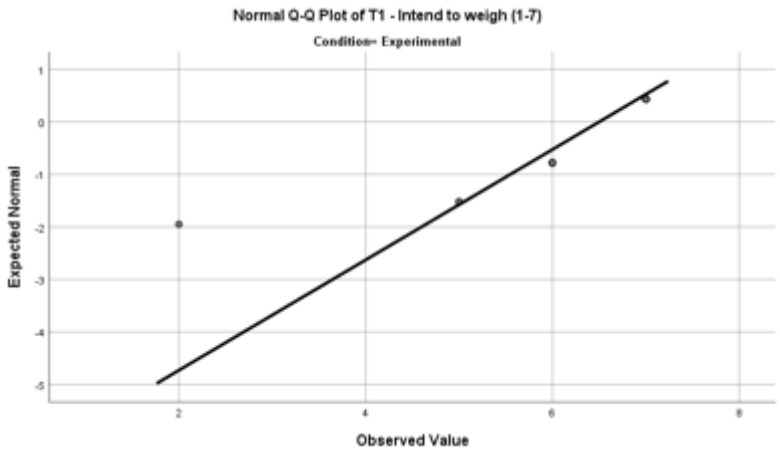


Histograms and Q-Q plots for baseline intentions to weigh.

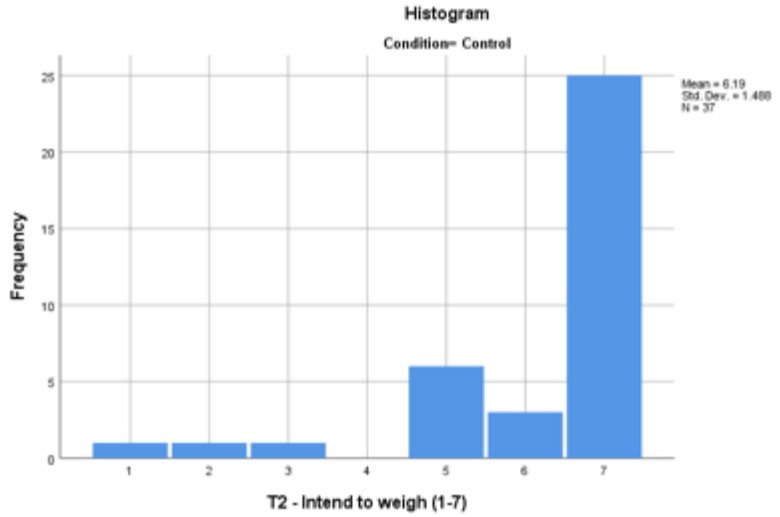

Histogram
Condition= Control
Mean = 6.17
Std. Dev. = 1.521
N = 36
Frequency
T1 - Intend to weigh (1-7)

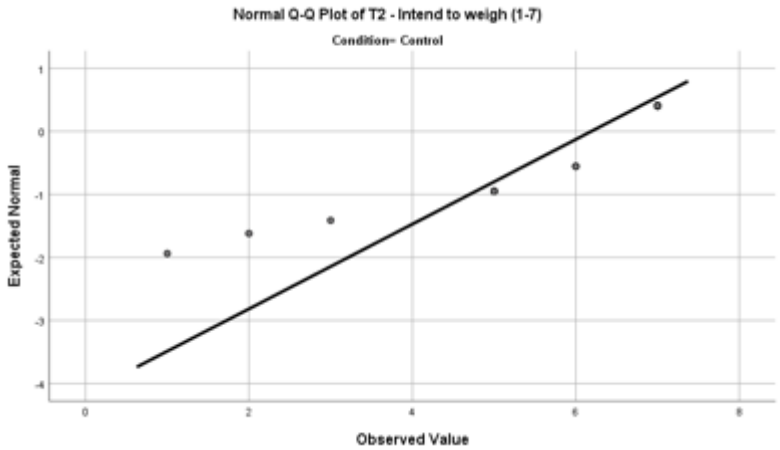

Normal Q-Q Plot of T1 - Intend to weigh (1-7)
Condition= Control
Expected Normal
Observed Value

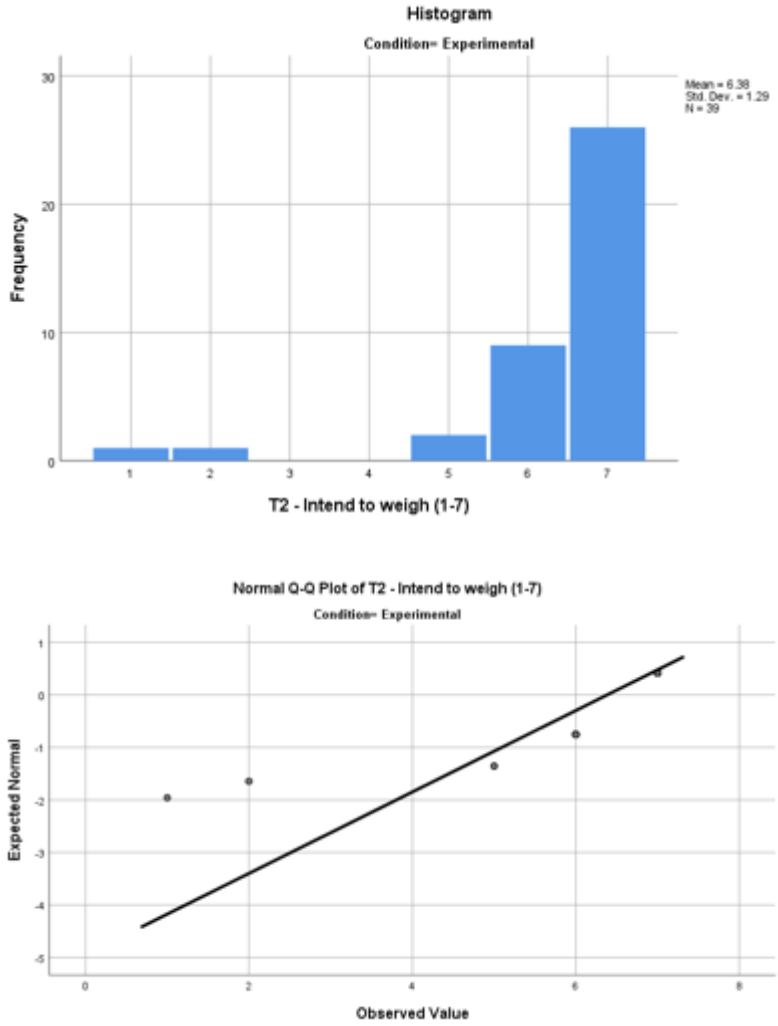

Histogram
Condition= Experimental
Mean = 6.5
Std. Dev. = .952
N = 38
Frequency
T1 - Intend to weigh (1-7)



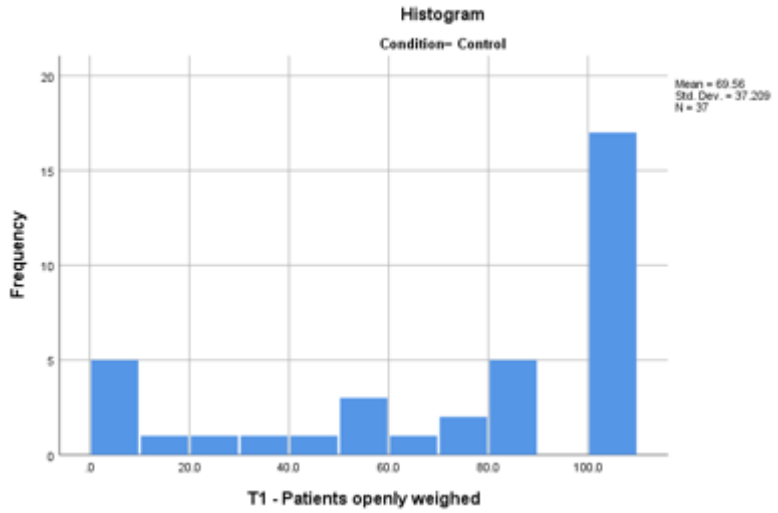
Histograms and Q-Q plots for post-intervention intentions to weigh.

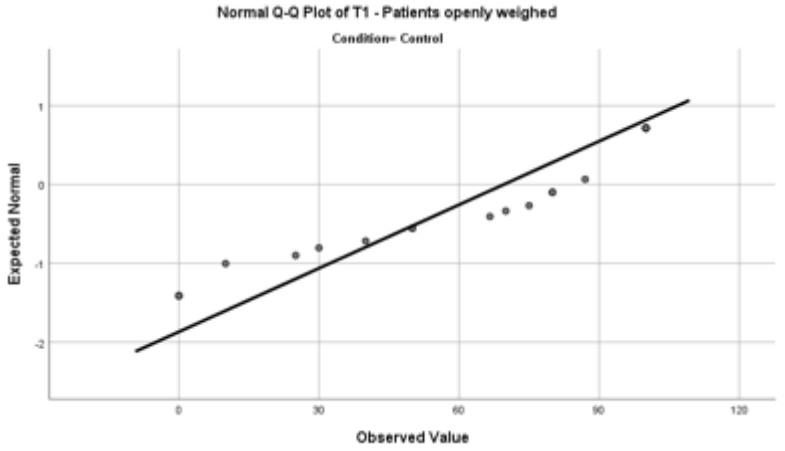


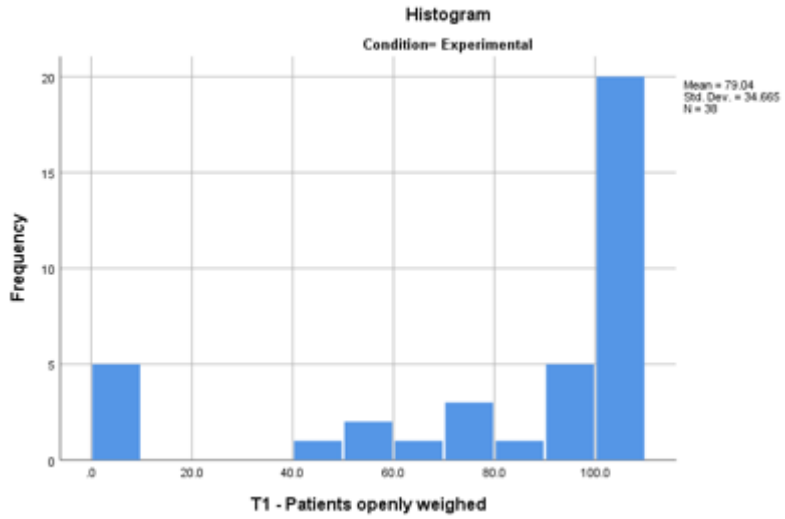


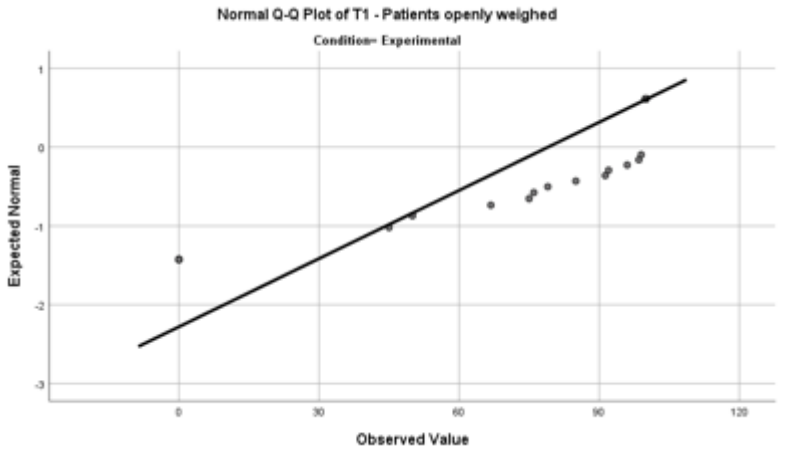


Histograms and Q-Q plots for baseline weighing behaviour.

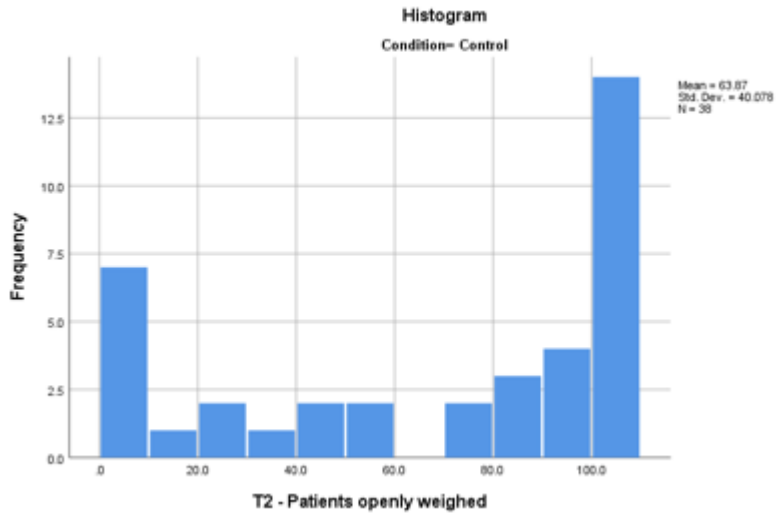


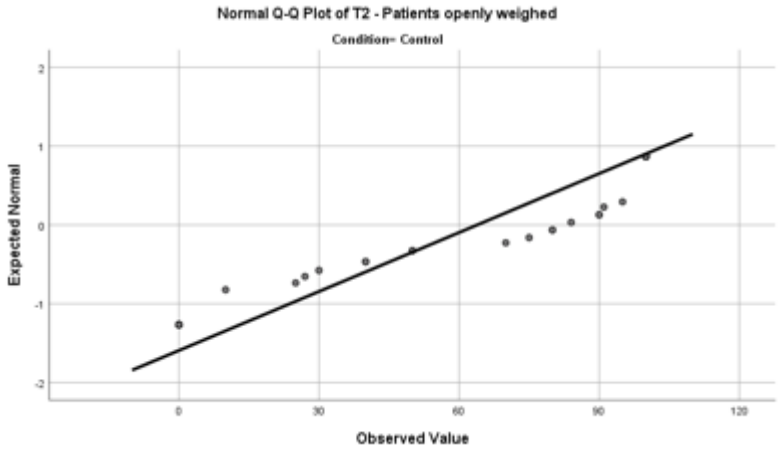


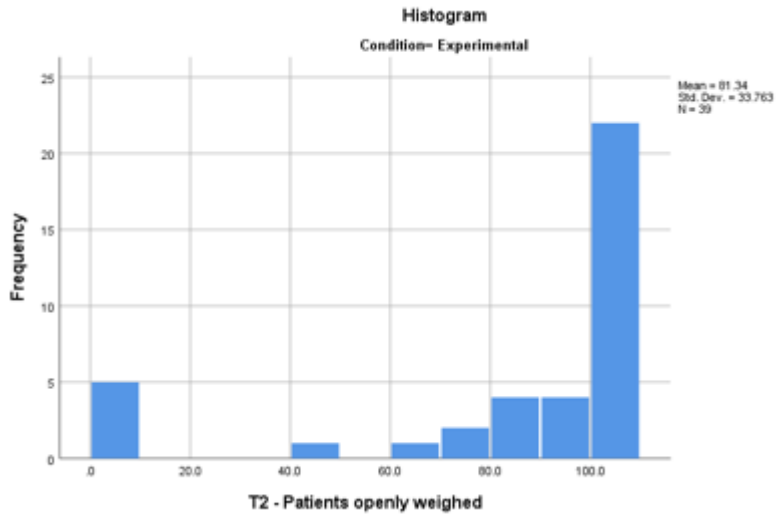


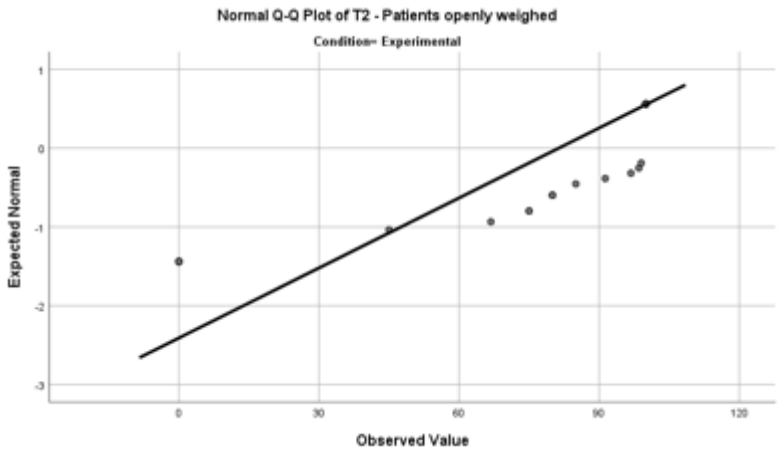


Histograms and Q-Q plots for post-intervention weighing behaviour.

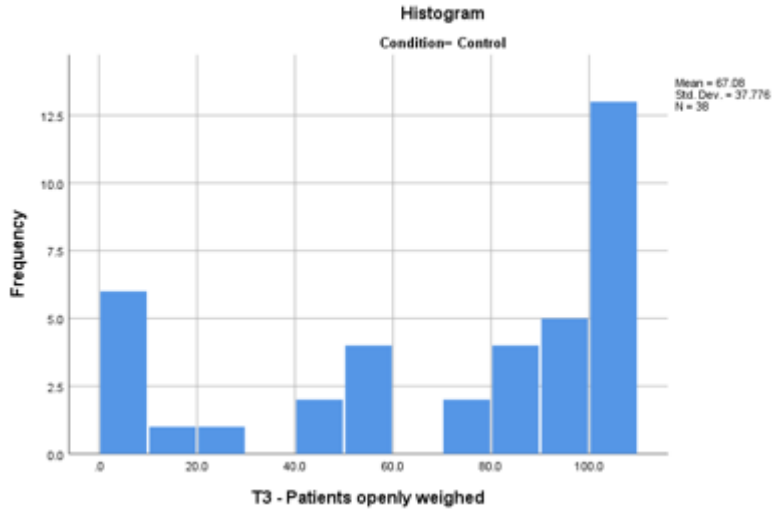

Histogram
Condition= Control
Mean = 63.87
Std. Dev. = 40.078
N = 38
Frequency
T2 - Patients openly weighed

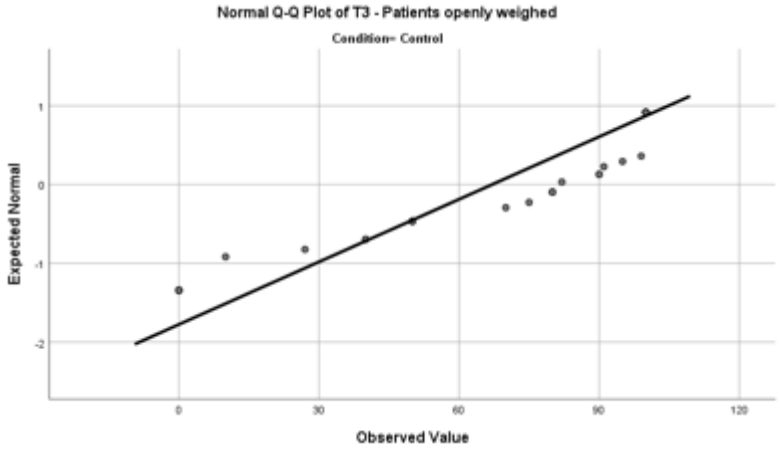

Normal Q-Q Plot of T2 - Patients openly weighed
Condition= Control
Expected Normal
Observed Value

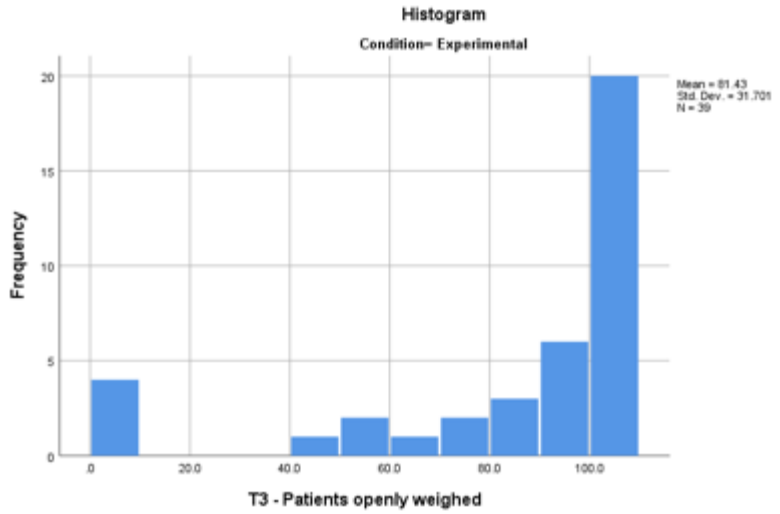

Histogram
Condition= Experimental
Mean = 81.34
Std. Dev. = 33.763
N = 39
Frequency
T2 - Patients openly weighed

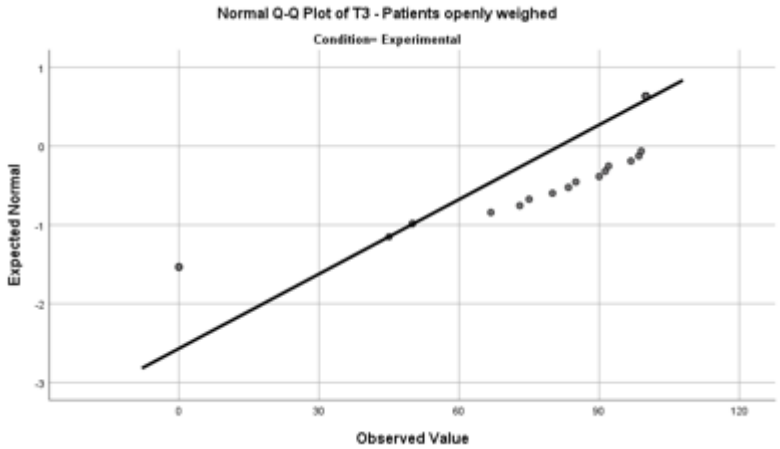


Histograms and Q-Q plots for follow-up weighing behaviour.




Normal Q-Q Plot of T3 - Patients openly weighed
Condition= Control
Expected Normal
Observed Value
1
0
-1
-2
0
30
60
90
120


Histogram
Condition= Experimental
Mean = 81.43
Std. Dev. = 31.701
N = 39
Frequency
T3 - Patients openly weighed
20
15
10
5
0
.0
20.0
40.0
60.0
80.0
100.0


Normal Q-Q Plot of T3 - Patients openly weighed
Condition= Experimental
Expected Normal
Observed Value
1
0
-1
-2
-3
0
30
60
90
120