
The University Of Sheffield.

# An Exploration of Sudden Gains in Psychotherapy

**Holly McGrellis**

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

The University of Sheffield

Faculty of Science

Clinical and Applied Psychology Unit, Department of Psychology

Submission Date: 31st May 2023

## Declaration

I, the author, declare that the thesis is my own work. The thesis has not been previously presented for an award at the University of Sheffield, or any other institution. I am aware of the University of Sheffield's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means) and declare that the thesis presented is original and all other sources have been referenced accordingly.

## Structure and Word Count

**Section One: Literature Review**

| | |
|---|---|
| Excluding references and tables: | 6,676 words |
| Including references and tables: | 11,537 words |

**Section Two: Research Report**

| | |
|---|---|
| Excluding references and tables: | 7,490 words |
| Including references and tables: | 10,170 words |

**Total Word Count:**

| | |
|---|---|
| Excluding references and tables: | 14,166 words |
| Including references and tables: | 21,707 words |

## Lay Summary

Sudden gains were identified by Tang and DeRubeis (1999) as large and stable reductions in symptoms that occur between therapy sessions. They found individuals who experienced sudden gains had better results at the end of treatment compared to those who did not have a sudden gain. Since their identification, many studies have investigated sudden gains and found they occur across different types of therapy, age groups, and in a variety of mental health diagnoses. The factors that cause sudden gains and the improved outcomes have been investigated and a new theory of sudden gains has been proposed (Aderka & Shalom, 2021). The new therapy suggests that sudden gains are caused by natural symptom fluctuation that interacts with therapy, and causes the large, stable reductions in symptoms known as sudden gains.

However, a challenge when researching sudden gains is the criteria that are used to identify the sudden gain. The original identification criteria were proposed by Tang and DeRubeis (1999). However, this criteria has been criticised and many adaptations have been made to the criteria. A large review of the literature (Shalom & Aderka, 2020) found that studies using adapted sudden gains criteria had a greater impact on treatment results compared to studies using the original criteria. They recommended caution when comparing sudden gains findings using different criteria and suggested that future research should conduct sudden gains research using the original and an altered criteria to develop a better understanding of the differences between the criteria.

The current research reviewed the literature to understand the adaptations that have been made to the sudden gains criteria and the impact that these adaptations have on treatment outcomes following SGs. The review found that a variety of adapted criteria have been used to identify sudden gains and the extent of the adaptations are potentially reducing the legitimacy of sudden gains research. The review found a difference between

the altered sudden gains criteria, with one criteria showing a large impact on outcomes and the other two criteria showing small impacts on outcomes. The differences between the criteria need to be understood further and, moving forwards, sudden gains research would benefit from considering how sudden gains criteria can be applied in real life to support the identification of sudden gains in a therapy setting.

A research study was also conducted to explore the new model of sudden gains, whilst continuing to think about the different sudden gains criteria. The aims of the research were to determine whether high levels of natural symptom fluctuation in depression during treatment led to sudden gains, and whether sudden gains led to improved treatment outcomes in two different therapies for depression. These aims were explored using two different sudden gains criteria to determine any differences between the criteria. The research found that using the original criteria, sudden gains occurred in 30.8% of the sample, and in 43.9% when using the altered criteria. Higher natural symptom fluctuation in depression was found to be linked to sudden gains, but only when using the altered sudden gains criteria. Sudden gains were found to be linked to improved treatment outcome, but only when using the original criteria. These findings provide mixed support for the revised theory of sudden gains, and the research further highlights the importance of creating a standardised criteria to identify sudden gains.

## Acknowledgements

Firstly, I would like to thank the entire department of Clinical Psychology at the University of Sheffield for getting me though three wonderfully challenging years of training. With special mention to my academic tutor Vyv Huddy, my clinical tutor Charlotte Merriman, and all of my brilliant placement supervisors.

This research project would not have been possible without my thesis supervisor, Gillian Hardy. Thank you for your support and guidance, and the important sudden gains research that you and your colleagues conducted, which paved the way for this research project. Thank you to Dave Saxon for being the guardian of the PRaCTICED trial data, and for supporting me with my statistical analysis. And thank you to Michael Barkham, who asked the hard questions, and encouraged me to delve deeper into the topic.

To my peers and friends Connie Newcombe and Kerry Ardern, I will forever be grateful for your support. You have both been there to discuss the challenges and always supported me find a solution. Connie, thank you for giving me a push to get started and for the many study days spent together in Couch. Kerry, thank you for sharing your research knowledge, and especially your expertise in the PRaCTICED trial.

Finally, I would like to thank my friends and family for supporting me throughout training. With special mention to my sisters; Jessica who is my loyal proof-reader, and Florence who always puts a smile on my face. Tom, you have been by my side through all the highs and lows and I cannot express how much that has meant to me, thank you.

## Contents

## Section One: Literature Review

# A Systematic Review and Meta-Analysis of the Altered Criteria Used to Identify Sudden Gains in Psychotherapy


## Abstract

**Objectives:** The aims of this review were to investigate how the criteria used to identify sudden gains (SGs) have been adapted, and what impact these adaptations have on the association between treatment outcomes and SGs.

**Methods:** A systematic review of the literature was conducted followed by data extraction, quality assessment, and a meta-analysis, with two subgroup analyses.

**Results:** A total of 17 papers were identified, which included three adapted SGs criteria. 15 adapted SGs criteria were identified, however the majority could not be included due to being used in fewer than three studies. Quality appraisal found the studies to be of moderate to low quality. The meta-analysis found a significant difference in outcomes between SG and non-SG groups, with an effect size of $g$ = .51. Subgroup analysis between the SGs criteria found that the combined Tang and Gaynor criteria had an effect size of $g$ = .94, whereas the Hardy and Kelly criteria had effect sizes of $g$ = .42 and $g$ = .39 respectively.

**Conclusions:** The extent of the adaptations to the SGs criteria is potentially reducing the validity of SGs research. The differences between the criteria need to be understood further. A strength of the combined Tang and Gaynor criteria may be its strictness combined with its ability to identify early SGs, which creates a good association with treatment outcomes. Moving forwards, SGs research must consider how the SG criteria can be applied in a clinical setting to support the improved treatment outcomes related to SGs.

**Keywords:** Sudden Gains; Sudden Gains Criteria; Psychotherapy.

**Practitioner Points**

- Clinicians should familiarise themselves with the SGs criteria, especially the first two criteria as they would be calculable in a clinical environment.
- Clinicians should administer sessional outcome measures with the intention of looking for large and stable improvements that occur between treatment sessions and may indicate that a SG has occurred.
- Clinicians should familiarise themselves with the 'upward spiral' (Aderka & Shalom, 2021) following a SG and aim to include these aspects into therapy following a potential SG.
- Clinicians should be aware that SGs are less likely to occur in treatment for anxiety disorders but can still occur and should still be monitored. Also, that SGs in other conditions occur more frequently, and can occur more than once throughout the course of therapy.
- Clinicians should be trying to identify SGs when using all models of therapy.

# A Systematic Review and Meta-Analysis of the Altered Criteria Used to Identify Sudden Gains in Psychotherapy

## Introduction

The phenomenon of sudden gains (SGs) was originally identified and defined by Tang and DeRubeis (1999). They described SGs as large and stable reductions in symptoms that occur between consecutive therapy session. For research purposes they created three criteria which are described below:

(a) the gain was at least 7 points on the Beck Depression Inventory (BDI; Beck & Steer, 1987);

(b) the gain represented at least 25% of the pre-gain session's BDI score;

(c) the mean BDI score of the three therapy sessions before the gain was significantly higher than the mean BDI score of the three therapy sessions after the gain using a two-sample t test, with an alpha of .05. When only two scores were available on either side of the SG the t-test could be run with 5 session scores instead of 6.

In their original investigation Tang and DeRubeis (1999) found that SGs occurred in 39.34% of participants and the magnitude of the SGs was 11.2 (*SD* = 4.4) BDI points. In addition, they found that participants who experienced a SG had superior outcomes compared with participants who did not have a SG by the end of treatment, and at six and 18 month follow up.

Tang and DeRubeis (1999) developed a model to explain the phenomenon of SGs which suggested that cognitive behavioural therapy (CBT) generates cognitive changes, which lead to improved alliance between the therapist and patient, which then leads to

additional cognitive change and facilitates rapid symptom improvement and overall superior outcomes.

Since this seminal work, the research area of SGs research has received considerable attention, with over 100 studies investigating the phenomenon. As the literature relating to SGs has grown, it has also become more diverse, exploring SGs in a range of treatment modalities (Doss et al., 2011), mental health conditions (Durland et al., 2018), age groups (Aderka et al., 2011), and treatment settings (Drymalski & Washburn, 2011). Additionally, the research has focussed on a variety of aspects of SGs for example, early SGs (Stiles et al., 2003), pre-gain processes (Durland et al., 2018), post-gain processes (Zilcha-Mano, 2019), as well as mediating and moderating factors associated with SGs (Shalom & Aderka, 2020). Many studies have replicated the results found by Tang and DeRubeis (1999) (Abel et al., 2016, Gaynor et al., 2003, Zilcha-Mano, 2019). However, many studies have found contradictory results. For example, Kelly et al. (2005) found that participants who experienced a SG did not have superior outcomes compared to participants without a SG.

**Sudden Gains Criteria**

Given the above evidence, it is generally agreed that SGs are a pan-theoretical phenomenon that occur across treatment modalities, mental health conditions, age groups, and in all treatment settings. However, a complicating factor in our understanding of SGs concerns the criteria used to identify the SG.

The original SGs criteria (Tang & DeRubeis, 1999) has been criticised on several aspects. Firstly, Tang and DeRubeis (1999) themselves described the seven-point cut off on the BDI as “arbitrary”. Stiles et al. (2003) recognised that seven points on the BDI is very close to the Reliable Change Index (RCI; Jacobson & Truax, 1991) score for this measure. Stiles recommended adapting the first criteria to reflect a change of the RCI

score. As well as providing empirical justification for the change score, this adaptation provided standardisation and allowed for the SGs criteria to be applied to measures other than the BDI. Since this time the majority of studies have applied this adaptation to the original criteria, especially studies which do not use the BDI.

Some criticisms have been made regarding the second criterion of the original SGs criteria (i.e., a gain representing at least 25% of the pre-gain session's BDI score). Hardy et al. (2005) described it as “problematic” as it assumes that the outcome measure being used is a ratio scale when, in fact, they are interval scales. Hardy et al. (2005) also noted that when including this criterion, no additional SGs were found when using the other two criteria, therefore making it redundant. The majority of research has included this criterion and no other criticisms have been made. However, this may be because this criterion has limited impact on the identification of SGs.

The third criterion (i.e., the comparison of pre and post SG scores) has received the largest number of criticisms and it is the criterion that has been most widely adapted. Vittengl (2005) described the use of a t-test as problematic because the three pre-gain and three post-gain scores are not independent observations, and therefore positive autocorrelation effects are likely to inflate the *t* value, whilst negative autocorrelation effects are likely to deflate the *t* value, so the comparison is not a valid inferential test. Hardy et al. (2005) echoed these concerns and as a result adapted the *t* value which they applied. Tang et al. (2005) recognised this issue and whilst they did not change how the criterion was applied, they adapted the wording of the criteria so that the mean difference between the BDI scores of the three sessions before the gain and the three sessions after the gain was at least 2.78 times greater than the pooled standard deviations of these two groups of sessions. This adaptation was equivalent to the original third criterion but reworded to better follow statistical convention.

Another criticism of this criterion has been its inability to identify SGs that occur early in therapy. Indeed, this is the main criticism as previous literature has established that significant change can occur early in therapy (Stiles et al., 2003). Gaynor et al. (2003) and Kelly et al. (2005) both made adaptations to this criterion which allowed for SGs to be identified between all therapy sessions.

However, the third criterion has also been misinterpreted due a lack of clarity in the original description. When applying the original criteria, some researchers appear not to be aware that Tang and DeRubeis (1999) allowed for SGs to be identified when only two pre-gain or two post-gain scores were available. Using this correct method, SGs cannot be identified following the first or before the final session. However, some researchers have applied the criterion with three sessions pre- and post-gain and have therefore been unable to identify SGs between the first two and final two sessions.

Following these original adaptations to the SGs criteria, many other adaptations have been made in order to suit the requirements of individual research. Whilst this diversification has allowed for different aspects of change to be explored (e.g., Cartwright et al. (2017) looked at SGs in changes in body weight), the increased variation poses the risk that SGs research using different criteria is investigating different underlying phenomena.

In a recent meta-analysis, Shalom and Aderka (2020) found that SGs identified using an altered SGs criteria have larger effect sizes ($g$ = .72) than SGs identified using the original criteria ($g$ = .63). Following this finding, it has been suggested that comparisons between SG research based on different criteria should be conducted with caution, and that SGs based on different criteria may represent partially distinct psychological constructs with unique characteristics. Shalom and Aderka (2020) stressed the need for standard criteria to be created and asked that future studies report findings

using both original and altered criteria to facilitate the creation of standard criteria based on empirical evidence.

There are a number of possible explanations for the larger effect sizes found using altered SGs criteria in Shalom and Aderka's (2020) review. One explanation could be the original criteria not including the identification of SGs in the first or last sessions, and sometimes the first two and last two sessions, whereas some altered criteria include the identification of SGs at all timepoints (Gaynor et al., 2003; Kelly et al., 2005). Some evidence has found that early gains may be especially predictive of outcome in psychotherapy (Singla et al., 2019), and it is possible that the identification of early gains increased the effect size. However, other evidence found that SGs which occur following the first session were not predictive of improved treatment outcomes (Clerkin et al., 2008). Therefore, a better understanding of early change and improved outcomes related to SGs is needed.

Another potential explanation for the finding is that the original criteria are too stringent and fail to identify some important gains in treatment. The third criterion has been described as particularly stringent, as the *t*-test only has three observations before and after the gain, it may be insufficiently powered to detect smaller changes. Whereas, the altered criteria may be able to detect smaller gains. For example, Hardy et al. (2005) relaxed this criterion so that $t\ (4) > 2.50$, $p < .05$ when comparing three pre-gain scores with the three post-gain scores, or $t\ (3) > 3.00$, $p < .05$, if only two pre-gain or two post-gain scores were available.

As the above explanations are speculative, more empirical research is needed to improve the understanding of the impact that altering the SGs criteria could have on the association between SGs and outcomes. The current review aims to investigate how the

SGs criteria have been adapted and what impact these adaptations have on treatment outcomes associated with SGs.

### Research Questions

1. Is there a difference in post treatment outcomes between SG and non-SG groups, in the studies which identify SGs using adapted criteria?
2. Do the different adapted criteria used to identify SGs moderate the effect of SGs on treatment outcome?

## Method

The protocol for this meta-analysis was formally registered on the Prospero database under the registration number: CRD42023409210.

### Identification and Selection of Studies

A population, intervention, comparison, outcomes and study (PICOS) framework (Amir-Behghadami & Janati, 2020) was adopted to refine the inclusion and exclusion criteria used to identify studies that were appropriate to be included in the current review. Table 1 presents the inclusion and exclusion criteria using the PICOS framework.

A systematic literature search was conducted on the electronic databases: Scopus; Medline (Via OvidSP); Web of Science; CINAHL (Via EBSCO); PyscInfo (Via OvidSP) and ProQuest with the final search taking place on 1st May, 2023. The Boolean operators 'AND' and 'OR' were used to combine the terms "sudden gain", "psychotherap*", "thera*", and "interven*". The search term "criteri*" was not included as it identified a large amount of studies which were not relevant. For PsycInfo and Medline, key word searches were performed and then combined. When searching the grey literature using ProQuest, the first 100 papers were screened, as they are sorted by relevance. This method has been deemed adequate by Stevinson and Lawlow (2004).

***Table 1***.

*Inclusion and Exclusion Criteria for Study Selection Based on the PICOS Framework.*

| | Inclusion | Exclusion |
|---|---|---|
| Population | Individuals undergoing psychotherapy<br>Individuals of any age | |
| Intervention | Any psychological therapy<br>Individuals receiving direct or group interventions | Non-psychological interventions<br>Systemic interventions e.g. couple therapy |
| Comparison | Any SGs criteria altered from the original Tang and DeRubeis (1999) criteria | The altered SGs criteria must be used in three or more studies<br>Details of how the SGs criteria have been adapted are not described |
| Outcome | Sessional outcome data measured using a validated measure | |
| Study Type | Quantitative studies<br>Published in the English language | Qualitative studies<br>Review studies<br>Case studies<br>Stimulated data<br>Re-analyses of published SGs data |

Note. SG = Sudden Gain

The search terms were applied to the title, abstract and keyword sections. The searches were filtered to only include studies that were published after 1999, when the

original Tang and DeRubeis (1999) study was published. The key recent systematic reviews in this area (Shalom & Aderka, 2021 and Silverstone et al. 2023) were reviewed using forward and backward citation searching to identify additional papers.

Search results from the electronic databases were exported to the reference manager, EndNote (2013). Duplications were removed, and the studies were screened by title and abstract against the eligibility criteria. The remaining studies were then screened by full text. Screening and selection tools were used in all stages to aid consistency. Titles and abstracts were screened by the first author only. A randomly selected subset of 25% of the full texts were reviewed by a peer (CN, trainee clinical psychologist) and 100% agreement was reached on which papers to include.

**Data Extraction**

An overview of the general characteristics of the studies were extracted, including: study design; country of study; study setting; diagnosis; mode of therapy; maximum number of sessions; sample size; percentage of females; and mean age and standard deviation (SD). An overview of the study characterises relating to SGs and treatment outcomes was also extracted including: SG criteria; measure used to identify SGs; SG percentage frequency; medium pre-gain session; mean magnitude of SGs; SG reversal rate; and details of any association between SG and outcome. The data was extracted by the first author and a randomly selected subset of studies (*n* = 7) were checked by peer (CN) and a 100% level of inter-rater reliability was achieved.

**Definition of the Sudden Gains Criteria**

In order to investigate the difference between the original criteria and altered SGs criteria, it was important to develop a thorough understanding of the original criteria and the subsequent adaptations that have been made. Table 2 was created to described the

different sudden gains criteria, how the criteria have been classified (original or adapted), how many and which studies used the criteria.

**Quality Appraisal**

This systematic review has been carried out and reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) checklist (Page et al., 2021, Appendix A)

The quality of the included studies was assessed using the Quality Assessment Tool for Quantitative Studies by the Effective Public Health Practice Project (EPHPP). The EPHPP tool focuses on six areas: selection bias, study design, confounders, blinding, data collection methods, withdrawals and dropouts. It defines the research as ‘strong’, ‘moderate’, or ‘weak’. Studies rated as ‘strong’ have no weak rating across the six areas, studies rated as ‘moderate’ have one weak rating across the six areas, and studies rated as ‘weak’ have two or more weak ratings across the six areas. This tool was selected as it has good construct validity (Thomas et al., 2004). It was also found to have ‘excellent’ inter-rater agreement for the final grade compared to the Cochrane Collaboration Risk of Bias Tool which had ‘fair’ inter-rater agreement for final grade (Armijo-Olivo et al., 2012).

The quality assessment was conducted on a study level as opposed to an outcome level and studies were not excluded based on their quality. The first author conducted the quality assessment and a randomly selected subset of studies ($n$ = 7) were checked by peer (CN) as most quality assessment tools have a high level of subjectivity. A Cohen’s *Kappa* Coefficient (1992) of .42 was found suggesting a moderate level of inter-rater reliability, as defined by Landis and Koch (1977). Disagreements were discussed between the raters and 100% agreement was reached.

***Table 2.***

*A Table Describing the Adapted SGs Criteria, How They Have Been Classified in this Review, and How Many and Which Papers Have Used the Criteria.*

| Sudden Gains Criteria | Description of the Criteria | Classification | Studies Using the Criteria *(n)* |
|---|---|---|---|
| Tang & DeRubis (1999) | (a) the gain was at least 7 BDI points (b) the gain represented at least 25% of the pre-gain sessions BDI score (c) the mean BDI score of the three therapy sessions before the gain was significantly higher than the mean BDI score of the three therapy sessions after the gain using a two-sample t test, with an alpha of .05. This definition included gains when only two pre-gain or two post-gain session scores were available. | Original | N/A |
| Stiles et al. (2003) | Noted that a 7 point gain on the BDI was close to the BDI's RCI and adapted the criterion to use the RCI instead to make it applicable to other measures. | Original. This alteration is essential to standardising the criteria and has been used in the majority of subsequent research, therefore studies that only change the criteria by using the RCI are classified as original. | N/A |

| Sudden Gains Criteria | Description of the Criteria | Classification | Studies Using the Criteria *(n)* |
|---|---|---|---|
| Gaynor et al. (2003) | Gaynor et al. (2003) maintained the original criteria but also looked at pre-session gains. To analyse this they did not use the third criterion and instead required that at least 50% of the gain be maintained during the next two sessions. | Adapted. Although Gaynor et al. (2003) used the original criteria, subsequent research has adopted the altered criteria. | Ietsugu et al. (2015) *(n= 1)* |
| Tang et al. (2005) | Tang at al. (2005) redefined the third criterion to "the mean difference between the BDI scores of the three sessions before the gain and the three sessions after the gain was at least 2.78 times greater than the pooled standard deviations of these two groups of sessions' BDI scores." | Original. This is a modification to the wording of the criteria, the analysis is conducted in the same way. | N/A |
| Hardy et al. (2005) | (a) used the RCI score (b) did not use the second criterion (c) $t(4)=2.50$, $p=.05$, comparing the three pre-gain scores with the three after-gain scores, or $t(3)=3.00$, $p=.05$, if only two pre-gain or two after-gain scores were available. | Altered. When studies have used this criteria but maintained the second criterion it has been classified as Hardy et al. (2005) as they found no difference when including or excluding this criterion. | Buchholz et al. (2019), Collins & Coles (2017), D'Arcy & Norton (2021), Hardy et al. (2005), Ryan (2018), Silverstone et al. (2023), Vincent & Norton (2019) *(n= 7)* |

| Sudden Gains Criteria | Description of the Criteria | Classification | Studies Using the Criteria *(n)* |
|---|---|---|---|
| Kelly et al. (2005) | Adapted the third criterion to “a gain must reflect an improvement of at least 1.5 standard deviations from the individual mean”. | Altered | Bisby et al. (2022), Clerkin et al. (2008), Kelly et al. (2005), Kelly et al. (2007), Singh et al. (2021), Thorisdottir et al. (2018) *(n=6)* |
| Drymalski & Washburn (2011) | Third criterion was changed to “any scores after the SG would not increase above half of the RCI”. | Altered | Drymalski & Washburn (2011) *(n= 1)* |
| Lorenzo et al. (2013) | The first criterion was changed to “equal or greater than the average change across all sessions plus one standard deviation”. The third criterion was changed to “not followed by any subsequent increase larger than 25% of the total item score”. | Altered | Lorenzo et al. (2013) *(n= 1)* |
| Cavallini & Spangler (2013) | Used the original criteria but without the second criterion. | Altered | Cavallini & Spangler (2013) *(n= 1)* |

| Sudden Gains Criteria | Description of the Criteria | Classification | Studies Using the Criteria *(n)* |
|---|---|---|---|
| Cartwright et al. (2017) | The second criterion was not used. An additional criterion of 'suddenness' was added which required the rate of gain between sessions N and N+1 to be ≥1.5 times the rate change between sessions N-1 and N. | Altered | Cartwright et al. (2017), Brockmeyer et al. (2019) *(n= 2)* |
| Singla et al. (2019) | Adapted the third criterion to "the subsequent sessions to the SG must all have a lower score than the sessions before the gain". | Altered | Singla et al. (2019) *(n= 1)* |
| Malins et al. (2020) | Adapted the third criterion to "the mean score of the three sessions prior to the gain should be at least a reliable change lower than the mean score for the three sessions following the sudden gain". | Altered | Malins et al. (2020) *(n= 1)* |
| Morrison (2020) | The second and third criteria were altered to "(b) a change greater than 1.5 times the average standard deviation (c) symptom changes did not exceed the first criterion (change in RCI) and the second criterion (1.5 times the average standard deviation)." | Altered | Morrison (2020) *(n= 1)* |

| Sudden Gains Criteria | Description of the Criteria. | Classification | Studies Using the Criteria *(n)* |
|---|---|---|---|
| Sloan et al. (2022) | The third criterion was changed to an "increase of reported symptoms of at least 50% of the original improvement of the sudden gain". | Altered | Sloan et al. (2022) *(n= 1)* |
| Combination of Tang & DeRubis (1999) and Gaynor et al. (2003) | Studies that used the original criteria but with the addition of Gaynor et al.'s (2003) alteration to identify early gains. | Altered | Bjureberg et al. (2020), Greenfield et al. (2011), Hamdeh et al. (2019), Hunnicutt-Ferguson et al. (2012) *(n= 4)* |
| Combination of Tang & DeRubis (1999) and Kelly et al. (2005) | Two studies added Kelly et al.'s (2005) criterion to the original criteria and applied them to all cases to detect a SG. | Altered | Hopko et al. (2009), Masterson et al. (2014) *(n= 2)* |
| Combination of Tang & DeRubis (1999) and Singla et al. (2019) | Oliviera et al. (2021) use the original criteria with the addition of the adaptation made by Singla et al. (2019). | Altered | Oliviera et al. (2021) *(n= 1)* |

## Meta-Analytic Strategy

The meta-analysis was conducted using Meta-Essential software version 1.4 (Suurmond et al., 2017). A random-effects meta-analysis, where studies were weighted using the inverse variance method, was used due to sampling error and variability in the population. This meta-analysis was primarily focused on the differences in post-treatment outcomes between those who did and did not have a SG, and therefore Hedge's *g* was used to calculate a standardised effect size. This was done automatically in Meta-Essentials using the post-treatment mean, *SD* and sample size for the SG and non-SG groups. When this data was not available the *F*, *t* or *d* values reported when comparing the outcome of SG and non-SG groups were used to calculate Hedge's *g*. An effect size of .2 was classified as 'small', .5 as 'medium' and .8 as 'large' (Cohen, 1969).

### *Heterogeneity*

As this meta-analysis included studies that employed different modes of therapy, diagnoses, experimental designs, outcome measures, and SGs criteria, it was important to account for between-study variance. Tests of heterogeneity were used to determine if further moderator or subgroup analysis were needed, with high levels of variance supporting the use of further analysis. The *Q*-statistic and the $I^2$ statistic were used to explore heterogeneity. The *Q*-statistic assesses the degree of variability among the pooled effect sizes, and a moderator analysis is indicated when this statistic is statistically significant (Borenstein et al., 2009). However, the *Q*-statistic is dependent on the number of studies included in the meta-analysis, and smaller study numbers increase the risk of erroneously assuming homogeneity (Higgins et al., 2003). The $I^2$ statistic is a measure of between-study variation that is not due to sampling error. Higgins et al. (2022) suggest that an $I^2$ of less than 40% may not be important, 30-60% indicates moderate heterogeneity, 50% to 90% denotes substantial heterogeneity, and 75% or over represents large heterogeneity. It was expected that heterogeneity would be high due to the differences in

the included studies. When the $I^2$ statistic is large then the combined effect size is no longer meaningful and the focus of the analysis should be on the dispersion of true effect sizes and of its moderators. Forest plots were used to visualise the effect sizes and 95% confidence intervals (CIs) were produced.

### *Subgroup Analysis*

Subgroup variables were specified *a priori* and ran when significant heterogeneity was present. These were determined based on variables explored in the literature around the predictors of SGs. Sub-group moderation analysis was used rather than meta-regression as the variables were categorical.

To explore differences in the effect sizes of SGs using the different SGs criteria, three sub-groups were created, identified by the original author names: (1) Hardy; (2) Kelly; and (3) Tang and Gaynor. Additionally, the diagnosis being treated were categorised as: anxiety disorders; depressive disorders; and obsessive-compulsive disorders (OCD). The studies by Bjureberg et al. (2020) and Greenfield et al. (2011) were not included in this subgroup analysis due to not fitting into the diagnosis categories.

As suggested by Card (2012), subgroup analysis based on categorical variables can only be performed when there were more than 10 studies in the analysis and the subgroups consisted of three or more studies. Due to this, a sub-group moderator analysis for therapy modality was not able to be conducted due to there being less than three studies in all therapy modalities other than CBT.

### *Publication Bias*

Whilst attempts have been made to reduce publication bias by including the grey literature, it is best practice to assess for publication bias due to this remaining a large threat to the validity of this meta-analysis. As this meta-analysis focuses on the difference between SG and non-SG groups in treatment outcome, there could be a bias towards the publication of papers that only report significant differences between the groups and this

could impact the included papers. Publication bias will be assessed through visual inspection of funnel plots, with the trim and fill method used when plots are asymmetrical. Egger's regression test, and the fail-safe *N* using the Rosenthal (1979) method will also assess publication bias.

## Results

### Study Selection

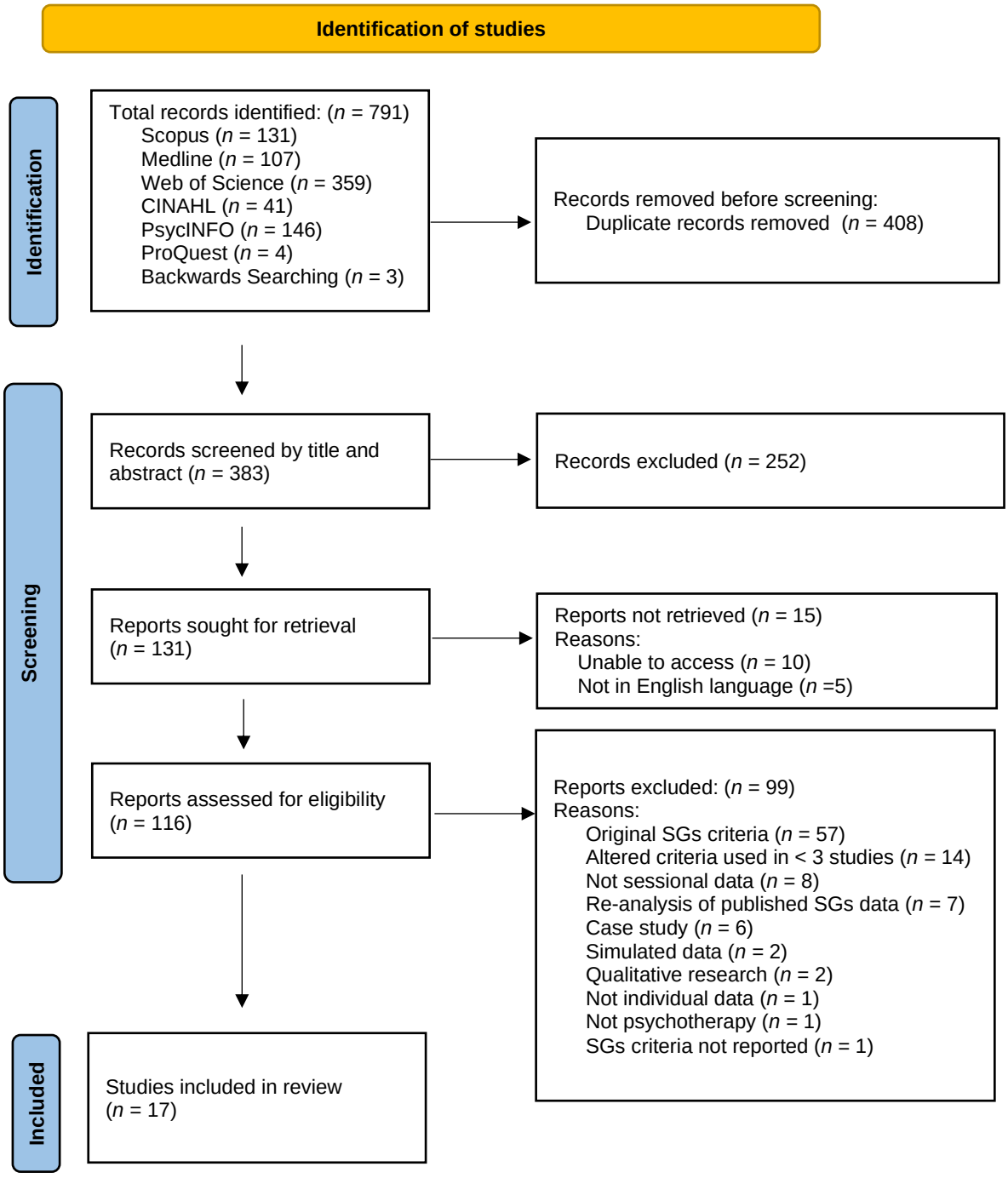
As presented in the PRISMA diagram (Figure 1), the literature search initially identified 788 articles through electronic database searching and three articles were identified through additional sources. After removing the duplicates (*n* = 408), 383 studies remained to be screened by the title and abstract. A total of 252 papers were excluded at this stage and a further 15 papers were inaccessible. This resulted in 116 papers being screened by full text using the eligibility criteria. The main reasons for not including articles in the review were: the original SGs criteria was used; the altered criteria was used in less than three studies; the data was not sessional; the paper was a re-analysis of published SGs research; and the paper described a case study. Following this final stage of screening, a total of 17 papers were eligible to be included in the review.

### Study Characteristics

Table 3 presents an overview of the general characteristics of the studies included in the meta-analysis. Of the 17 studies (*k* = 20), 12 were conducted in North America, three in Northern Europe, two in Australia, and one in the United Kingdom. Seven studies used secondary data analysis of an RCT, one study used secondary data analysis of a controlled clinical trial, and the remaining nine studies used cohort methodology. All studies took place in outpatient settings. One study was an unpublished thesis, and the remaining studies were published in peer-reviewed research journals.

***Figure 1***.

*PRISMA Diagram Detailing the Study Selection Process.*




Eight studies focussed on anxiety-based disorders (general anxiety disorder *n* = 4, social anxiety disorder *n* = 2, and panic disorder *n* = 2). Seven studies focussed on major depressive disorder, three studies focussed on OCD, one study focussed on body

dysmorphic disorder, and one study focussed on diverse disorders. A total of 15 studies included CBT based treatments (online CBT *n* = 2, Exposure and Response Prevention *n* = 2, transdiagnostic CBT (*t*CBT) *n* = 2, individual CBT *n* = 1, cognitive therapy *n* = 2, behavioural activation *n* = 1, group CBT *n* = 3, group *t*CBT *n* = 1). One study used interpersonal therapy, one study used experiential therapy, and one study used group psychotherapy. The maximum number of sessions ranged from eight to 30, and the sample sizes ranged from 23 to 259 participants. All included studies were conducted with adults. One study included only females and the remaining studies included male and female participants.

Table 4 presents an overview of the study characteristics related to SGs and treatment outcomes. Seven studies used the SGs criteria defined by Hardy, six studies used the SGs criteria defined by Kelly, and four studies used a combination of the Tang and DeRubeis and Gaynor criteria. Thirteen different validated, self-report measures were used to identify SGs, with the most common being the BDI-II (Beck et al., 1996; *n* = 5). The frequency of SGs ranged from 14% to 63.9%, and the medium pre-gain session ranged from 2-11. The mean magnitude of SGs ranged from .96 to 2.3 however, this data was unavailable for some studies. The percentage of SG reversals ranges from 8% to 58%, again this data was unavailable for some studies. Finally, SG were associated with improved treatment outcomes when compared non-SG groups in 13 out of 20 studies.

***Table 3.***

*A General Overview of the Characteristics of the Studies Included in the Review.*

| Study Authors And Year | Design | Country | Setting | Mental Health Condition | Mode of Treatment | Maximum number of Sessions | *N* | % Female | Mean Age (*SD*) |
|---|---|---|---|---|---|---|---|---|---|
| Bisby et al. (2022) | Secondary analysis of RCT | Australia | Outpatient | GAD | Online CBT | 8 | 259 | 75 | 44.84 (11) |
| Bisby et al. (2022) | Secondary analysis of RCT | Australia | Outpatient | Panic Disorder | Online CBT | 8 | 109 | 83 | 43.23 (11.04) |
| Bisby et al. (2022) | Secondary analysis of RCT | Australia | Outpatient | SAD | Online CBT | 8 | 175 | 59 | 42.95 (10.24) |
| Bisby et al. (2022) | Secondary analysis of RCT | Australia | Outpatient | MDD | Online CBT | 8 | 209 | 74 | 44.95 (11.9) |
| Bjureberg et al. (2020) | Secondary analysis of RCT | Sweden | Outpatient | Body Dysmorphic Disorder | Online CBT | 12 | 47 | 82.67 | 33.67 (13.33) |

| Study Authors And Year | Design | Country | Setting | Mental Health Condition | Mode of Treatment | Maximum number of Sessions | *N* | % Female | Mean Age (*SD*) |
|---|---|---|---|---|---|---|---|---|---|
| Buchholz et al. (2019) | Secondary analysis of RCT | USA | Outpatient | OCD | EPR | 16 | 44 | 63.64 | 27.19 |
| Clerkin et al. (2008) | Cohort study | USA | Outpatient | Panic Disorder | Group CBT | 12 | 30 | 70 | 40.63 (14.93) |
| Collins & Coles (2017) | Cohort study | USA | Outpatient | OCD | ERP | 16 | 23 | 51.85 | 32.3 (13.8) |
| D'Arcy & Norton (2021) | Cohort study | USA | Outpatient | GAD | tCBT | 12 | 58 | 48 | 32.5 (10.56) |
| Greenfield et al. (2011) | Cohort study | USA | Outpatient | Diverse Disorders | CBT | 30 | 106 | 64.2 | 34.5 (12.6) |
| Hamdeh et al. (2019) | Secondary analysis of RCT | Sweden | Outpatient | OCD | Online CBT | 12 | 170 | 77.34 | 34.4 (11.43) |
| Hardy et al. (2005) | Cohort study | UK | Outpatient | MDD | CT | 20 | 76 | 70 | 35.18 (9.86) |
| Hunnicutt-Ferguson et al. (2012) | Cohort study | USA | Outpatient | MDD | BA | 16 | 42 | 69.05 | 35.38 |

| Study Authors And Year | Design | Country | Setting | Mental Health Condition | Mode of Treatment | Maximum number of Sessions | *N* | % Female | Mean Age (SD) |
|---|---|---|---|---|---|---|---|---|---|
| Kelly et al. (2005) | Cohort study | USA | Outpatient | MDD | Group CBT | 12 | 31 | 61.3 | 41.6 |
| Kelly et al. (2007) | Cohort study | USA | Outpatient | MDD | IPT | 12 | 185 | 100 | 37.7 (9.93) |
| Ryan (2018) | Cohort study | USA | Outpatient | MDD | CT | 16 | 41 | 56 | 37.2 (12.1) |
| Silverstone et al. (2023) | Secondary analysis of RCT | Canada | Outpatient | GAD | Group tCBT | 12 | 117 | 86.3 | 37.82 (12.19) |
| Singh et al. (2021) | Secondary analysis of RCT | Canada | Outpatient | MDD | Experiential Therapy | 20 | 38 | 68.6 | 39.2 (10.8) |
| Thorisdottir et al. (2018) | Secondary analysis of RCT | Iceland | Outpatient | SAD | Group CBT, Group Psychotherapy | 8 | 45 | 46.67 | 19.82 (1.57) |
| Vincent & Norton (2019) | Secondary analysis of CCT | Australia | Outpatient | GAD | tCBT | 12 | 58 | 48.3 | 32.5 (10.56) |

Note. RCT = Randomised Control Trial, GAD = Generalised Anxiety Disorder, CBT = Cognitive Behavioural Therapy, SAD = Social Anxiety Disorder, MDD = Major Depressive Disorder, OCD = Obsessive Compulsive Disorder, EPR = Exposure and Response

Prevention, tCBT = Transdiagnostic CBT, CT = Cognitive Therapy, BA = Behavioural Activation, IPT = Interpersonal Therapy, CCT = Controlled Clinical Trial.

***Table 4.***

*An Overview of the Study Characteristics Relating to SGs and Treatment Outcome.*

| Study Authors and Year | SGs Criteria | SGs Measure | % of SGs | Medium Pre-Gain Session | Mean Magnitude of SGs | SG Reversals (%) | SG Association with Treatment Outcomes |
|---|---|---|---|---|---|---|---|
| Bisby et al. (2022) | Kelly | GAD-7 | 34 | 2 | | 24 | Yes. Participants who had a SG reported a greater reduction in symptoms compared to non-SGers across treatment ($p$ = .02). |
| Bisby et al. (2022) | Kelly | PDSS-SR | 37 | 2 | | 55 | No. No differences in treatment change were observed based on SG status ($p$ = .32). |
| Bisby et al. (2022) | Kelly | Mini-SPIN | 14 | 2 | | 42 | Yes. Participants who has a SG reported greater reduction in symptoms compared to non-SGers ($p$ = .002). |
| Bisby et al. (2022) | Kelly | PHQ-9 | 19 | 2 | | 13 | No. There were no difference in treatment outcomes based on SG status ($p$ = .8). |

| Study Authors and Year | SGs Criteria | SGs Measure | % of SGs | Medium Pre-Gain Session | Mean Magnitude of SGs | SG Reversals (%) | SG Association with Treatment Outcomes |
|---|---|---|---|---|---|---|---|
| Bjureberg et al. (2020) | Tang and Gaynor | Y-BOCS modified for BDD | 25.5 | 4 | 2.3 | 33.3 | Yes. Scores were significantly lower for the SG group post-treatment compared to non-SGers ($p < .001$) . |
| Buchholz et al. (2019) | Hardy | DOCS | 27.3 | 6 | 1.96 | | Yes. Participants who had a SG had greater symptom reduction than those without a SG ($F$ (1, 41) = 6.64, $p$ = .01). |
| Clerkin et al. (2008) | Kelly | PDSS-SR | 43.3 | 3 | | 33.3 | Yes. Individuals who had a SG had significantly lower scores at the end of treatment compared to non-SGers ($t$ (27) = 2.53 $p$ = .02). |
| Collins & Coles (2017) | Hardy | OCSCI | 52 | 11 | 0.96 | 21.4 | No. A significant time by SG status interaction was not found ($F$ (1, 20) = .05, $p$ = .83). |
| D'Arcy & Norton (2021) | Hardy | STAI-S | 19 | 5 | | | Yes. The interaction between time and SG status was significant ($F$ (1, 56) = 7.07, $p$ = .01). |

| Study Authors and Year | SGs Criteria | SGs Measure | % of SGs | Medium Pre-Gain Session | Mean Magnitude of SGs | SG Reversals (%) | SG Association with Treatment Outcomes |
|---|---|---|---|---|---|---|---|
| Greenfield et al. (2011) | Tang and Gaynor | OQ-45 | 23 | 4 | 1.36 | 36 | Yes. Significant time by SG status interaction was found ($F$ (1, 69) = 6.24, $p$ = .02). |
| Hamdeh et al. (2019) | Tang and Gaynor | Y-BOCS | 38 | 5 | 1.79 | 8 | Yes. Significant time by SG group interaction was found ($F$ (1, 128) = 34.7, $p$ < .001). |
| Hardy et al. (2005) | Hardy | BDI-II | 41 | 5 | 1.76 | 30 | Yes. Mean scores for SG group were significantly lower than non-SGers post-treatment ($t$ (74) = 3.62, $p$ = .001). |
| Hunnicutt-Ferguson et al. (2012) | Tang and Gaynor | QIDS- SR | 35.7 | 4 | 1.53 | 13.33 | Yes. The SG and non-SG groups differed in average change across treatment ($t$ (40) = 3.22, $p$ = .01). |
| Kelly et al. (2005) | Kelly | BDI-II | 41.9 | | | 53.85 | No. Participants who had a SG did not have better outcomes than non-SGers, ($F$ (1, 2) = 1.31, $p$ = .13). |
| Kelly et al. (2007) | Kelly | BDI-II | 33.5 | | 1.8 | 53 | No. Participants who had SGs did not differ significantly from non-SGers post-treatment, ($t$ (183) = .19, $p$ = .85). |

| Study Authors and Year | SGs Criteria | SGs Measure | % of SGs | Medium Pre-Gain Session | Mean Magnitude of SGs | SG Reversals (%) | SG Association with Treatment Outcomes |
|---|---|---|---|---|---|---|---|
| Ryan (2018) | Hardy | BDI-II | 58.54 | 4 | | 58 | No. Participants who had SGs had greater improvement at a trend level but no statistical difference ($t$ (39) = .6, $p$ = .55). |
| Silverstone et al. (2023) | Hardy | ADDQ-W | 18.8 | 6 | | | Yes. The main effect of SGs was significant ($F$ (1,81) = 7.18, $p$ = .01). |
| Singh et al. (2021) | Kelly | BDI-II | 63.9 | | 1.99 | 28.6 | Yes. Mean improvement was greater for SGers than non-SGers ($F$ (1, 34) = 8.39, $p$ < .001). |
| Thorisdottir et al. (2018) | Kelly | SIPS | 22.2 | 5 | | | No. SGs were not more likely to be associated with improved treatment outcome ($\chi^2$ (4) = 2.31, $p$ = .67). |
| Vincent & Norton (2019) | Hardy | STAI-S | 21 | 5 | 1.27 | | Yes. SGers had significantly greater improvement than non-SGers ($t$ (56) = −2.59, $p$ = .02). |

Note. GAD-7 = General Anxiety Disorder 7-Item, PDSS-SR = Panic Disorder Severity Scale–Self Report, Mini-SPIN = Mini-Social Phobia Inventory, PHQ-9 = Patient Health Questionnaire 9-Item, Y-BOCS = Yale-Brown Obsessive-Compulsive Scale, BDD = Body Dysmorphic Disorder, DOCS = Dimensional Obsessive-Compulsive Scale, OCSCI = Obsessive Compulsive Session Change Index, STAI-S = State-Trait Anxiety Inventory-State Version, OQ-45 = Outcome Questionnaire-45, BDI-II = Beck Depression Inventory-II, QIDS-SR = Quick Inventory of Depressive Symptomatology-Self Rated, ADDQ-W = Anxiety Disorders Diagnostic Questionnaire - Weekly, SIPS = Social Interaction Phobia Scale.

## Quality Appraisal

Appendix B presents an overview of the quality appraisal of the studies included in this review. Based on the quality appraisal, seven studies were rated as moderate and the remaining 10 studies were rated as weak.

All of the studies used validated outcomes measures and therefore they all scored strongly in this domain. The studies tended to use a strong or moderate research design (eight rated as strong and nine rated as moderate). The main area of bias was in participant selection (five rated moderate and 12 rated weak).

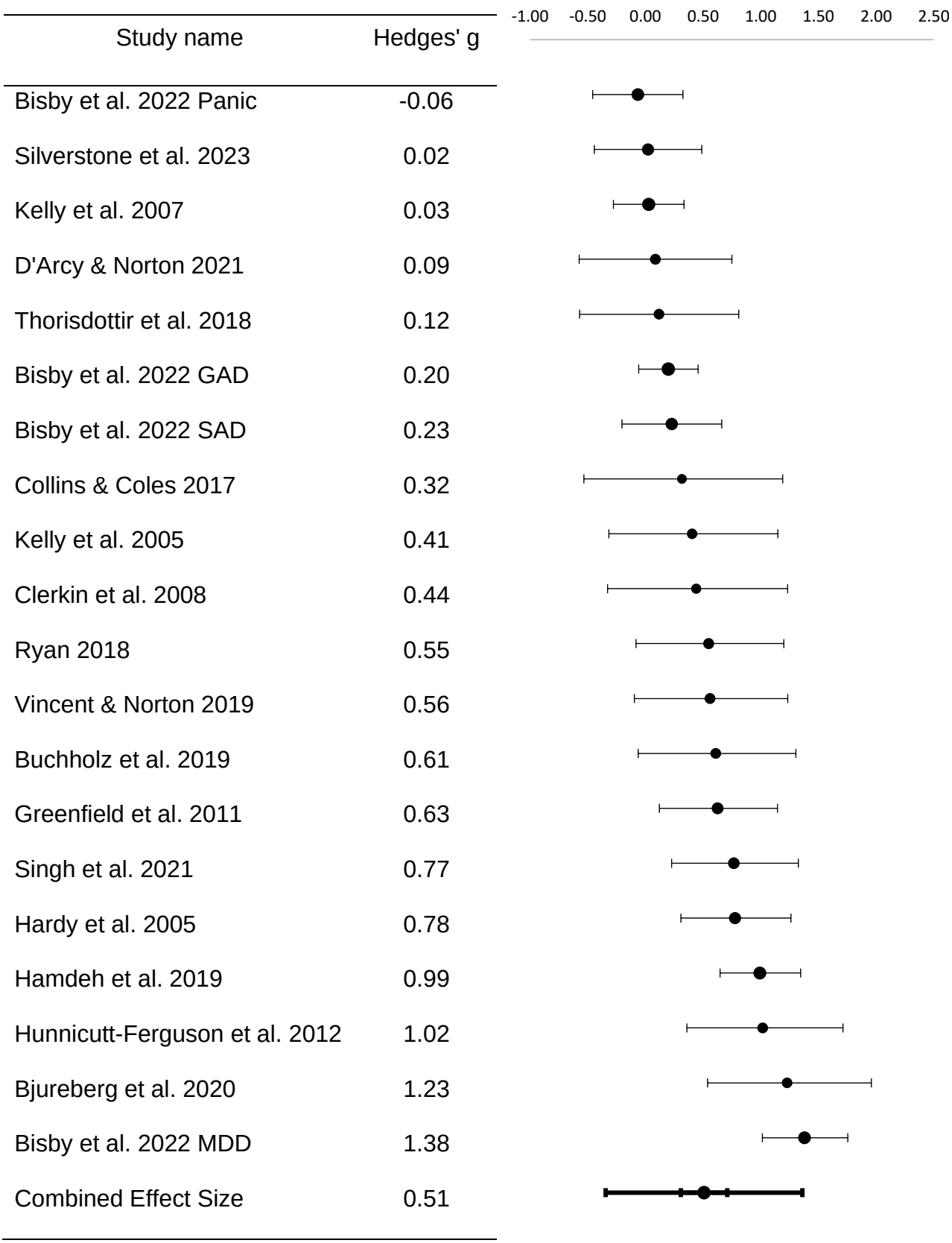
In general, it was found that the studies did not report on confounding variables, blinding methods, and withdrawals and dropout rates, and therefore the studies were rated poorly in these areas. This lack of reporting may be due to the use of secondary data analysis. Where possible, information from the original studies was used to support the quality appraisal.

## Meta-Analysis Results

A total of 17 studies ($k$ = 20), representing $N$ = 1848 participants, provided data related to treatment outcomes when comparing individuals who had a SG and individuals who did not. The Forest Plot showing the association between SGs and treatment outcomes is presented in Figure 2. The meta-analysis revealed a significant difference between SG and non-SG groups following treatment, $Z$ = 5.28, $p$ <.001, with an effect size of $g$ = .51 (*95% CI* [.31, .71]). Tests of heterogeneity were significant and showed substantial variability, $Q$ = 70.1, $p$ < .001; $I^2$ = 72.9%. Therefore, a moderator analysis was conducted to investigate the source of the heterogeneity, including exploring the SGs criteria used to identify SGs.

***Figure 2.***

*Forest Plot Showing Effect Sizes of the Studies Included in the Meta-Analysis.*



## Subgroup Analysis

### *Sudden Gains Criteria*

Of the 20 studies included in the meta-analysis, three different criteria were used to identify SGs: Kelly (*k* = 9); Hardy (*k* = 7); and Tang and Gaynor (*k* = 4). The subgroup analyses revealed that the effects gathered from studies using different SGs criteria were significantly different in SG outcomes: $Q_{between}$ (2) = 11.19, *p* = .004. The results from this analysis are presented in Figure 3 and Table 4. Studies which used the Tang and Gaynor criteria had an effect size of *g* = .94 (*95% CI* [.72, 1.15]) and no heterogeneity, *Q* = 2.31, *p* = .52, $I^2$ = 0%. Studies using the Hardy criteria had an effect size of *g* = .42 (*95% CI* [.18, .65]), and heterogeneity was non-significant and deemed unimportant, *Q* = 6.81, *p* = .34, $I^2$ = 11.9%. Studies using the Kelly criteria had an effect size of *g* = .39 (*95% CI* [.08, .69]), and large heterogeneity, *Q* = 43.25, *p* < .001, $I^2$ = 85.5%.

### *Diagnosis*

Of the 20 studies included in the meta-analysis, 18 studies focussed on three different diagnoses: anxiety disorders (*k* = 8); depressive disorders (*k* = 7) and OCD (*k* = 3). The subgroup analyses revealed that the effects gathered from studies focussing on diagnosis were significantly different in SG outcomes $Q_{between}$ (2) = 12.27, *p* = .002; see Figure 4 and Table 5. Studies which focussed on OCD had an effect size of *g* = .77 (95% CI [0.40, 1.15]) and non-significant heterogeneity *Q* = 2.8, *p* = .25, $I^2$ = 28.45%. Studies focussing on anxiety disorders had an effect size of *g* = .16, (*95% CI* [.04, .28]) and non-significant heterogeneity *Q* = 3.9, *p* = .79, $I^2$ = 0%. Studies focussing on depressive disorders had an effect size *g* = .70 (*95% CI* [.37, 1.04]. However, heterogeneity in this condition was significant *Q* = 33.35, *p* < .001, $I^2$ = 82.01%.

***Figure 3.***

*Forest Plot Showing the Effect Sizes in the SGs Criteria Subgroup Analysis*

| Study name / Subgroup name | Hedges' *g* |
|---|---|
| Buchholz et al. 2019 | 0.61 |
| Collins & Coles 2017 | 0.32 |
| D'Arcy & Norton 2021 | 0.09 |
| Hardy et al. 2005 | 0.78 |
| Ryan 2018 | 0.55 |
| Silverstone et al. 2023 | 0.02 |
| Vincent & Norton 2019 | 0.56 |
| **Hardy** | **0.42** |
| Bisby et al. 2022 GAD | 0.20 |
| Bisby et al. 2022 Panic | -0.06 |
| Bisby et al. 2022 SAD | 0.23 |
| Bisby et al. 2022 MDD | 1.38 |
| Clerkin et al. 2008 | 0.44 |
| Kelly et al. 2005 | 0.41 |
| Kelly et al. 2007 | 0.03 |
| Singh et al. 2021 | 0.77 |
| Thorisdottir et al. 2018 | 0.12 |
| **Kelly** | **0.39** |
| Bjureberg et al. 2020 | 1.23 |
| Greenfield et al. 2011 | 0.63 |
| Hamdeh et al. 2019 | 0.99 |
| Hunnicutt-Ferguson et al. 2012 | 1.02 |
| **Tang & Gaynor** | **0.94** |
| **Combined Effect Size** | **0.51** |

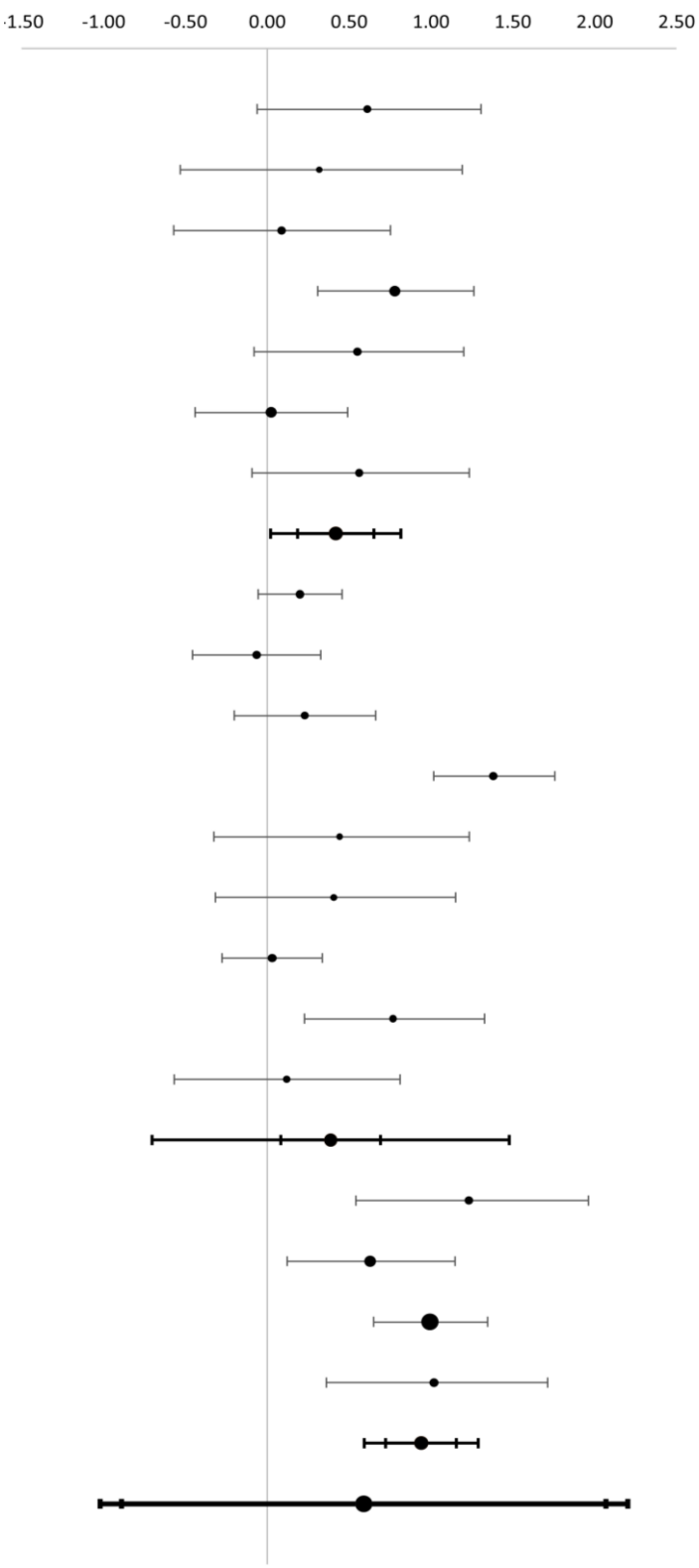


***Figure 4.***

*Forest Plot Showing the Effect Sizes in the Mental Health Condition Subgroup Analysis.*

| Study name / Subgroup name | Hedges' g |
|---|---|
| Bisby et al. 2022 GAD | 0.20 |
| Bisby et al. 2022 Panic | -0.06 |
| Bisby et al. 2022 SAD | 0.23 |
| Clerkin et al. 2008 | 0.44 |
| D'Arcy & Norton 2021 | 0.09 |
| Silverstone et al. 2023 | 0.02 |
| Thorisdottir et al. 2018 | 0.12 |
| Vincent & Norton 2019 | 0.56 |
| **Anxiety** | **0.16** |
| Bisby et al. 2022 MDD | 1.38 |
| Hardy et al. 2005 | 0.78 |
| Hunnicutt-Ferguson et al. 2012 | 1.02 |
| Kelly et al. 2005 | 0.41 |
| Kelly et al. 2007 | 0.03 |
| Ryan 2018 | 0.55 |
| Singh et al. 2021 | 0.77 |
| **Depression** | **0.70** |
| Buchholz et al. 2019 | 0.61 |
| Collins & Coles 2017 | 0.32 |
| Hamdeh et al. 2019 | 0.99 |
| **OCD** | **0.77** |
| **Combined Effect Size** | **0.51** |

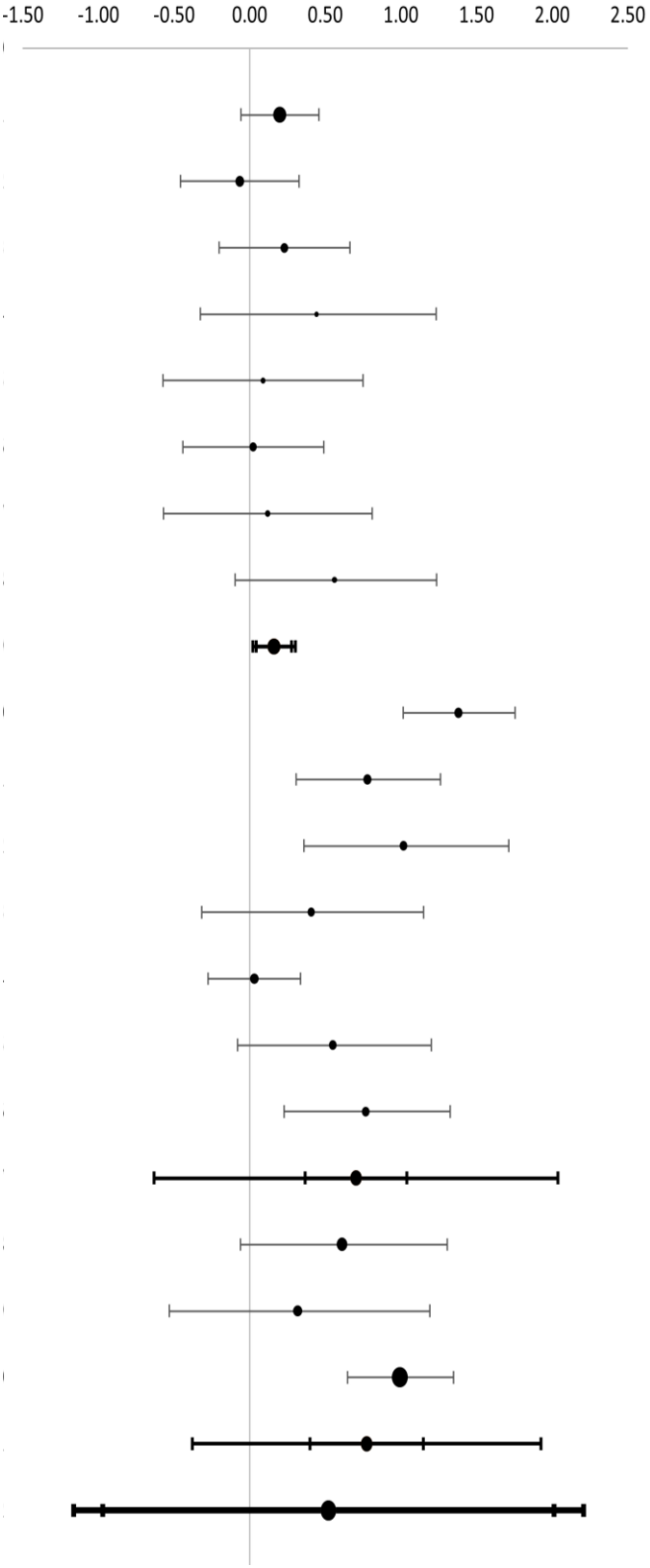


***Table 5.***

*A Table Presenting the Sub-Group Moderator Analyses.*

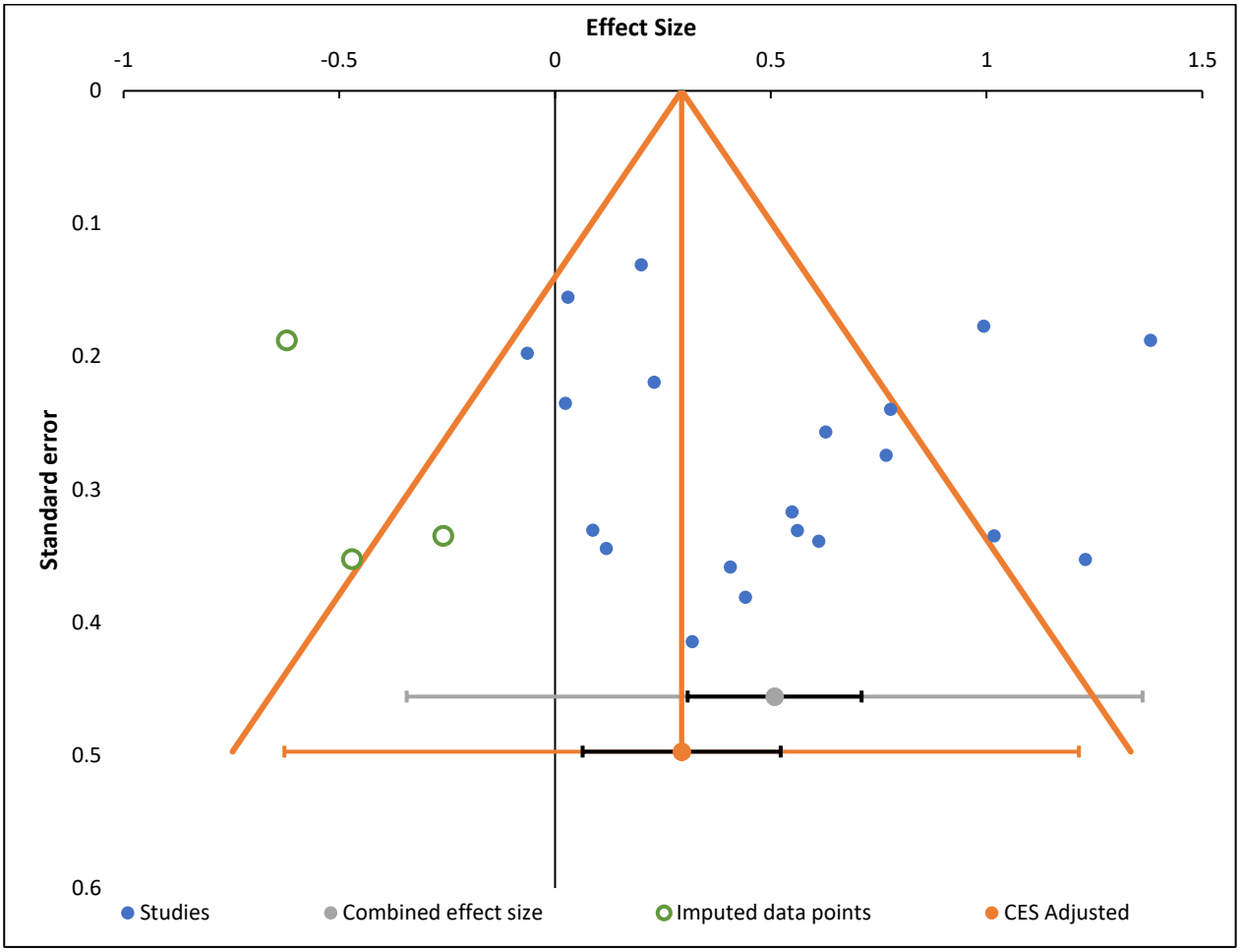
| Moderator | *k* | *n* | *g* | *95% CI* | *Q* | *p* | *I² (%)* |
|---|---|---|---|---|---|---|---|
| **Criteria** | | | | | | | |
| Kelly | 9 | 1079 | .39 | .08, .69 | 43.25 | <.001 | 81.5 |
| Hardy | 7 | 417 | .42 | .18, .65 | 6.81 | .34 | 11.9 |
| Tang & Gaynor | 4 | 365 | .94 | .72, 1.15 | 2.31 | .51 | 0 |
| **Mental Health Condition** | | | | | | | |
| Anxiety | 8 | 851 | .16 | .04, .28 | 3.9 | .79 | 0 |
| Depression | 7 | 620 | .70 | .37, 1.04 | 33.35 | <.001 | 82.01 |
| OCD | 3 | 237 | .77 | .40, 1.15 | 2.8 | .25 | 28.45 |

**Publication Bias**

To assess for publication bias, Rosenthal's fail-safe *N* analysis showed that 554 studies with a null result would be needed to reduce the significance of the effects to greater than .05. This is above the threshold of 295 studies suggested using the Rosenthal (1979) method. Egger's test was also found to be non-significant ($t$ = .36, $p$ = .72). However, a visual inspection of a funnel plot (Figure 5) showed asymmetry and the trim-and-fill test resulted in three studies being imputed. This suggests some publication bias, specifically of studies which do not find a significant difference between SG and non-SG groups. When the overall effect size is adjusted to account for this bias it appears to fall below $g$ = .50, reducing the effect size. However, this adjustment is speculation.

***Figure 5.***

*A Funnel Plot Assessing the Level of Publication Bias.*




## Discussion

The results from the current systematic review show that a large amount of research has been conducted in the area of sudden gains (SGs), and many different criteria have been used to identify SGs. This review included 17 studies which met the inclusion and exclusion criteria. Approximately 57 studies were identified which used the original SGs criteria. An additional 14 studies were identified which used adapted SGs criteria, however because the specific criteria used in these studies was applied to less than three studies they could not be included in this review. The extent of the adaptations

to the SGs criteria has been considerable and has reduced the current review's ability to compare the altered SGs criteria.

The findings from the meta-analysis indicated that there is a significant difference in outcomes between SG and non-SG groups following psychotherapy, with SGs being associated with improved outcomes (*g* = .51). However, substantial variability and heterogeneity were found in the results. This was expected due to the variability in the studies and indicated that further subgroup analysis was necessary.

The subgroup analysis between the SGs criteria was of primary importance to this investigation. It showed that studies using a combination of the Tang and Gaynor criteria yielded large effects whereas studies using the Hardy and Kelly criteria had smaller effects. One explanation for this result may be that the Tang and Gaynor criteria maintains the stringent nature of the original criteria whilst also being able to identify early SGs. Whereas the Hardy criteria is unable to identify early SGs, the Kelly criteria has been criticised for not being as stringent as the original criteria. This investigation found no heterogeneity between the studies using the Tang and Gaynor and the Hardy criteria. However, when identifying SGs using the Kelly criteria, a large amount of heterogeneity was found. One possible reason for this is that the sample size was much larger using the Kelly criteria, which would likely generate greater heterogeneity.

The subgroup analysis based on diagnosis found a significant difference in the association between SGs and treatment outcome depending on the diagnosis being treated. Studies focussing on OCD found a medium effect size and no heterogeneity. However, it must be noted that this sample was comparatively small. Studies focussing on anxiety found a small effect size and no heterogeneity. This may be explained by the fact that many of the studies which focussed on anxiety appeared to find no or minimal differences between the SG and non-SG conditions. Studies focussing on depression found a medium effect size but also high levels of heterogeneity. This may be because of

general variability in the sample, or specific factors that that were not included in the sungroup analysis, such as treatment modality and research design. It must be noted that two studies were unable to be included in this analysis as they investigated diverse disorders and body dysmorphic disorder which did not fit into the subgroups.

A subgroup analysis focussing on the different models of therapy was planned but it was not possible to conduct this analysis as the majority of studies used a CBT-based treatments. However, the importance of this subgroup analysis is reduced as there is an increasing amount of evidence to suggest that SGs occur across all psychological therapies and do not differ depending on the model of therapy used (Aderka & Shalom, 2021; Brockmeyer et al., 2023). Despite this, when considering which SGs criteria to apply, it is important to consider the evidence base around the timing of change in individual therapies (e.g., Lutz et al., 2014). For example, some therapy modalities have found that change can occur very early in therapy and therefore using SGs criteria which can be applied following the first session may be important.

When comparing the current results with the Shalom and Aderka (2020) meta-analysis, categorising the altered SGs criteria into three subgroups rather than comparing original to altered SGs criteria allowed more in-depth exploration of the specific criteria used to identify SGs and the following association between SGs and treatment outcome. When comparing criteria, the Tang and Gaynor criteria yielded a large effect, whereas the Hardy and Kelly criteria both had smaller effects. This is contradictory to the Shalom and Aderka (2020) findings. One explanation for the difference between the criteria could be that the Tang and Gaynor criteria has the strictness of the original criteria and therefore is more likely to identify SGs that are significantly related to improved outcomes, whilst also allowing early SGs to be identified. In contrast, the Kelly and Hardy criteria are less stringent and therefore the SGs identified using these criteria may be less likely to be significantly associated with improved treatment outcomes.

Another difference with the Shalom & Aderka (2020) meta-analysis appears when considering diagnosis. The current review findings suggest a small effect when focussing on anxiety disorders, whereas Shalom and Aderka (2020) found a medium effect. This may be related to their larger sample size. A recent meta-analysis comparing the number of SGs in anxiety and depressive disorders (Silverstone et al., 2023) found that depressive disorders have significantly more SGs than anxiety disorders and this may be another explanation for the smaller effects in anxiety disorders compared to depressive disorders and OCD found in the current review.

**Limitations**

Evidence of publication bias was found in this review, and one possible explanation for this is that studies which do not find a significant association between SG and treatment outcome are less likely to be published. This effects the validity of the results reported here and as stated above when the effect size is adjusted to account for this bias it becomes a small effect size.

The number of studies included in this review is a limitation because it reduces the power of the review. The decision to exclude the 14 studies that used altered criteria in less than three studies was made because the number of studies using the different criteria would be too small to make any meaningful inferences from. However, excluding these studies greatly reduced the sample. It may have been beneficial to compare the altered SGs criteria with the studies identified using the original criteria. However, this was out of the scope of the current review.

The quality of the studies included in the current review ranged from ‘moderate’ to ‘weak’ with no studies rated as ‘strong’. This suggests that research into SGs is generally low in quality and requires improvement. However, it is also important to note that the EPHPP tool may not have been the most suitable for evaluating the included studies. A quality appraisal tool which focuses on secondary analysis of data may have been more

appropriate for assessing the quality of the studies in the current review. For example, the EPHPP tool places an emphasis on blinding which is not relevant to SGs research. In fact, it may be beneficial when the occurrence of a SGs is shared with the patient. It also places emphasis on selection bias, and as the samples used are not generalisable to the general population, the included studies generally scored low. However, the samples used are appropriate for the research purposes as they often comprise the people who are seeking therapy, which makes the finding applicable to this specific population.

A final criticism of this review is the choice to look at how SGs relate to treatment outcomes, when other options would have been to look at the prevalence or magnitude of SGs. This decision was made based on the fundamental link between SGs and treatment outcomes. This association is the reason why this is such a large area of research and understanding this link is the driving force behind the literature. Looking at the prevalence and magnitude of SGs is interesting, especially as prevalence appears to change based on diagnosis, however the link with improved treatment outcomes remains more significant.

**Research and Clinical Implications**

A strength of the current review is its careful exploration of the criteria used to identify SGs. This is the first review to systematically review, document and tabulate the changes that have been made to the SGs criteria. The aim of this tool is to aid researchers with their decision-making process when choosing which SGs criteria to use, to support research with continued exploration to find a standardised SGs criteria, and to improve our understanding of how the SGs criteria have been altered and why. The table also indicates the extent of the alterations that are occurring, which may be having a detrimental impact on the validity of SGs research. As suggested by Shalom and Aderka (2020), research should continue to explore the differences between the original and altered SGs criteria to

continue developing our understanding of using altered criteria in identification of SGs and the association between SGs and improved treatment outcomes.

Clinicians should aim to administer sessional outcome measures with the intention of looking for large improvements between sessions that may indicate that a SG has occurred. As this review has found, SGs in treatment significantly improve treatment outcomes and it is important for clinicians to be aware of such an impact. It will be beneficial for clinicians to familiarise themselves with the SGs criteria, especially the first two criteria as they would be simple and easy to calculate in a clinical environment. Clinicians should also familiarise themselves with the 'upward spiral' (Aderka & Shalom, 2021) following a SG and aim to include these aspects into therapy following a potential SG.

Clinicians should be aware that SGs are less likely to occur in treatment for anxiety disorders but that they do occur and should still be monitored for. Whereas SGs in the treatment of other conditions appear to occur more frequently and can occur more than once throughout the course of therapy. Clinicians should be trying to identify SGs in all models of therapy.

The identification of SGs appears to be primarily a research focus, and while it is known that SGs are occurring in clinical practice, they are not well understood by clinicians. The routine monitoring and identification of potential SGs in clinical practice should be the aim of the next stage of SGs research and this is important to consider when thinking about SGs criteria. SGs can be easily identified in a research setting but going forward it is important for SGs to be easily identified in a clinical setting, to make this known phenomenon a useful clinical tool. Therefore, it may be helpful for future SGs research to continue developing an understanding of the criteria used to identify SGs, but also how these criteria can be applied clinically to support the phenomenon of SGs to become more clinically meaningful.

## Conclusion

In conclusion, a number of criteria have been used to identify SGs and the extent of the adaptations is potentially reducing the validity of SGs research. The current meta-analysis added to existing literature by finding that individuals who experience SGs have improved treatment outcomes compared to those who do not experience a SG. Three adapted SGs criteria were compared, and the Tang and Gaynor criteria was found to have a large effect, whilst the Kelly and Hardy criteria were found to have smaller effects. The differences between the criteria needs to be understood further, but a strength of the Tang and Gaynor criteria may be its strictness combined with its ability to identify early SGs, which creates a good association between SGs and improved treatment outcomes. Moving forward, it is important for SGs research to become more clinically applicable, and when considering the SGs criteria it may be helpful to consider how the criteria can be applied in a clinical setting to support treatment outcomes.

## References


Abel, A., Hayes, A. M., Henley, W., & Kuyken, W. (2016). Sudden gains in cognitive-behavior therapy for treatment-resistant depression: Processes of change. *Journal of Consulting and Clinical Psychology, 84*(8), 726-737. https://doi.org/10.1037/ccp0000101

Aderka, I. M., Appelbaum-Namdar, E., Shafran, N., & Gilboa-Schechtman, E. (2011). Sudden gains in prolonged exposure for children and adolescents with posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 79*(4), 441-446. https://doi.org/10.1037/a0024112

Aderka, I. M., & Shalom, J. G. (2021). A revised theory of sudden gains in psychological treatments. *Behavioural Research and Therapy, 139*, 103830. https://doi.org/10.1016/j.brat.2021.103830

Amir-Behghadami, M., & Janati A. (2020). Population, intervention, comparison, outcomes and study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. *Emergency Medicine Journal, 37*, 387. http://dx.doi.org/10.1136/emermed-2020-209567

Armijo-Olivo, S., Stiles, C. R., Hagen, N. A., Biondo, P. D., & Cummings, G. G. (2012). Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *Journal of Evaluation in Clinical Practice, 18*, 12-18. https://doi: 10.1111/j.1365-2753.2010.01516.x

Beck, A. T., & Steer, R. A. (1987). *Beck Depression Inventory manual*. Harcourt Brace.

Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck Depression Inventory–II (BDI-II) [Database record]. *APA PsycTests*. https://doi.org/10.1037/t00742-000

Bisby, M. A., Scott, A. J., Hathway, T., Dudeney, J., Fisher, A., Gandy, M., Heriseanu, A. I., Karin, E., Titov, N., & Dear, B. F. (2022). Sudden gains in therapist-guided versus self-guided online treatments for anxiety or depression. *Journal of Consulting and Clinical Psychology, 90*(11), 861–871. https://doi.org/10.1037/ccp0000771

Bjureberg, J., Enander, J., Andersson, E., Ivanov, V. Z., Rück, C., & Fernández de la Cruz, L. (2020). Sudden gains in internet-based cognitive behavior therapy for body dysmorphic disorder. *Behavior Therapy, 51*(5), 753–763. https://doi.org/10.1016/j.beth.2019.11.002

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

Brockmeyer, T., Titzmann, M., Zipfel, S., Wild, B., Resmark, G., Teufel, M., Giel, K., de Zwaan, M., Dinkel, A., Herpertz, S., Burgmer, M., Löwe, B., Tagay, S., Rothermund, E., Zeeck, A., Herzog, W., & Friederich, H. C. (2023). The role of general change mechanisms in sudden gains in the treatment of anorexia nervosa. *Behaviour Research and Therapy, 163*, 104285. https://doi.org/10.1016/j.brat.2023.104285

Buchholz, J. L., Abramowitz, J. S., Blakey, S. M., Reuman, L., & Twohig, M. P. (2019). Sudden gains: How important are they during exposure and response prevention for obsessive-compulsive disorder? *Behavior Therapy, 50*(3), 672–681. https://doi.org/10.1016/j.beth.2018.10.004

Card N. A. (2012). *Applied meta-analysis for social science research*. Guilford.

Cartwright, A., Cheng, Y. P., Schmidt, U., & Landau. S. (2017). Sudden gains in the outpatient treatment of anorexia nervosa: A process-outcome study. *International Journal of Eating Disorders, 50*(10), 1162-1171. https://doi:10.1002/eat.22773

Cavallini, A. Q., & Spangler, D. L. (2013). Sudden gains in cognitive behavioral therapy for eating disorders. *International Journal of Cognitive Therapy, 6*(3), 292–310. https://doi.org/10.1521/ijct.2013.6.3.292

Clerkin, E. M., Teachman, B. A., & Smith-Janik, S. B. (2008). Sudden gains in group cognitive-behavioral therapy for panic disorder. *Behaviour Research and Therapy, 46*(11), 1244–1250. https://doi.org/10.1016/j.brat.2008.08.002

Cohen, J. (1969). *Statistical power for the behavioural sciences*. Academic Press What Works Clearinghouse Procedures Handbook Version 4.0.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98–101. https://doi.org/10.1111/1467-8721.ep10768783

Collins, L. M. & Coles, M. E. (2017). Sudden gains in exposure therapy for obsessive-compulsive disorder. *Behaviour Research and Therapy. 93*, 1-5. https://10.1016/J.Brat.2017.03.003

D'Arcy, S., & Norton, P. (2021). The effect of comorbid depression on sudden gains during transdiagnostic cognitive-behavioural therapy for anxiety disorders. *Behaviour Change, 38*(2), 84-94. https://doi.10.1017/bec.2020.20

Doss, B. D., Rowe, L. S., Carhart, K., Madsen, J. W., & Georgia, E. J. (2011). Sudden gains in treatment-as-usual couple therapy for military veterans. *Behavior Therapy, 42*(3), 509–520. https://doi.org/10.1016/j.beth.2010.12.001

Durland, P. H., Wyszynski, C. M., & Chu, B. C. (2018). Predictors and outcomes of sudden gains and sudden regressions in cognitive behavioral therapy for youth anxiety. *Behavior Therapy, 49*(5), 823–835. https://doi.org/10.1016/j.beth.2018.01.006

Drymalski, W. M., & Washburn, J. J. (2011). Sudden gains in the treatment of depression in a partial hospitalization program. *Journal of Consulting and Clinical Psychology, 79*(3), 364–368. https://doi.org/10.1037/a0022973

Gaynor, S. T., Weersing, V. R., Kolko, D. J., Birmaher, B., Heo, J., & Brent, D. A. (2003). The prevalence and impact of large sudden improvements during adolescent therapy for depression: A comparison across cognitive-behavioral, family, and supportive therapy. *Journal of Consulting and Clinical Psychology, 71*(2), 386. https://doi.org/10.1037/0022-006X.71.2.386

Greenfield, M. F., Gunthert, K. C., & Haaga, D. A. F. (2011). Sudden gains versus gradual gains in a psychotherapy training clinic. *Journal of Clinical Psychology, 67*(1), 17–30. https://doi.org/10.1002/jclp.20748

Hamdeh, A. A., Bjureberg, J., Lenhard, F., Hedman-Lagerlöf, E., Flygare, O., Lundström, L., Ljótsson, B., Mataix-Cols, D., Rück, C. & Andersson, E. (2019). Sudden gains in internet-based cognitive behavior therapy for obsessive-compulsive disorder. *Journal of Obsessive-Compulsive and Related Disorders, 21*, 75–81. https://doi.org/10.1016/j.jocrd.2018.12.005

Hardy, G. E., Cahill, J., Stiles, W. B., Ispan, C., Macaskill, N., & Barkham, M. (2005). Sudden gains in cognitive therapy for depression: A replication and extension. *Journal of Consulting and Clinical Psychology, 73*(1), 59–67. https://doi.org/10.1037/ 0022-006x.73.1.59

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook.

Hopko, D. R., Robertson, S. M. C., & Carvalho, J. P. (2009). Sudden gains in depressed cancer patients treated with behavioral activation therapy. *Behavior Therapy, 40*(4), 346–356. https://doi.org/10.1016/j.beth.2008.09.001

Hunnicutt-Ferguson, K., Hoxha, D., & Gollan, J. (2012). Exploring sudden gains in behavioral activation therapy for major depressive disorder. *Behaviour Research and Therapy, 50*(3), 223–230. https://doi.org/10.1016/j.brat.2012.01.005

Ietsugu, T., Crane, C., Hackmann, A., Brennan, K., Gross, M., Crane, R. S., Silverton, S., Radford, S., Eames, C., Fennell, M. J. V., Williams, J. M. G., & Barnhofer, T. (2015). Gradually getting better: Trajectories of change in rumination and anxious worry in mindfulness-based cognitive therapy for prevention of relapse to recurrent depression. *Mindfulness, 6*(5), 1088–1094. https://doi.org/10.1007/s12671-014-0358-3

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19. https://doi.org/10.1037//0022-006x.59.1.12

Kelly, M. A. R., Cyranowski, J. M., & Frank, E. (2007). Sudden gains in interpersonal psychotherapy for depression. *Behaviour Research and Therapy, 45*(11), 2563–2572. https://doi.org/10.1016/j.brat.2007.07.007

Kelly, M. A. R., Roberts, J. E., & Ciesla, J. A. (2005). Sudden gains in cognitive behavioral treatment for depression: When do they occur and do they matter? *Behaviour Research and Therapy, 43*(6), 703–714. https://doi.org/10.1016/j.brat.2004.06.002

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174. https://doi.org/10.2307/2529310

Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., Woods, S. W. & Barlow, D. H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology, 82*(2), 287-297. https://doi.10.1037/a0035535

Malins, S., Moghaddam, N., Morriss, R., Schröder, T., Brown, P., & Boycott, N. (2021). Predicting outcomes and sudden gains from initial in-session interactions during remote cognitive-behavioural therapy for severe health anxiety. *Clinical Psychology and Psychotherapy*, *28*(4), 891–906. https://doi.org/10.1002/cpp.2543

Masterson, C., Ekers, D., Gilbody, S., Richards, D., Toner-Clewes, B., & McMillan, D. (2014). Sudden gains in behavioural activation for depression. *Behaviour Research and Therapy*, *60*, 34–38. https://doi.org/10.1016/j.brat.2014.06.008

Morrison, O. P. (2020). *What clients say about their single largest change in treatment: Comparing sudden gains across treatment approaches and phases of therapy.* [Doctoral Thesis, University of Windsor]. ProQuest. https://www.proquest.com/openview/e4d7035de7052386698209f7aaa53631/1?pq-origsite=gscholar&cbl=18750&diss=y

Page, M.J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T.C., Mulrow, C. D., Shamseer, L., Tetzlaff, J M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal, 372*. https://doi.10.1136/bmj.n71

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Ryan, E. T. (2018). *Sudden gains and sudden losses in cognitive therapy for major depressive disorder.* [Doctoral Thesis, Ohio State University]. Ohiolink.edu. https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=osu1354636220

Shalom, J. G., & Aderka, I. M. (2020). A meta-analysis of sudden gains in psychotherapy: Outcome and moderators. *Clinical Psychology Review, 76*, 101827. https://doi.org/10.1016/j.cpr.2020.101827

Silverstone, U. R., Roberge, P., Provencher, M. D., & Norton, P. J. (2023). An examination of sudden gain prevalence across cognitive-behavioural therapy for anxiety and depressive disorders: A quantitative analysis and meta-analytic review. *Journal of Anxiety Disorders, 95*, 102697. https://doi.org/10.1016/j.janxdis.2023.102697

Singh, T., Pascual-Leone, A., Morrison, O. P., & Greenberg, L. (2021). Working with emotion predicts sudden gains during experiential therapy for depression. *Psychotherapy Research*, *31*(7), 895-908. https://doi.org/10.1080/10503307.2020.1866784

Singla, D. R., Hollon, S. D., Fairburn, C. G., Dimidjian, S., & Patel, V. (2019). The roles of early response and sudden gains on depression outcomes: Findings from a randomized controlled trial of behavioral activation in Goa, India. *Clinical Psychological Science, 7*(4), 768–777. https://doi.org/10.1177/2167702619825860

Sloan, D. M., Thompson-Hollands, J., Hayes, A. M., Lee, D. J., Alpert, E., & Marx, B. P. (2022). Sudden gains in two trauma-focused treatments for posttraumatic stress disorder. *Behavior Therapy*, *53*(2), 255-266. https://doi.org/10.1016/j.beth.2021.08.003

Stevinson, C., & Lawlor, D. A. (2004). Searching multiple databases for systematic reviews: added value or diminishing returns? *Complementary Therapies in Medicine, 12*(4), 228-232. https://doi.org/10.1016/j.ctim.2004.09.003

Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., & Hardy, G. E. (2003). Early sudden gains in psychotherapy under routine clinic conditions: Practice-based evidence. *Journal of Consulting and Clinical Psychology, 71*(1), 14-21. https://doi.org/10.1037/0022-006x.71.1.14.

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*(4), 537-553. https://doi.org/10.1002/jrsm.1260

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*(6), 894–904. https://doi.org/10.1037/0022-006X.67.6.894

Tang, T. Z., DeRubeis, R. J., Beberman, R., & Pham, T. (2005). Cognitive changes, critical sessions, and sudden gains in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 73*(1), 168. https://doi.org/10.1037/0022-006X.73.1.168

The EndNote Team (2013). *EndNote* (EndNote X9) [64 bit]. Calrivate.

Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing, 1*, 176-184. https://doi.10.1111/j.1524-475X.2004.04006.x

Thorisdottir, A. S., Tryggvadottir, A., Saevarsson, S. T., & Bjornsson, A. S. (2018). Brief report: Sudden gains in cognitive-behavioral group therapy and group psychotherapy for social anxiety disorder among college students. *Cognitive Behaviour Therapy*, *47*(6), 462-469. https://doi.org/10.1080/16506073.2018.1466909

Vincent, L., & Norton, P. J. (2019). Predictors and consequences of sudden gains in transdiagnostic cognitive-behavioural therapy for anxiety disorders. *Cognitive Behaviour Therapy*, *48*(4), 265-284. https://doi.org/10.1080/16506073.2018.1513557

Vittengl, J. R., Clark, L. A., & Jarrett, R. B. (2005). Validity of sudden gains in acute phase treatment of depression. *Journal of Consulting and Clinical Psychology, 7*(1), 173–182. https://doi.org/10.1037/0022-006x.73.1.173

Zilcha-Mano, S., Errazuriz, P., Yaffe-Herbst, L., German, R. E., & DeRubeis, R. J. (2019). Are there any robust predictors of “sudden gainers,” and how is sustained improvement in treatment outcome achieved following a gain? *Journal of Consulting and Clinical Psychology, 87*(6), 491–500. https://doi.org/10.1037/ccp0000401

# Appendices

## Appendix A- PRISMA checklist

| Section and Topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| **TITLE** | | | |
| Title | 1 | Identify the report as a systematic review. | 1 |
| **ABSTRACT** | | | |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist. | 1 |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | 7 |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | 8 |
| **METHODS** | | | |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | 9 |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | 8 |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | 8 |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | 9 |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | 9 |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | 10 |
| | 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | 10 |
| Study risk of bias assessment | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | 11 |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | 12 |
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | 10 |
| | 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | 12 |

| Section and Topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| | 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | 10 |
| | 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | 11, 12 |
| | 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | 12, 13 |
| | 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | N/A |
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | 13 |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | 13 |
| **RESULTS** | | | |
| Study selection | 16a | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram. | 14, 15 |
| | 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. | 15 |
| Study characteristics | 17 | Cite each included study and present its characteristics. | 14-23 |
| Risk of bias in studies | 18 | Present assessments of risk of bias for each included study. | 24 |
| Results of individual studies | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. | 20-13 |
| Results of syntheses | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | 24 |
| | 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. | 24 |
| | 20c | Present results of all investigations of possible causes of heterogeneity among study results. | 24-29 |
| | 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results. | N/A |
| Reporting biases | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed. | 29, 30 |
| Certainty of evidence | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | N/A |
| **DISCUSSION** | | | |
| Discussion | 23a | Provide a general interpretation of the results in the context of other evidence. | 30-33 |
| | 23b | Discuss any limitations of the evidence included in the review. | 33, 34 |
| | 23c | Discuss any limitations of the review processes used. | 33, 34 |
| | 23d | Discuss implications of the results for practice, policy, and future research. | 34,35 |
| **OTHER INFORMATION** | | | |

| Section and Topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| Registration and protocol | 24a | Provide registration information for the review, including register name and registration number, or state that the review was not registered. | Unpublished manuscript. |
| | 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared. | PROSPERO CRD42023409210. |
| | 24c | Describe and explain any amendments to information provided at registration or in the protocol. | N/A |
| Support | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | N/A |
| Competing interests | 26 | Declare any competing interests of review authors. | N/A |
| Availability of data, code and other materials | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review. | Non are publicly available. |

*From:* Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/

**Appendix B- Quality Appraisal.**

| Paper | Selection bias | Study design | Confounders | Blinding | Data collection method | Withdrawals and drop-outs | Global rating |
|---|---|---|---|---|---|---|---|
| Bisby et al. (2022) | 3 | 1 | 1 | 2 | 1 | 1 | 2 |
| Bjureberg et al. (2020) | 3 | 1 | 1 | 1 | 1 | 1 | 2 |
| Buchholz et al. (2019) | 3 | 1 | 1 | 2 | 1 | 2 | 2 |
| Clerkin et al. (2008) | 3 | 2 | 1 | 2 | 1 | 2 | 2 |
| Collins & Coles (2017) | 3 | 2 | 3 | 2 | 1 | 1 | 3 |
| D'Arcy & Norton (2021) | 3 | 2 | 3 | 2 | 1 | 3 | 3 |
| Greenfield et al. (2011) | 2 | 2 | 3 | 2 | 1 | 3 | 3 |
| Hamdeh et al. (2019) | 3 | 1 | 3 | 1 | 1 | 3 | 3 |
| Hardy et al. (2005) | 2 | 2 | 3 | 2 | 1 | 2 | 2 |
| Hunnicutt-Ferguson et al. (2012) | 3 | 2 | 1 | 2 | 1 | 3 | 3 |
| Kelly et al. (2005) | 2 | 2 | 3 | 2 | 1 | 3 | 3 |
| Kelly et al. (2007) | 3 | 2 | 1 | 2 | 1 | 3 | 3 |
| Ryan (2013) | 2 | 2 | 3 | 2 | 1 | 1 | 2 |
| Silverstone et al. (2023) | 2 | 1 | 3 | 2 | 1 | 3 | 3 |
| Singh et al. (2021) | 3 | 1 | 1 | 2 | 1 | 1 | 2 |
| Thorisdottir et al. (2018) | 3 | 1 | 1 | 2 | 1 | 3 | 3 |
| Vincent & Norton (2019) | 3 | 1 | 3 | 2 | 1 | 3 | 3 |

All papers

10- weak

7- moderate

0- strong

## Section Two: Empirical Study

# Sudden Gains and Intraindividual variability in Cognitive Behavioural Therapy and Person Centred Experiential Therapy for Depression


## Abstract

**Objectives:** The aims of the current research were to determine whether intraindividual variability in depression symptomatology during treatment predicted sudden gains (SGs), and whether SGs predicted improved treatment outcomes in two contrasting psychological therapies for depression. These aims were explored using two different SGs criteria to determine their differential impact in the identification of SGs.

**Design:** A secondary data-analysis design was adopted, using data from a recently published pragmatic, non-inferiority randomised controlled trial (PRaCTICED, Barkham et al., 2021).

**Methods:** SGs were identified using the original Tang and DeRubeis (1999) criteria (*n* = 208), and the criteria adapted by Kelly (2005; *n* = 246) which allows for sudden gains to be identified earlier. Participants completed a weekly PHQ-9 measure, which was used to identify SGs. A regression analysis identified predictors of SGs and treatment outcomes.

**Results:** Using the original criteria, SGs occurred in 30.8% of the total sample and in 43.9% when using the altered criteria. Intraindividual variability in depression symptomology was significantly associated with SGs, but only when using the altered SGs criteria. SGs were significantly associated with improved treatment

outcome, but only when identified using the original SGs criteria. No difference was found relating to SGs and the two treatment modalities.

**Conclusion:** These findings provide mixed support for the revised theory of SGs, and suggest that SGs occur across therapeutic modalities. The current research highlights the importance of creating standardised criteria to identify SGs.

**Keywords:** Sudden Gains, Intraindividual variability, Psychotherapy, Cognitive Behavioural Therapy, Person Centred Experiential Therapy

**Practitioner Points**

- Clinicians, at intake, can assess how much symptoms fluctuate in patients as this could be an early indicator that a SGs may be more likely to occur throughout therapy.
- Clinicians can use sessional outcome measures which are directly related to the condition being treated or goals of therapy. This will allow for changes in symptoms to be monitored including large, sudden reductions which could indicate a SG.
- If the clinician suspects that a SG has occurred, the therapist can support the patient in trying to achieve the improved treatment outcomes related to SGs by actively pursuing the ‘upward spiral’, described in the original and revised theory of SGs.
- These practitioner points can be applied across treatment and therapy modalities and in a range of diverse settings.

# Sudden Gains and Intraindividual variability in Cognitive Behavioural Therapy and Person Centred Experiential Therapy for Depression

## Introduction

The phenomenon of sudden gains (SGs) was first identified by Tang and DeRubeis (1999) who defined SGs as “a rapid reduction in symptoms that occurs between consecutive treatment sessions”. They developed a model to explain the phenomenon, which proposed that cognitive behavioural therapy (CBT) generates cognitive changes that lead to improved alliance between the therapist and patient, which in turn lead to additional cognitive changes and thereby facilitates rapid symptom improvement and overall superior outcomes. The significance of the phenomenon of SGs is highlighted by the findings of two meta-analyses which found that individuals who experience SGs have better post treatment and follow-up outcomes compared to individuals who did not experience SGs (Aderka et al., 2012; Shalom & Aderka, 2020).

Since their original investigation, SGs have been widely researched and supporting evidence has been found throughout the lifespan (Aderka et al., 2012), in a range of treatment settings (Drymalski & Washburn, 2011) and in a variety of mental health conditions (Aderka et al., 2011). Although the research area originated in CBT, SGs have been found across a range of treatment modalities (Bisby et al., 2022). In their recent meta-analysis, Shalom and Aderka (2020) compared CBT therapies with non-CBT therapies as a moderator of SGs. They found that the effects of SGs were not significantly different in CBT ($g$ = .72) compared to non-CBT interventions ($g$ = .57) although there was a trend towards larger effects in CBT.

These findings suggest that SGs may predict outcomes beyond the therapeutic modality used.

In the original study, Tang and DeRubeis (1999) found the prevalence of SGs to be approximately 40% in CBT for depression. However, the prevalence of SGs has varied across samples. For example, Bohn et al. (2013) investigated SGs in interpersonal therapy for social anxiety disorder and found a prevalence rate of 14.3%, whereas Gibby (2015) investigated SGs in CBT for post-traumatic stress disorder and found a prevalence rate of 62.2%.

In addition, evidence for the Tang and DeRubeis' (1999) model of SGs has been inconsistent. Some studies have found evidence of cognitive changes preceding a SG (Cavallini & Spangler, 2013), whereas other studies have found no evidence of cognitive changes prior to SGs (Bohn et al., 2013). Similarly, predictive factors for SGs have received mixed evidence. For example, Storch et al. (2019) found that ethnicity is a predictor for SGs, whereas Thorisdottir et al. (2018) found that it was not. This mixed evidence has been the same for age (Hamdeh et al., 2019; Storch et al., 2019), pre-treatment symptom levels (Hamdeh et al., 2019; Hofmann et al., 2006), and a variety of other variables (Hardy et al., 2005; Kelly et al., 2007).

The processes following a SG have received less research attention and some evidence has, in part, supported the Tang and DeRubeis (1999) model. For example, in a large sample, Zilcha-Mano et al. (2019) found that SGs were mediated by increased alliance. Whilst Bohn et al. (2013) found no evidence of cognitive changes prior to the SG, they did find significant cognitive changes following the SG.

### Aderka and Shalom's (2021) Revised Theory of Sudden Gains

Given the conflicting evidence around the causes of SGs, much of which challenges the Tang and DeRubeis (1999) model, Aderka and Shalom (2021) proposed a revised model of SGs. They proposed that SGs are caused by natural symptom fluctuation (referred to as intraindividual variability in their model) rather than as the result of treatment factors. They suggest that during treatment, natural symptom fluctuations occur around a reducing mean and interacts with treatment. This causes a rapid decrease in symptoms which is reinforced by treatment factors.

The revised model is supported by evidence that SGs occur in individuals receiving no treatment (Krüger et al., 2014) and those receiving a placebo treatment (Vittengl et al., 2005). During the development of the revised model, Shalom et al. (2018) analysed three independent data sets to investigate whether intraindividual variability predicted SGs. The data included a randomised controlled trial (RCT) of prolonged exposure therapy for posttraumatic stress disorder in children and adolescents, an RCT of cognitive and behavioural therapies for obsessive compulsive disorder in adults, and psychodynamic therapy delivered under routine clinical conditions in a naturalistic setting for diverse disorders. They found that higher intraindividual variability in symptoms significantly predicted SGs in all three data sets. This indicated that SGs are significantly associated with natural variability in symptoms, in diverse modes of therapy, contexts, and populations. In further research, Shalom et al. (2020) investigated internet-delivered treatment for social anxiety disorder and found that higher intraindividual variability in symptoms significantly predicted SGs when variability was measured before treatment and during treatment.

Kuck et al. (2023) is, to the researcher's knowledge, the first independent published study to attempt to replicate these findings. They explored intraindividual variability and SGs in eye-movement desensitisation and reprocessing and imagery rescripting for PTSD and found that intraindividual variability was not associated with the occurrence of sudden gains in either treatment modality. They suggest that the findings by Shalom et al. (2018) do not easily generalise across treatments, disorders and populations. However, more independent investigation is needed to explore the relationship between intraindividual variability and SGs and provide further empirical evidence for or against Aderka and Shalom's (2021) revised model of SGs.

**Criteria for Identifying Sudden Gains**

One of the most significant criticisms of the SGs literature is related to the criteria used to identify SGs. The original definition by Tang and DeRubeis (1999) used three criteria: the reduction is large in absolute terms; it represents a 25% drop in symptoms, and it is stable. They themselves described elements of this definition as arbitrary and although it has been used in much of the SGs research, the definition has also been adapted. In particular, the third criterion has often been adapted as it excludes SGs which could occur between the first and second therapy sessions (Kelly et al., 2005).

The SGs criteria has been adapted in a variety of ways. One of the most used adaptations is from Kelly et al. (2005), who maintained the first two criteria whilst modifying the third criteria to 'an improvement of at least 1.5 standard deviations from the individual mean'. This allowed the research to identify early gains in therapy, whilst still including a criterion that would look at the stability of the SG. The ability to identify early SGs is particularly important as evidence has shown that SGs

are most likely to occur in the first three sessions (Dour et al., 2013, Keinonen et al., 2018; Masterson et al., 2014).

The meta-analysis conducted by Shalom and Aderka (2020) found that studies using the adapted SGs criteria had larger effect sizes compared to studies using the original criteria. Shalom and Aderka (2020) suggested that future research should investigate SGs using both the original criteria and the adapted criteria to develop a fuller understanding of what the different criteria are observing, with the intention of an empirically based standard criteria being developed.

**Research Aims**

The primary aim:

1. To determine whether intraindividual variability of depression symptomatology during treatment predicted SGs. As intraindividual variability appears to be more prevalent in depressive disorders (Shalom & Aderka, 2020), this study used data from a trial investigating the efficacy of contrasting psychological interventions for depression (Barkham et al., 2021). The presence of other predictive factors for SGs were also investigated, for example treatment modality. Based on the revised model of SGs (Aderka & Shalom, 2021), it was predicted that intraindividual variability would predict SGs using both SGs criteria.

The secondary aims:

2. To determine whether SGs predicted improved treatment outcomes in two different psychological therapies for depression. It was predicted that individuals who experience a SG will achieve better treatment outcomes

compared to those who do not experience a SG. It was predicted that this will be the same in both treatment modalities and using both SGs criteria.

3. To compare two different SGs criteria: the original criteria (Tang & DeRubis, 1999) and the altered criteria as defined by Kelly et al. (2005) and determine differences between the criteria and consider the clinical implications of any differences. It was predicted that the altered criteria would identify a higher prevalence of SGs compared to the original criteria due to its ability to identify early gains.

# Method

## Design

A quantitative secondary data analysis design was used. The data has been drawn from the PRaCTICED trial (Barkham et al., 2021), which adopted a pragmatic, randomised, non-inferiority design to investigate the comparative efficacy of CBT versus person-centred experiential therapy (PCET) for moderate to severe depression in the Sheffield Increasing Access to Psychological Therapies (IAPT) service (now named NHS Talking Therapies for Anxiety and Depression).

## Approvals

National Health Service (NHS) ethical approval was granted for the PRaCTICED trial (Barkham et al., 2021) by the Health Research Authority (Research Ethics Committee 14/YH/0001) and the research topic was specified in the protocol. Informed consent was gained from the participants. All the data has been anonymised to protect the identity of participants. No research safety issues were identified.

All the data has been stored on encrypted files within a shared drive, which was only accessible to the researchers. There was no need for data to be printed or stored in any written format. A research data management plan was created to ensure all data was safely managed. Ethical approval from the University of Sheffield was gained for this study (Reference Number 044245, Appendix A).

The current research project was conducted and reported in line with the Journal Article Reporting Standards (JARS; APA, 2008), specifically the General quantitative reporting standards (Appendix B).

**Service User Involvement**

The PRaCTICED trial was supported by a Patient and Public Involvement (PPI) group throughout the research process. As the PPI group has now disbanded additional input from this group is not possible.

**PRaCTICED Trial Overview**

The PRaCTICED trial took place in Sheffield, a city with a population of 584,280 (Sheffield City Council, 2019) which has average demographics in comparison to other cities across the UK (Saxon et al., 2017). Sheffield's IAPT service was set up by Sheffield Health and Social Care NHS Foundation Trust (SHSC) in 2009. The service offered both CBT and PCET prior to the trial.

The Sheffield IAPT service comprised approximately 30 counsellors and 35 high intensity CBT therapists. PCET is a form of person-centred therapy and in the IAPT service it was originally named 'counselling for depression' and subsequently re-named person-centred experiential counselling for depression (Murphy, 2019). The current study uses the term PCET, consistent with the PRaCTICED trial. All trial counsellors (*N* = 18) received PCET training prior to taking on trial patients. All CBT

therapists (*N* = 27) were trained in Beckian CBT for depression (Beck, 2011). Both PCET counsellors and CBT therapists received top-up training during the trial and received ongoing supervision throughout, consistent with the IAPT model of practice.

**Participant Procedure**

The participants in the PRaCTICED trial completed a full battery of measures at assessment. Acceptance into the trial was based on participants having moderate to severe depression as determined by the Clinical Interview Schedule-Revised (CIS-R; Lewis et al., 1992). Participant demographics and some clinical information was collected including: age; gender; a measure of deprivation (IMD); employment status; use of psychiatric medication; referral source; and ethnicity. The measures were repeated again at six and 12-months post randomisation.

Following the assessment, eligible participants were randomised to either CBT or PCET. Participants were offered weekly one-to-one therapy sessions and could have up to 20 sessions, although this was flexible depending on clinical need and standard IAPT practice for determining discharge. Participants completed the following weekly measures at the beginning of their therapy sessions: Generalised Anxiety Disorder- 7 item (Spitzer et al., 2006); Work and Social Adjustment Scale (Mundt et al, 2002); and Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2010). Data from the PHQ-9 is the only measure that was included in the current study given the primary focus on depression.

**Measures**

***Patient Health Questionnaire-9*** (PHQ-9; Kroenke et al., 2010)

The PHQ-9 (Appendix C) was the primary measure for determining the state of depression in the current study. It consists of nine questions and asks participants

to rate their experiences over the past two weeks on a four-point scale ranging from "not at all" to "nearly every day". Scores can range from 0-27 and scores of 5, 10, 15 and 20 represent the criteria for mild, moderate, moderately severe, and severe depression. The PHQ-9 has a Cronbach's $\alpha$ = .86, suggesting excellent internal reliability and the test-retest reliability has also been found to be excellent (Kroenke et al., 2001).

***Clinical Interview Schedule-Revised*** (CIS-R; Lewis et al., 1992)

The CIS-R was used to determine the initial level of depression severity as being either moderate or severe. A milder categorisation of depression was an exclusion criterion. The CIS-R is a computerised diagnostic tool which was created to improve the standardisation of diagnosis. It has been found to be a valid instrument for detection of common mental health disorders and has excellent specificity (.97) but lower sensitivity (.49; Subramaniam et al., 2006).

**Sudden Gains Criteria**

The original SGs criteria (Tang & DeRubeis, 1999) were used as follows:

1. The reduction is large in absolute terms. Tang and DeRubeis (1999) defined this as a change of seven points on the Beck Depression Inventory (BDI; Beck et al., 1987). However, the original authors described their use of seven points as "arbitrary" and this criterion was modified by Stiles et al. (2003) who recognised that 7 points on the BDI is nearly equivalent to its reliable change index (RCI; Jacobson & Truax, 1991) score. Using the RCI allows this criterion to be adapted and applied to measures other than the BDI, whilst maintaining this original criteria.
2. It represents a 25% drop in symptoms.

3. It is stable. Tang and DeRubis (1999) operationalised this by stating that the symptom levels in the three sessions following the gain are significantly lower than the symptom levels in the three sessions preceding the gain using an independent samples t-test. When the first two criteria were met but only two scores were available either pre or post gain this analysis could be conducted with five scores.

The altered criteria (Kelly et al., 2005) used the same first two criteria from the above definition. However, the third criterion was changed to:

3. A gain must reflect an improvement of at least 1.5 standard deviations from the individual mean.

The current research used the PHQ-9, as this is the standard measure of depression in many services. With regards to the first criteria, “the reduction is large in absolute terms”, a transformation was needed from the BDI to the PHQ-9. The RCI for the PHQ-9 was originally identified as a score of five points (McMillan et al., 2010). However, this has been adapted more recently to a score of six points, which is cited in the IAPT manual (NHS England, 2021) and used clinically. Additionally, SGs research using the PHQ-9 has also used an RCI score of six (Aderka et al., 2021).

## Participants

### *PRaCTICED Trial Participants*

During the PRaCTICED trial the participants were screened for eligibility and were required to give informed consent. Individuals who did not meet the criteria for the trial received treatment as usual within the Sheffield IAPT service. Participants were required to meet the following criteria:

- Aged 18 years or older
- A moderate to severe or severe rating of depression on the CIS-R
- No strongly-held treatment preference such that they would decline treatment if it were not their preferred treatment

Participants were excluded from the PRaCTICED trial if they met any of the following criteria:

- The presence of a long-term physical health condition
- A previous diagnosis of personality disorder, schizophrenia or bipolar disorder
- Drug or alcohol dependency
- An elevated clinical risk of suicide

A total of 761 participants were assessed between November 2014 and August 2018 and of these, 510 participants met the eligibility criteria and were randomly assigned to one of the two active treatment conditions.

***The Altered Sudden Gains Criteria Participants***

In order to meet the altered SGs criteria, further exclusion criteria were needed. The additional criteria excluded participants if they met any of the following:

- Attended fewer than three therapy sessions
- Had missing weekly PHQ-9 scores
- Switched between therapy modalities during treatment
- Did not progress to high intensity therapy

Using these additional exclusion criteria, a further 264 participants were excluded leaving a total of 246 participants to be analysed using the altered criteria. The participant allocation flow diagram is presented in Figure 1. The participants in this sample were aged between 18 and 69 years with an average age of 38.22 years

(*SD* = 12.38). The participants had between 3 and 23 therapy sessions with a mean of 10.54 (*SD* = 5.71). 53.7% of participants received PCET with the remainder receiving CBT. Full participant demographics are presented in Table 1.

***Figure 1.***

*Participant Allocation Flow Diagram Detailing the Reasons for Exclusion in Both SGs Criteria*

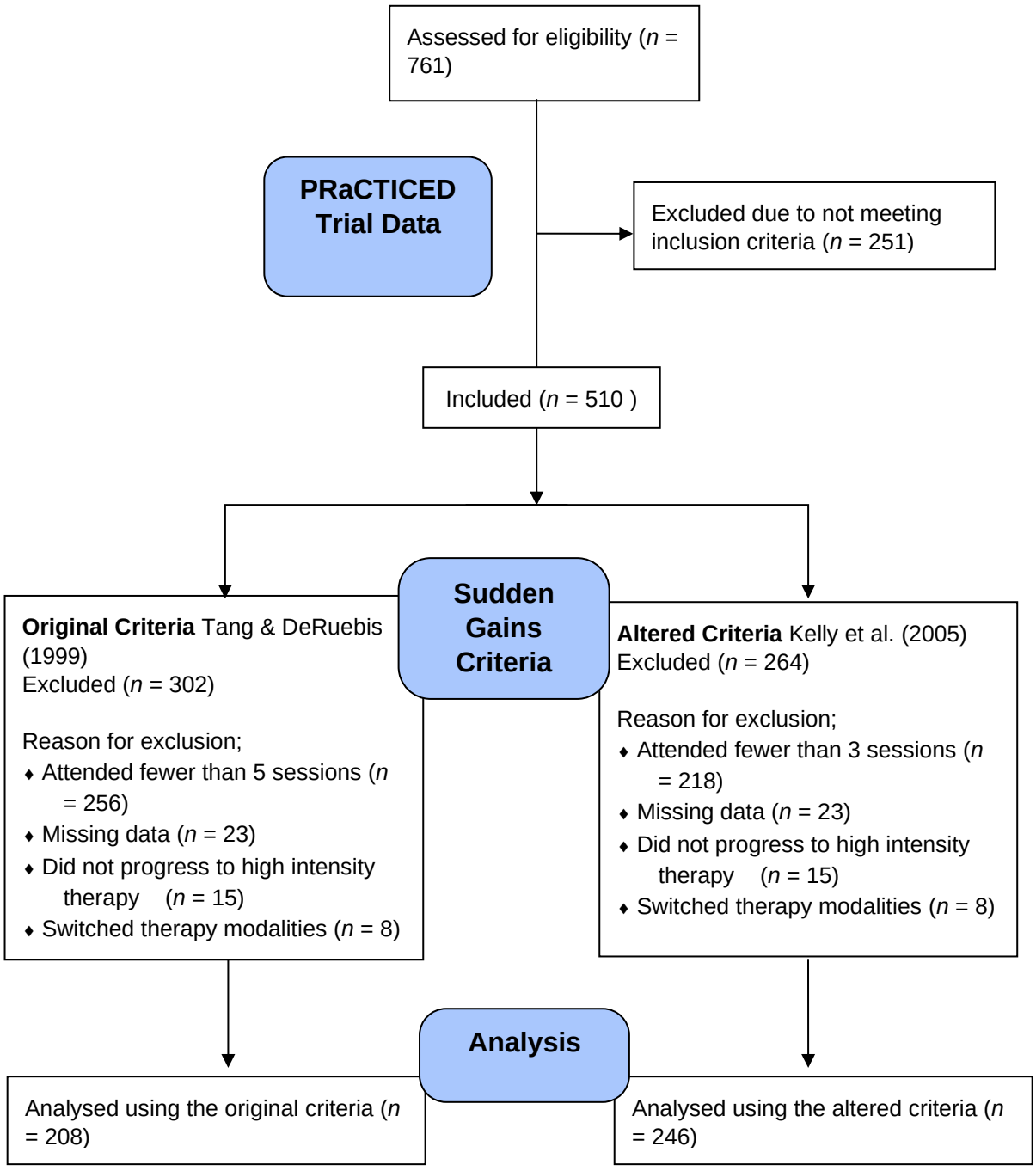


***Table 1.***

*Demographics for Participants in Both SG Criteria Samples*

| Variable | Original Criteria (*n* = 208) | Altered Criteria (*n* = 246) |
|---|---|---|
| **Gender** (%) | | |
| Male | 41.8 | 40.2 |
| Female | 58.2 | 59.8 |
| **Age** M (SD) | 38.46 (11.96) | 38.22 (12.38) |
| **Ethnicity** (%) | | |
| White British | 84.1 | 84.1 |
| Not Asked | 9.1 | 8.5 |
| Black/Black British Caribbean | 1.4 | 1.2 |
| Mixed White and Black Caribbean | 1.4 | 1.2 |
| White Irish | 1.0 | 0.8 |
| White Other | 0.5 | 0.8 |
| Asian/Asian British Indian | 0.5 | 0.4 |
| Asian Other | 0.5 | 0.4 |
| Black/Black British African | 0 | 0.4 |
| Other | 0 | 0.4 |
| Refused to Answer | 0 | 0.4 |
| Unable to Answer | 0.5 | 0.4 |
| **Employment Status** (%) | | |
| Employed | 55.8 | 56.3 |
| Seeking Work | 6.7 | 9.4 |
| Long-term Sick/Disabled | 9.1 | 7.8 |
| Full-time Student | 4.8 | 5.3 |
| Retired | 1.9 | 2.0 |
| Homemaker | 1.9 | 1.6 |
| Part-time Employed | 0.5 | 0.4 |
| Not seeking work | 0.5 | 0.4 |
| Unknown | 18.8 | 16.7 |
| **IMD** M (SD) | 5.5 (3.31) | 5.3 (3.3) |
| **Psychotropic Medication** | | |
| Prescribed and Taking | 58.2 | 56.5 |
| Not Prescribed | 29.8 | 30.5 |
| Unknown | 9.1 | 8.9 |
| Prescribed not Taking | 2.9 | 4.1 |
| **Referral Source** (%) | | |
| GP | 96.2 | 95.9 |
| Self-referral | 1.4 | 2.0 |
| Unknow | 2.4 | 2.0 |

Note. SG= Sudden Gain; CBT= Cognitive Behaviour Therapy; PCET= Person Centred Experiential Therapy; IMD= Indices of Multiple Deprivation

### *The Original Sudden Gains Criteria Participants.*

In order to meet the original SGs criteria one further exclusion criterion was needed:

- Attended less than five therapy sessions

Using this additional exclusion criterion, a further 38 participants were excluded, leaving a total of 208 participants to be analysed using the original criteria. The participants in this sample were aged between 19 and 65 years with an average age of 38.46 years (*SD* = 11.96). The participants had between 5 and 23 therapy sessions with a mean of 11.81 sessions (*SD* = 5.5). 56.3% of participants received PCET with the remainder receiving CBT.

## Analysis

### *Identifying Sudden Gains*

Identifying SGs was the first stage of analysis as this was central to all subsequent analysis. SGs were identified using the two criteria discussed previously. Part of this analysis was conducted in collaboration with peer CN, see collaboration statement (Appendix D). When all of the criteria were met, a SG had occurred. This analysis was conducted using the two different SGs criteria to create two independent data sets.

### *Intraindividual variability and Sudden Gains*

The primary research aim was analysed using multiple logistic regressions to determine whether intraindividual variability in depression symptomology predicted SGs using the different criteria. In this analysis SGs were the dependent variable and the independent variables were intraindividual variability scores and all other demographic and clinical variables (initial depression severity, age, gender, IMD,

employment status, use of psychiatric medication, referral source, ethnicity, therapist, number of treatment sessions and mode of therapy). To increase the robustness of the analyses, they were run with bootstrapping, and the confidence intervals and standard errors were based on 1000 bootstrap samples.

The method used to create a measure of intraindividual variability was replicated from Shalom et al. (2018). This consisted of calculating the change scores between consecutive sessions prior to the SG. These scores were converted to absolute scores as the direction of change was not relevant to the analysis. Then the average sum of the change scores was calculated to create an overall intraindividual variability score.

An intraindividual variability score was also created for participants who did not have a SG. Again, this was replicated from Shalom et al. (2018). The average number of sessions prior to the SG was identified for each SG criteria, then the consecutive change scores were used to calculate an intraindividual variability score for participants who did not have a SG using the same method as above. A minimum of three consecutive change scores were needed to calculate the intraindividual variability score.

### *Sudden Gains and Outcomes*

Multiple linear regression analyses were used to investigate whether SGs predicted improved treatment outcomes using the different SGs criteria. In this analysis the treatment outcome was defined as the final PHQ-9 score. The initial PHQ-9 score was included in all analyses and therefore controlled for the impact of initial severity on outcome. For this investigation, the treatment outcome was the dependent variable and SGs and other demographic and clinical factors were included as independent variables (intraindividual variability, age, gender, IMD,

employment status, use of psychiatric medication, referral source, ethnicity, therapist, number of treatments sessions, and mode of therapy). To increase the robustness of the analysis, they were run with bootstrapping, and the confidence intervals and standard errors are based on 1000 bootstrap samples.

***Power Analysis***

A post-hoc power analysis was conducted using G*Power 3.1.9.7 (Faul et al., 2007) for the logistic and linear regressions. This analysis was conducted post-hoc as it was secondary analysis of an existing dataset and therefore the sample size could not be altered. However, the post-hoc analysis would indicate whether the analyses were sufficiently powered to find a significant result.

For the logistic regression using the original criteria, a power value of 1-$\beta$ = .96 was found, suggesting a sufficient level of statistical power. For the logistic regression using the altered criteria, a power value of 1-$\beta$ = .29 was found. As this is lower than .8, it suggests that this analysis may have been insufficient powered to find a significant result if one exists (Cohen, 1992).

For the linear regression analysis using the original criteria, a power value of 1-$\beta$ = .99 was found suggesting a sufficient level of statistical power. For the linear regression using the altered criteria, a power value of 1-$\beta$ = .99 was found also suggesting a sufficient level of statistical power for this analysis.

***Statistical Assumptions***

Box plots were used to identify outliers (Appendix E). Using the original criteria, eight outliers were identified; one in the first session PHQ-9 scores, one in the PHQ-9 end of therapy scores, and four in the intraindividual variability scores. Using the altered criteria nine outliers were identified; one in the first session PHQ-9 scores, one in the PHQ-9 end of therapy scores, and seven in the intraindividual

variability scores. When comparing the two criteria, there was overlap in the outliers: the first session PHQ-9 score had the same outlier in both samples; the PHQ-9 end of therapy score had the same outlier in both samples; and two of the outliers for intraindividual variability scores were the same in both samples. As none of these outliers were extreme and none were repeatedly the same participants, it was decided not to remove them from the sample.

As proposed by George and Mallory (2010), the acceptable range for the data to be normally distributed for skewness and kurtosis is ±1.96. In both criteria all variables were within these acceptable limits (see Appendix E). Therefore it was deemed appropriate to proceed with parametric analysis (Glen, 2022).

For the regression analysis, the assumption of independence of residuals was met; the Durbin-Watson statistics were very close to a value of 2 (logistic regression: original criteria = 1.87; logistic regression altered criteria = 2. Linear regression: original criteria = 2.1; altered criteria = 2.03). The assumption of multicollinearity was met for all the regressions, with all Variance Inflation Factor statistics less than ten and tolerance values greater than .1 (Appendix F). For the logistic regression, linearity of the logit was assessed and no interaction terms were significant, suggesting that the assumption was met (Appendix G). Therefore, the data has been considered suitable for the regression analyses.

When the linear regression was initially conducted the same outlier in both criteria was identified as having significant influence on the model (original criteria: Mahalanobis Distance = 14.03; Cook's Distance = .29; Centred Leverage Value = .08. Altered criteria: Mahalanobis Distance = 15.75; Cook's Distance = .14; Centred Leverage Value = .06). Therefore, this outlier was removed and not included in the

linear regression analysis. The same investigation was conducted for the logistic regression and no participants were identified as having a significant influence on the model.

## Results

### Identifying Sudden Gains

Using the original criteria, 74 SGs occurred in 64 participants (30.8% of the sample), of whom 54 participants had one SG and 10 participants had two SGs. The mode number of SGs occurred between session three and four (16.22%). Sudden gains occurred in 45.05% of participants receiving CBT and 28.21% of participants receiving PCET. A 2x2 $\chi^2$ analysis was carried out to discover whether there was a significant relationship between the number of SGs in the two therapy modalities. The $\chi^2$ (1) = 2.29, $p$ = .13, suggesting no association between the number of SGs and therapy modality.

Using the altered criteria, 128 SGs occurred in 108 participants (43.9% of the sample), with 90 participants having one SG, 16 participants two SGs, and two participants having three SGs. The mode number of SGs occurred between session 1 and 2 (14.84%). SGs occurred in 42.98% of participants receiving CBT and 44.7% of participants receiving PCET. A 2x2 $\chi^2$ analysis was carried out to discover whether there was a significant relationship between the number of SGs in the two therapy modalities. The $\chi^2$ (1) = .07, $p$ = .79, suggesting no association between the number of SGs and therapy modality. Table 2 summarises the differences between the two criteria when identifying SGs. Figure 2 shows the number of SGs occurring in each treatment session for both SGs criteria and for CBT and PCET.

***Table 2***.

*The Differences Between the Two Criteria When Identifying SGs*

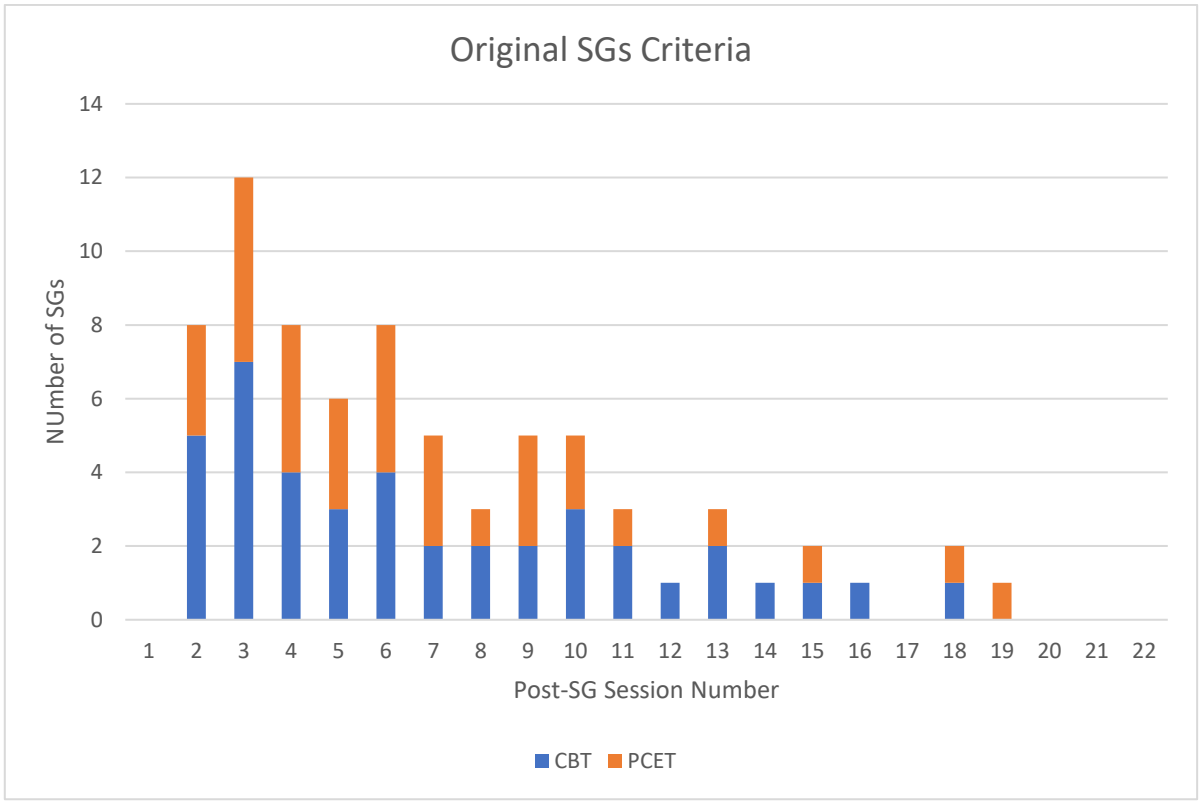
| | Original SGs Criteria | | | Altered SGs Criteria | | |
|---|---|---|---|---|---|---|
| | Overall | CBT | PCET | Overall | CBT | PCET |
| SG Frequency (%) | 35.58 | 45.05 | 28.21 | 43.9 | 42.89 | 44.7 |
| Medium Pre-Gain Session | 6 | 6 | 6 | 5 | 5 | 4 |
| SG Reversals (%) | 28.38 | 31.71 | 24.24 | 39.06 | 42.65 | 35 |
| SG Regains (%) | 71.43 | 76.92 | 62.5 | 47.74 | 34.48 | 52.38 |
| Mean Magnitude of SGs (SD) | 8.04 (2.53) | 8.23 (2.54) | 7.84 (2.37) | 8.51 (2.71) | 8.75 (2.71) | 8.31 (2.68) |

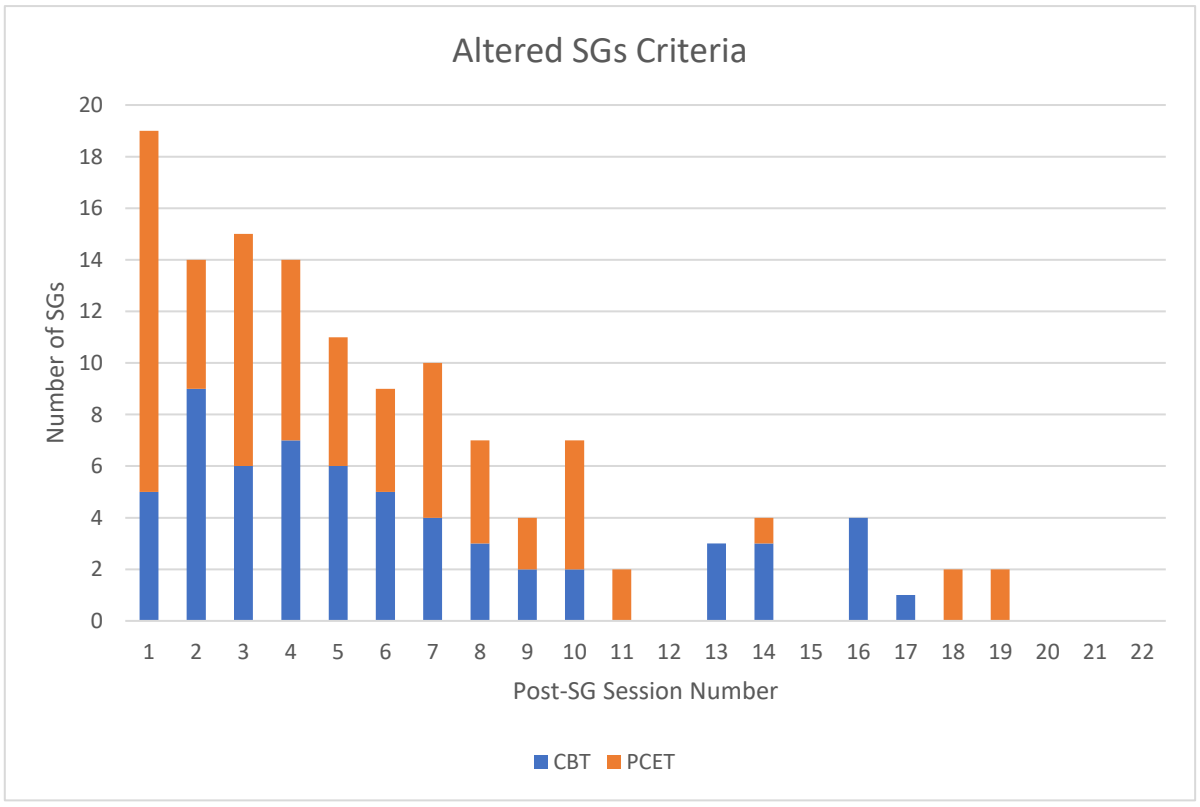
Note. SG = Sudden Gain; CBT = Cognitive Behaviour Therapy; PCET = Person Centred Experiential Therapy.

A 2x2 $\chi^2$ analysis was carried out to discover whether there was a significant relationship between the SG criteria. The $\chi^2$ (1) = 8.26, $p$ = .004, suggests that an association is extremely unlikely to have arisen as a result of sampling error. Cramer's $V$ = .14, suggesting that nearly 2% of the variance in the frequencies of SGs can be explained by the criteria used to identify the SG. Due to differences between the SGs criteria, all further analyses were conducted using both criteria to help identify how these differences may impact SGs research.

***Figure 2*.**

*The Number of SGs Occurring in Each Treatment Session for CBT and PCET Using Both SGs Criteria.*



## Intraindividual variability and Sudden Gains

### *Original Criteria*

Table 3 shows the logistic model of predictors of SGs using the original criteria. The final regression model was: $\chi^2(4) = 14.54$, $p = .006$. The model was able to correctly classify SGs 75% of the time, with non-SGs being classified more accurately than SGs (97.9% compared to 8.2%).

Using the original criteria, the initial PHQ-9 score was significantly associated with whether an individual had a SG. Specifically, individuals who had higher depression severity were more likely to have a SG. Employment status was also significantly associated with whether an individual had a SG. Specifically, it was found that individuals who were categorised as 'not working' were less likely to have a SG.

***Table 3***.

*Logistic Regression Model of Predictors of SGs Using the Original Criteria*

| | *$R^2$ (Cox & Snell)* | *B (95% CI)* | *Std. Error* | *Wald ʒ* | *Odds Ratio ( 95% CI)* | *p* |
|---|---|---|---|---|---|---|
| Constant | | -2.63 (-4.13, -1.54) | .71 | 12.17 | .07 | <.001 |
| 1st PHQ-9 Score | .03 | .09 (.02, .19) | .04 | 5.45 | 1.1 (1.02, 1.19) | .008 |
| Employment Status- not working | .07 | -1.14 (-2.61, -.3) | 2.17 | 3.95 | .32 (.1, .98) | .02 |
| *IV Score | .08 | .12 (-.2, .4) | .16 | .71 | 1.12 (.86, 1.47) | .44 |

*Not included in the final regression model.

Note. PHQ-9 = Patient Health Questionnaire-9, IV = Intraindividual variability.

Intraindividual variability in depression symptomology was not significantly associated with SGs using the original criteria. No interaction effects were found between variables. No difference was found between intraindividual variability scores for participants receiving CBT or PCET using the original SGs criteria $t(190) = .25$, $p = .80$.

### *Altered Criteria*

Table 4 shows the final regression model of predictors of SGs using the altered criteria. The final regression model was: $\chi^2$ (2) = 12.87, $p = .002$. The model was able to correctly classify SGs 57% of the time. The model was more accurate with predicting non-SGs (68.3%) compared to SGs (43.8%).

***Table 4.***

*Logistic Regression Model of Predictors of SGs Using the Altered Criteria.*

| | *$R^2$ (Cox & Snell)* | *B (95% CI)* | *Std. Error* | *Wald ʒ* | *Odds Ratio ( 95% CI)* | *p* |
|---|---|---|---|---|---|---|
| Constant | | .53 (-.77, 1.76) | .65 | .69 | 1.69 | .41 |
| 1st PHQ-9 Score | .04 | -.09 (-.15, -.03) | .03 | 7.23 | .92 (.86, .98) | .007 |
| IV Score | .07 | .30 (.04, .67) | .13 | 5.41 | 1.35 (1.05, 1.73) | .02 |

Note. PHQ-9 = Patient Health Questionnaire-9, IV = Intraindividual variability.

Using the altered SGs criteria the initial PHQ-9 score was significantly associated with whether an individual had a SG. More specifically individuals with higher initial severity scores were less likely to experience a SG. This is the reverse from the findings using the original criteria. It was also found that intraindividual variability in depression symptomology was significantly associated with whether an individual had a SG. Specifically, when intraindividual variability was higher and there was a greater amount of fluctuation in depression symptomology, an individual was more likely to experience a SG.

None of the other variables included in the model and described above were significantly associated with SGs using the altered criteria, and no interaction effects were found between variables. There was no difference in intraindividual variability scores for participants receiving CBT or PCET using the altered SGs criteria $t(187) = 1.81$, $p = .07$.

**Sudden Gains and Treatment Outcomes**

Table 5 shows the initial PHQ-9 scores and PHQ-9 change scores from the first to last treatment sessions overall and using both criteria, for the SG and non-SG groups. Figure 3 presents the mean PHQ-9 scores for the SG and non-SG groups from the first to final session, using both SGs criteria. When using the original SGs criteria there was a significant difference in the initial severity of depression for those who had a sudden gain and those who did not: $t(206) = -2.96$, $p = .002$, with participants who had a SG having higher initial depression severity scores. Using the original criteria there was also a significant difference in the PHQ-9 change scores for the SG and non-SG groups: $t(206) = -6.33$, $p < .001$. With the SG group having a greater reduction in PHQ-9 scores, suggesting a greater reduction in depression

severity. No other differences were found between the groups on initial severity of depression or change in PHQ-9 scores throughout treatment.

Table 6 shows the initial PHQ-9 scores and PHQ-9 change scores overall and for the CBT and PCET group. No significant differences were found between the groups on initial severity of depression or PHQ-9 change scores throughout treatment.

***Table 5.***

*The Differences Between the SG and Non-SG Groups Using Both SG Criteria*

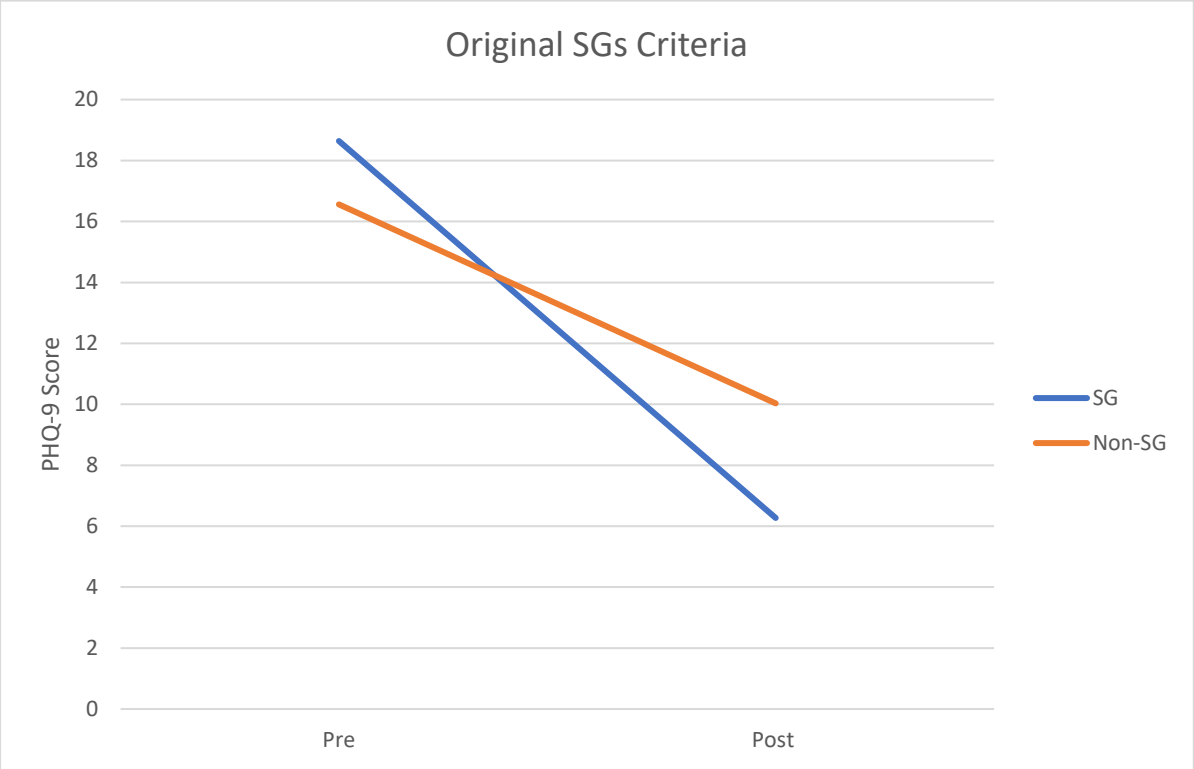
| | Original SGs Criteria | | | | Altered SGs Criteria | | | |
|---|---|---|---|---|---|---|---|---|
| | Overall | SG | Non-SG | Level of significance | Overall | SG | Non-SG | Level of significance |
| 1st PHQ-9 M (SD) | 17.2 (4.77) | 18.64 (4.08) | 16.56 (4.93) | *p*=.003* | 16.82 (4.96) | 16.64 (4.59) | 16.69 (5.23) | *p*=.61 |
| PHQ-9 Change M (SD) | -8.35 (6.37) | -12.19 (5.54) | -6.64 (5.97) | *p*<.001** | -7.47 (6.54) | -8.18 (5.55) | -6.91 (7.19) | *p*=.13 |

*significant at p=.05, ** significant at p<.001.

Note. SG= Sudden Gain, PHQ-9= Patient Health Questionnaire-9.

***Figure 3.***

*The Mean PHQ-9 Scores for the SG and Non-SG Groups From the First to Final Session Using Both SGs Criteria.*



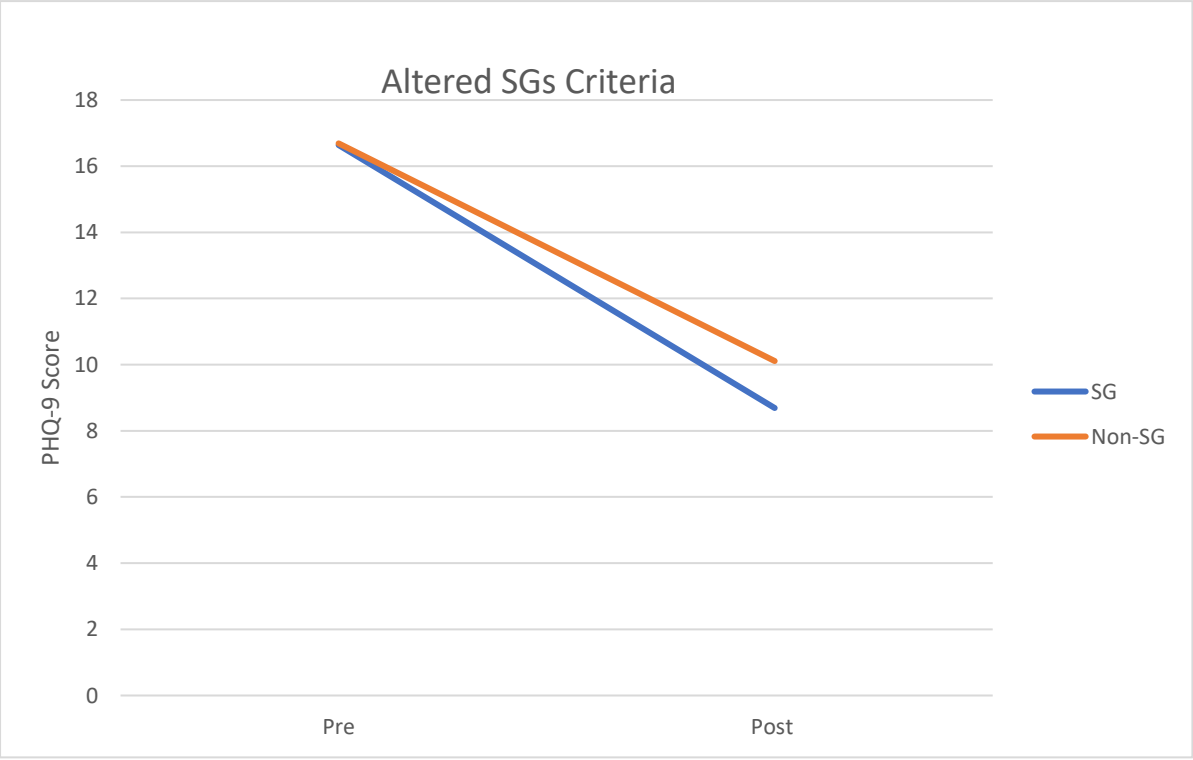



***Table 6.***

*The Differences Between the CBT and PCET groups for Both SGs Criteria in Initial Depression Severity and Change Over Treatment.*

| | **Original SGs Criteria** | | | **Altered SGs Criteria** | | |
|---|---|---|---|---|---|---|
| | **CBT** | **PCET** | **Level of significance** | **CBT** | **PCET** | **Level of significance** |
| 1st PHQ-9 Mean (SD) | 16.6 (5.2) | 17.58 (4.39) | p=.19 | 16.27 (5.31) | 17.3 (4.6) | p=.11 |
| PHQ-9 Change Mean (SD) | -8.63 (7.06) | -8.13 (5.79) | p=.58 | -7.54 (7.09) | -7.4 (6.05) | p=.87 |

Note. SG= Sudden Gain; CBT= Cognitive Behaviour Therapy; PCET; Person Centred Experiential Therapy; PHQ-9= Patient Health Questionnaire-9.

***Original Criteria***

Table 7 shows the linear model of predictors of treatment outcomes using the original SGs criteria. The overall regression was statistically significant ($R^2$ = .28, $F$(4, 173) = 17.4, $p$ < .001), with the associated variables accounting for 28% of the variance.

It was found that the first session PHQ-9 score was significantly associated with treatment outcome and individually explained 15% of the variance, suggesting that higher initial PHQ-9 scores resulted in higher final PHQ-9 scores, and worse treatment outcomes. It was also found that SGs using the original criteria were

***Table 7***.

*Linear Model of Predictors of Treatment Outcomes using the Original Criteria*

| | *Adjusted* $R^2$ | *B (95% CI)* | *β* | *Std. Error* | *t* | *p* |
|---|---|---|---|---|---|---|
| Constant | | 2.85 (-.15, 5.65) | | 1.6 | 1.54 | .08 |
| 1st PHQ-9 Score | .15 | .52 (.37, .7) | .4 | .08 | 6.16 | <.001 |
| Sudden Gain | .22 | -3.46 (-4.94, -2.12) | -.26 | .82 | -3.98 | <.001 |
| IV Score | .26 | -1.01 (-1.69, -.29) | -.21 | .33 | -3.17 | .003 |
| Gender | .27 | 1.57 (.09, 3.19) | .13 | .77 | 2.03 | .05 |

Note. PHQ-9 = Patient Health Questionnaire-9; IV = Intraindividual variability

significantly associated with improved treatment outcomes and explained an additional 7% of the variance.

The model found that intraindividual variability was significantly associated with treatment outcomes and explained an additional 7% of the variance. Specifically, lower levels of intraindividual variability in depression symptomology resulted lower final PHQ-9 scores and therefore improved treatment outcomes. Gender was also associated with improved treatment outcomes. The level of variance explained by gender was small at 2%. Specifically, females were more likely to have a better treatment outcome. An independent t-test showed a significant difference between males and females at the end of therapy ($t(206) = -2.21$, $p = .03$). No difference was found between gender in initial depression severity scores ($t(206) = -.62$, $p = .54$). No significant interactions were found between variables and SGs did not moderate the relationship between intraindividual variability and outcomes.

***Altered Criteria***

Table 8 shows the linear model of predictors of treatment outcomes using the altered SGs criteria. The overall regression was statistically significant ($R^2$ = .25, $F$(3, 188) = 21.25, $p$ < .001), with three associated variables accounting for 24% of the variance.

The first session PHQ-9 score was significantly associated with treatment outcome and individually explained 14% of the variance, suggesting that higher initial PHQ-9 scores resulted in higher final PHQ9-scores and therefore worse treatment outcomes. Intraindividual variability in depression symptomology was significantly associated with treatment outcome and individually explained an additional 8% of the variance, suggesting that higher levels of intraindividual variability resulted in lower final PHQ-9 scores and therefore improved treatment outcomes. Both of these variables had a similar impact on treatment outcomes in both SGs criteria.

***Table 8.***

*Linear Model of Predictors of Treatment Outcomes Using the Altered SGs Criteria*

| | *Adjusted $R^2$* | *B (95% CI)* | *β* | *Std. Error* | *t* | *p* |
|---|---|---|---|---|---|---|
| Constant | | -1.61 (-7.64, 4.94) | | 3.13 | -.58 | .59 |
| 1st PHQ-9 Score | .13 | .49 (.33, .65) | .39 | .08 | 6.11 | <.001 |
| IV Score | .22 | -1.45 (-2.17, -.78) | -.29 | .37 | -4.61 | <.001 |
| Ethnicity | .24 | 2.9 (.57, 5.26) | .17 | 1.22 | 2.73 | .01 |
| *Sudden Gain | .24 | .52 (-1.06, 2.05) | .04 | .74 | .65 | .49 |

*Not included in final regression model.

Note. PHQ-9= Patient Health Questionnaire-9; IV = Intraindividual variability.

In this model, ethnicity was also associated with treatment outcomes, suggesting that participants who identified as White British had improved treatment outcomes. The level of variance explained by ethnicity was small at 3%. No other variables were found to have significant associations with treatment outcomes, including SGs using the altered criteria. No significant interactions were found between variables and SGs did not moderate the relationship between intraindividual variability and outcomes.

## Discussion

In order to provide further evidence for or against Aderka and Shalom's (2021) revised theory of sudden gains, the primary aim of the current investigation was to determine whether intraindividual variability in depression symptomatology during treatment predicted SGs. The findings from this investigation are mixed, depending on the criteria used to identify SGs. When using the original criteria, intraindividual variability in depression symptomology was not significantly associated with SGs, contradicting the model. However, when using the altered criteria, intraindividual variability did predict SGs supporting the predictions of this study and Aderka and Shalom's (2021) revised theory of SGs.

A secondary aim of this investigation was to determine whether SGs predicted improved treatment outcomes in two different psychological therapies for depression. Supporting previous literature in this area, SGs were significantly associated with improved treatment outcomes, but only when identified using the original SGs criteria. SGs identified using the altered criteria were not found to be associated with improved treatment outcome. In addition, there was no difference between SGs identified in the different therapy modalities, adding support to the

growing literature that SGs are a common, pantheoretical phenomenon (Aderka & Shalom, 2021; Brockmeyer et al., 2023).

The final research aim was to compare two different SGs criteria, as suggested in the most recent meta-analysis (Shalom & Aderka, 2020). In contrast to other research projects which have identified SGs using more than one criterion, very different findings were found depending on the criteria used. It was expected that more SGs would be found using the altered criteria but the contrasting findings in the regression analyses were unexpected.

**Intraindividual variability and Sudden Gains**

Using the original criteria, initial treatment severity and employment status were associated with the occurrence of SGs. Specifically, individuals who were not working were less likely to experience a SGs. This category was condensed from individuals who were retired, seeking work and unable to work due to disability or sickness. One possible explanation for this finding is that those who were unemployed were experiencing higher levels of depression which impacted their ability to work. Using the original criteria, higher rates of depression were associated with experiencing a SG. To the researcher's knowledge, the association between employment and SGs has not been identified previously and it will be important for this finding to be replicated and to explore further what it is about not working that is associated with SGs. It is also important to note that no association was found between employment status and SGs using the altered criteria.

Using the altered criteria, intraindividual variability in depression symptomatology and initial treatment severity were the only predictors of SGs. As mentioned above, this provides evidence for the revised model of SGs (Aderka & Shalom, 2021). However, it is difficult to determine why this result was found using

one set of criteria but not the other. One possibility is that due to the intraindividual variability score not being able to be calculated for every participant due to the method, the analysis was underpowered in the original criteria. Another possibility is that the ability to identify SGs in fewer sessions using the altered criteria gave larger mean intraindividual variability scores and these larger scores were more likely to predict the large sudden reductions that classify SGs. Either way, more investigation is needed into the revised model of SGs.

**Sudden Gains and Treatment Outcomes**

When reviewing the literature, it has been found that SGs identified using the Kelly (2005) criteria have predicted improved treatment outcomes in three out of six investigations (Bisby et al. 2022; Clerkin et al., 2008; Singh et al., 2021). Suggesting that although it is widely documented that SGs are associated with improved treatment outcomes, when using the Kelly criteria this association is only found half of the time. Using information extracted from the Shalom and Aderka (2020) meta-analysis, the association between improved treatment outcomes and SGs using the original criteria is not found in 13-33% of investigations, which is much less than when using the Kelly et al. (2005) criteria.

There could be many different reasons behind this finding, however, as the most prominent difference between the two criteria is the ability to identify early SGs in therapy. This is a helpful place to begin. A large body of evidence has found that early gains in therapy are important. However, some evidence suggests that very early SGs may not predict overall treatment outcome. For example, Clerkin et al. (2008) compared SGs that occurred following session one and session two and found that the SGs following session one were not predictive of outcome but that the SGs following session two were. From a clinical perspective this is understandable

as the patient may experience an initial improvement from being able to talk about their difficulties. However, this may not be maintained. Additionally, the altered criteria have been identified as being less stringent and therefore it may be identifying small natural symptom fluctuations in treatment as SGs which are not predictive of improved treatment outcomes.

Using the original criteria, initial treatment severity, being female, and high levels of intraindividual variability predicted improved treatment outcomes. There was a significant difference between the SG and non-SG groups in their initial depression severity which supports Silverstone's (2023) recent finding that individuals with higher initial depression severity are more likely to have a SG. The findings around gender and IMD may be linked to specific factors with the IAPT service or the therapists providing treatment. Using the altered criteria, initial treatment severity and being White British was also found to be associated with improved treatment outcomes. Again, this may be due to service and therapist factors.

Interestingly, both criteria found that intraindividual variability in depression symptomology is linked with improved treatment outcomes. The fact that this was found using both criteria suggests that it has a strong connection to outcome. Although SGs were not found to mediate the association between intraindividual variability and improved outcome, this finding provides support for the revised theory of SGs. As it suggests that natural symptom fluctuation around a reducing mean can be picked up and built upon by therapists, whereas when there is less symptom fluctuation there is less improvement.

**Sudden Gains Criteria**

As the current research has found contradictory findings depending on the SGs criteria, it is important to note the large differences between the findings when using a sample which is mostly the same. This has important implications for the SGs research, as previously differences between the criteria appeared to be small and therefore researchers have used a variety of SGs criteria based on the research needs and made further adaptations to the criteria. This research adds support to the suggestion that SGs using different criteria are distinct and non-comparable phenomenon. While there are similarities, the factors that lead to and result from SGs when identified using different criteria are different and it is important for this to be recognised when conducting SGs research.

The importance of creating a standardised SGs criteria must be stressed. Based on the current research, it would seem sensible to base this on the original Tang and DeRubeis (1999) criteria as this has a much larger body of evidence and has been found to predict improved outcomes related to SGs more consistently, which is the most important and clinically relevant element of SGs research.

**Limitations**

Some limitations relating to the identification of SGs should be mentioned. Firstly, participants were not excluded based on the initial treatment score, only their depression severity at screening. A variety of time periods passed for participants between screening and the beginning of treatment which means that initial depression severity changed and for some participants put them below the clinical threshold; for example, one participant started treatment with a PHQ-9 score of 0. Other SGs research has applied this exclusion criteria from the beginning of treatment; this would have affected the replicability of the current research. Also, it

would have provided a more consistent baseline between participants at the beginning of treatment.

Similar to the above point, some SGs research only included participants who attended eight therapy sessions. It was decided that this would not be applied to the current study in order to increase the sample size and subsequent statistical power. However, a criticism could be made that participants who had less than eight sessions may have benefitted less from the therapy process.

When comparing the SGs criteria, it must be acknowledged that the two samples had a different number of participants. This decision was made because it was deemed more accurate when making comparisons to other SGs research using the different exclusion criteria. Also, as the two samples were not compared statistically, there was no risk that the differences found between the criteria would be due to differences in the sample size. However, the difference in sample size does create a power difference between the criteria, which was explored in post-hoc analysis. The only analysis that was found to be underpowered was the logistic regression using the altered criteria. As this analysis found a significant result, this does not appear to have been an issue.

**Clinical Implications**

Research into the phenomenon of SGs has important clinical implications, as obtaining a better understanding of the factors which lead to SGs could help inform and improve clinical practice. These specific factors could then actively be pursued in therapy to support the achievement of the improved outcomes which have been linked to SGs.

Using the findings from the current research it may be helpful if clinicians are able to identify at intake or early in therapy whether patients' symptoms appear to fluctuate frequently. As this could be an early indicator that a SGs may be more likely to occur throughout the course of treatment. This knowledge could support clinicians to make helpful treatment decisions regarding the therapeutic process and encourage the use of symptom fluctuation monitoring using regular outcome measures. Then, if a SG occurs, the therapist could support the patient by actively pursuing the 'upward spiral', as described in the original and revised theory of SGs. This research can be applied across treatment and therapy modalities and in a range of diverse settings.

## Conclusion

In conclusion, intraindividual variability in depression symptomology was associated with SGs but only when using altered criteria to identify SGs. In contrast, SGs were associated with improved treatment outcomes in CBT and PCET for depression, but only when using the original SGs criteria. Higher levels of intraindividual variability were associated with improved treatment outcomes using both SGs criteria. These findings provide mixed support for the revised theory of SGs, and more research is needed to explore these findings further. The current research highlights the importance of creating standardised criteria to identify SGs. Until this occurs there is a risk that the different criteria are exploring and comparing distinct phenomenon.

## References


Aderka, I. M., Appelbaum-Namdar, E., Shafran, N., & Gilboa-Schechtman, E. (2011). Sudden gains in prolonged exposure for children and adolescents with posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 79*(4), 441-446. https://doi.org/10.1037/a0024112

Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(1), 93-101. https://doi.org/10.1037/a0026455

Aderka, I. M. & Shalom, J. G. (2021). A revised theory of sudden gains in psychological treatments. *Behavioural Research and Therapy, 139*, 103830. https://doi.org/10.1016/j.brat.2021.103830

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *American Psychologist, 73*(1), 3–25. https://doi.org/10.1037/amp0000191

Barkham, M., Saxon, D., Hardy, G. E., Bradburn, M., Galloway, D., Wickramasekera, N., Keetharuth, A. D., Bower, P., King, M., Elliott, R., Gabriel, L., Kellett, S., Shaw, S., Wilkinson, T., Connell, J., Harrison, P., Ardern, K., Bishop-Edwards, L., Ashley, K., Ohlsen, S., Pilling, S., Waller, G. & Brazier, J. E. (2021). Person-centred experiential therapy versus cognitive behavioural therapy delivered in the English Improving Access to Psychological Therapies service for the treatment of moderate or severe depression (PRaCTICED): A

pragmatic, randomised, non-inferiority trial. *Lancet Psychiatry, 8*(6), 487-499. https://10.1016/S2215-0366

Beck, A. T., & Steer, R. A. (1987). *Beck Depression Inventory Manual*. Harcourt Brace.

Bisby, M. A., Scott, A. J., Hathway, T., Dudeney, J., Fisher, A., Gandy, M., Heriseanu, A. I., Karin, E., Titov, N., & Dear, B. F. (2022). Sudden gains in therapist-guided versus self-guided online treatments for anxiety or depression. *Journal of Consulting and Clinical Psychology, 90*(11), 861–871. https://doi.org/10.1037/ccp0000771

Bohn, C., Aderka, I. M., Schreiber, F., Stangier, U., & Hofmann, S. G. (2013). Sudden gains in cognitive therapy and interpersonal therapy for social anxiety disorder. *Journal of Consulting and Clinical Psychology, 81*(1), 177–182. https://doi.org/ 10.1037/a0031198

Brockmeyer, T., Titzmann, M., Zipfel, S., Wild, B., Resmark, G., Teufel, M., Giel, K., de Zwaan, M., Dinkel, A., Herpertz, S., Burgmer, M., Löwe, B., Tagay, S., Rothermund, E., Zeeck, A., Herzog, W., & Friederich, H. C. (2023). The role of general change mechanisms in sudden gains in the treatment of anorexia nervosa. *Behaviour Research and Therapy*, *163*, 104285. https://doi.org/10.1016/j.brat.2023.104285

Cavallini, A. Q., & Spangler, D. L. (2013). Sudden gains in cognitive-behavioral therapy for eating disorders. *International Journal of Cognitive Therapy, 6*(3), 292–310. https://doi.org/10.1521/ijct.2013.6.3.292

Clerkin, E. M., Teachman, B. A., & Smith-Janik, S. B. (2008). Sudden gains in group cognitive-behavioral therapy for panic disorder. *Behaviour Research and Therapy, 46*(11), 1244–1250. https://doi.org/10.1016/j.brat.2008.08.002

Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science, 1*(3), 98–101. https://doi.org/10.1111/1467-8721.ep10768783

Dour, H. J., Chorpita, B. F., Lee, S. & Weisz, J. R. (2013).Sudden gains as a long-term predictor of treatment improvement among children in community mental health organizations. *Behaviour Research and Therapy, 51*(9), 564-572. https://doi.org/10.1016/j.brat.2013.05.012

Drymalski, W. M., & Washburn, J. J. (2011). Sudden gains in the treatment of depression in a partial hospitalization program. *Journal of Consulting and Clinical Psychology, 79*(3), 364-368. https://doi.org/10.1037/a0022973

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

George, D. & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference 17.0 update.* Pearson.

Gibby, B. (2015). *Sudden gains in trauma-focused cognitive-behavioral therapy for child and adolescent trauma victims.* [Master's dissertation, Towson University]. Towson University Institutional Repository.

Glen, S. (2022, January 28). *Non-normal Distribution.* Statisticshowto.com. https://www.statisticshowto.com/probability-and-statistics/non-normal-distributions/

Hamdeh, A. A., Bjureberg, J., Lenhard, F., Hedman-Lagerlöf, E., Flygare, O., Lundström, L., Ljótsson, B., Mataix-Cols, D., Rück, C. & Andersson, E. (2019). Sudden gains in internet-based cognitive behavior therapy for obsessive-compulsive disorder. *Journal of Obsessive-Compulsive and Related Disorders, 21*, 75–81. https://doi.org/10.1016/j.jocrd.2018.12.005

Hardy, G. E., Cahill, J., Stiles, W. B., Ispan, C., Macaskill, N., & Barkham, M. (2005). Sudden gains in cognitive therapy for depression: A replication and extension. *Journal of Consulting and Clinical Psychology, 73*(1), 59–67. https://doi.org/10.1037/ 0022-006x.73.1.59

Hofmann, S. G., Schulz, S. M., Meuret, A. E., Moscovitch, D. A., & Suvak, M. (2006). Sudden gains during therapy of social phobia. *Journal of Consulting and Clinical Psychology, 74*(4), 687–697. https://doi.org/10.1037/0022-006x.74.4.687

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19. https://doi.org/10.1037//0022-006x.59.1.12

Keinonen, K. Kyllönen, H. Astikainen, P. & Lappalainen, R. (2018). Early sudden gains in an acceptance and values-based intervention: Effects on treatment outcome for depression and psychological flexibility. *Journal of Contextual Behavioral Science, 10*, 24-30. https://doi.org/10.1016/j.jcbs.2018.07.010

Kelly, M. A. R., Roberts, J. E., & Ciesla, J. A. (2005). Sudden gains in cognitive behavioral treatment for depression: When do they occur and do they matter? *Behaviour Research and Therapy, 43*(6), 703–714. https://doi:10.1016/j.brat.2004.06.002

Kelly, M. A. R., Roberts, J. E., & Bottonari, K. A. (2007). Non-treatment-related sudden gains in depression: The role of self-evaluation. *Behaviour Research and Therapy, 45*(4), 737–747. https://doi.org/10.1016/j.brat.2006.06.008

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

Krüger, A., Ehring, T., Priebe, K., Dyer, A. S., Steil, R., & Bohus, M. (2014). Sudden losses and sudden gains during a DBT-PTSD treatment for posttraumatic stress disorder following childhood sexual abuse. *European Journal of Psychotraumatology, 5*, 10. https://doi.org/10.3402/ejpt.v5.24470, 3402/ejpt.v5.24470

Kuck, S., Arntz, A., Rameckers, S. A., Lee, C. W., Boterhoven de Haan, K. L., Fassbinder, E., & Morina, N. (2023). Intraindividual variability and emotional change as predictors of sudden gains in imagery rescripting and EMDR for PTSD in adult survivors of childhood abuse. *Clinical Psychology & Psychotherapy.* Advance online publication. https://doi.org/10.1002/cpp.2855

Lewis, G., Pelosi, A. J., Araya, R. & Dunn, G. (1992). Measuring psychiatric disorder in the community: A standardized assessment for use by lay interviewers. *Psychological Medicine, 22*(2), 465-86. https://doi.10.1017/s0033291700030415

Masterson, C., Ekers, D., Gilbody, S., Richards, D., Toner-Clewes, B., & McMillan, D. (2014). Sudden gains in behavioural activation for depression. *Behaviour Research and Therapy, 60,* 34–38. https://doi.org/10.1016/j.brat.2014.06.008

McMillan, D., Gilbody, S., & Richards D. (2010). Defining successful treatment outcome in depression using the PHQ-9: A comparison of methods. Journal of Affective Disorders, 127(1-3), 122–129. https://doi.10.1016/j.jad.2010.04.030

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry: The Journal of Mental Science*, *180*, 461–464. https://doi.org/10.1192/bjp.180.5.461

Murphy, D. (2019). *Person-centred experiential counselling for depression* (2nd ed.). Sage Publishing; British Association for Counseling & Psychotherapy.

National Health Service England. (2021). *The Improving Access to Psychological Therapies Manual.* The National Collaborating Centre for Mental Health. https://www.england.nhs.uk/wp-content/uploads/2018/06/the-nhs-talking-therapies-manual-v6.pdf

Saxon, D., Ashley, K., Bishop-Edwards, L., Connell, J., Harrison, P., Ohlsen, S., Hardy, G., E., Kellett, S., Mukuria, C., Mank, T., Bower, P., Bradburn, M., Brazier, J., Elliott, R., Gabriel, L., King, M., Pilling, S., Shaw, S., Waller, G. & Barkham, M. (2017). A pragmatic randomised controlled trial assessing the non-inferiority of counselling for depression versus cognitive-behaviour therapy for patients in primary care meeting a diagnosis of moderate or severe depression (PRaCTICED): Study protocol for a randomised controlled trial. *Trials, 18*, 93. https://doi.org/10.1186/s13063-017-1834-6

Shalom, J. G., & Aderka, I. M. (2020). A meta-analysis of sudden gains in psychotherapy: Outcome and moderators. *Clinical Psychology Review, 76*, 101827. https://doi.org/ 10.1016/j.cpr.2020.101827

Shalom, J.G., Strauss, A.Y., Huppert, J.D., Andersson, G., Agassi, O.D., & Aderka, I.M. (2020). Predicting sudden gains before treatment begins: An examination of pretreatment intraindividual variability in symptoms. *Journal of Consulting and Clinical Psychology, 88*(9), 809–817. https://doi.10.1037/ccp0000587

Shalom, J. G., Gilboa-Schechtman, E., Atzil-Slonim, D., Bar-Kalifa, E., Hasson-Ohayon, I., van Oppen, P., van Balkom, A. J. L. M., & Aderka, I. M. (2018). Intraindividual variability in symptoms consistently predicts sudden gains: An examination of three independent datasets. *Journal of Consulting and Clinical Psychology, 86*(11), 892–902. https://doi.org/10.1037/ccp0000344

Sheffield City Council (2023, March 5). *Sheffield's Population.* www.sheffield.gov.uk/home/your-city-council/population-insheffield.html

Silverstone, U. R., Roberge, P., Provencher, M. D., & Norton, P. J. (2023). An examination of sudden gain prevalence across cognitive-behavioural therapy for anxiety and depressive disorders: A quantitative analysis and meta-analytic review. *Journal of Anxiety Disorders*, 95, 102697. https://doi.org/10.1016/j.janxdis.2023.102697

Singh, T., Pascual-Leone, A., Morrison, O. P., & Greenberg, L. (2021). Working with emotion predicts sudden gains during experiential therapy for depression. *Psychotherapy Research*, *31*(7), 895-908. https://doi.org/10.1080/10503307.2020.1866784

Spitzer, R. L., Kroenke, K., Williams, J. B. W. & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., Iveson, M., & Hardy, G. E. (2003). Early sudden gains in psychotherapy under routine

clinic conditions: Practice-based evidence. *Journal of Consulting and Clinical Psychology, 71*(1), 14–21. https://doi.org/10.1037/0022-006X.71.1.14

Storch, E. A., McGuire, J. F., Schneider, S. C., Small, B. J., Murphy, T. K., Wilhelm, S., & Geller, D. A. (2019). Sudden gains in cognitive behavioral therapy among children and adolescents with obsessive compulsive disorder. *Journal of Behavior Therapy and Experimental Psychiatry, 64*, 92–98. https://doi.org/10.1016/j.jbtep.2019.03.003

Subramaniam, K., Krishnaswamy, S., Jemain, A. A., Hamid, A., & Patel, V. (2006). The Clinical Interview Schedule-Revised (CIS-R)-Malay Version, Clinical Validation. *The Malaysian Journal of Medical Sciences: MJMS, 13*(1), 58–62.

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*(6), 894–904. https://doi.org/10.1037/0022-006X.67.6.894

Tang, T. Z., DeRubeis, R. J., Beberman, R., & Pham, T. (2005). Cognitive changes, critical sessions, and sudden gains in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 73*(1), 168. https://doi.org/10.1037/0022-006X. 73.1.168

Thorisdottir, A. S., Tryggvadottir, A., Saevarsson, S. T., & Bjornsson, A. S. (2018). Brief report: Sudden gains in cognitive-behavioral group therapy and group psychotherapy for social anxiety disorder among college students. *Cognitive Behaviour Therapy, 47*(6), 462–469. https://doi.org/10.1080/16506073.2018.1466909

Vittengl, J. R., Clark, L. A., & Jarrett, R. B. (2005). Validity of sudden gains in acute phase treatment of depression. *Journal of Consulting and Clinical Psychology, 7*(1), 173–182. https://doi.org/10.1037/0022-006x.73.1.173

Zilcha-Mano, S., Errazuriz, P., Yaffe-Herbst, L., German, R. E., & DeRubeis, R. J. (2019). Are there any robust predictors of “sudden gainers,” and how is sustained improvement in treatment outcome achieved following a gain? *Journal of Consulting and Clinical Psychology, 87*(6), 491–500. https://doi.org/10.1037/ccp0000401

# Appendices

## Appendix A- University Ethics Approval


Downloaded: 07/12/2021
Approved: 07/12/2021

Holly McGrellis
Registration number: 200183655
Psychology
Programme: Doctorate in Clinical Psychology

Dear Holly

**PROJECT TITLE:** An exploration of Aderka & Shalom's revised model of sudden gains: A secondary analysis of the PRaCTICED trial data set.
**APPLICATION:** Reference Number 044245

This letter confirms that you have signed a University Research Ethics Committee-approved self-declaration to confirm that your research will involve only existing research, clinical or other data that has been robustly anonymised. You have judged it to be unlikely that this project would cause offence to those who originally provided the data, should they become aware of it.

As such, on behalf of the University Research Ethics Committee, I can confirm that your project can go ahead on the basis of this self-declaration.

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since full ethical review may be required.

Yours sincerely

Department Of Psychology Research Ethics Committee
Departmental Ethics Administrator

## Appendix B- Journal Article Reporting Standards (JARS)

### Title and Title Page

#### Title

Identify main variables and theoretical issues under investigation and the relationships between them.

Identify the populations studied.

Provide acknowledgment and explanation of any special circumstances, including

– registration information if the study has been registered

– use of data also appearing in previous publications

– prior reporting of the fundamental data in dissertations or conference papers

– sources of funding or other support

– relationships or affiliations that may be perceived as conflicts of interest

– previous (or current) affiliation of authors if different from location where the study was conducted

– contact information for the corresponding author

– additional information of importance to the reader that may not be appropriately included in other sections of the paper

### Abstract

#### Objectives

State the problem under investigation, including main hypotheses.

Participants

• Describe subjects (nonhuman animal research) or participants (human research), specifying their pertinent characteristics for the study; in animal research, include genus and species. Participants are described in greater detail in the body of the paper.

Study Method

• Describe the study method, including

– research design (e.g., experiment, observational study)

– sample size

– materials used (e.g., instruments, apparatus)

– outcome measures

– data-gathering procedures, including a brief description of the source of any secondary data. If the study is a secondary data analysis, so indicate.

Findings

• Report findings, including effect sizes and confidence intervals or statistical significance levels.

Conclusions

State conclusions, beyond just results, and report the implications or applications.

## Introduction

### Problem

State the importance of the problem, including theoretical or practical implications.

Review of Relevant Scholarship

- Provide a succinct review of relevant scholarship, including

– relation to previous work

– differences between the current report and earlier reports if some aspects of this study have been reported on previously

Hypothesis, Aims, and Objectives

State specific hypotheses, aims, and objectives, including

– theories or other means used to derive hypotheses

– primary and secondary hypotheses

– other planned analyses

State how hypotheses and research design relate to one another.

## Method

### Inclusion and Exclusion

Report inclusion and exclusion criteria, including any restrictions based on demographic characteristics.

Participant Characteristics

Report major demographic characteristics (e.g., age, sex, ethnicity, socioeconomic status) and important topic-specific characteristics (e.g., achievement level in studies of educational interventions).

In the case of animal research, report the genus, species, and strain number or other specific identification, such as the name and location of the supplier and the stock designation. Give the number of animals and the animals' sex, age, weight, physiological condition, genetic modification status, genotype, health–immune status, drug or test naïveté, and previous procedures to which the animal may have been subjected.

Describe procedures for selecting participants, including

– sampling method if a systematic sampling plan was implemented

– percentage of sample approached that actually participated

– whether self-selection into the study occurred (either by individuals or by units, such as schools or clinics)

Describe settings and locations where data were collected as well as dates of data collection.

Describe agreements and payments made to participants.

Describe institutional review board agreements, ethical standards met, and safety monitoring.

Sample Size, Power, and Precision

• Describe the sample size, power, and precision, including – intended sample size

– achieved sample size, if different from the intended sample size

– determination of sample size, including

› power analysis, or methods used to determine precision of parameter estimates

› explanation of any interim analyses and stopping rules employed

Measures and Covariates

Define all primary and secondary measures and covariates, including measures collected but not included in the report.

Data Collection

Describe methods used to collect data.

Quality of Measurements

• Describe methods used to enhance the quality of measurements, including

– training and reliability of data collectors

– use of multiple observations

Instrumentation

• Provide information on validated or ad hoc instruments created for individual studies, for individual studies (e.g., psychometric and biometric properties).

Masking

Report whether participants, those administering the experimental manipulations, and those assessing the outcomes were aware of condition assignments.

If masking took place, provide a statement regarding how it was accomplished and whether and how the success of masking was evaluated.

Estimate and report values of reliability coefficients for the scores analyzed (i.e., the researcher's sample), if possible. Provide estimates of convergent and discriminant validity where relevant.

Report estimates related to the reliability of measures, including

– interrater reliability for subjectively scored measures and ratings

– test–retest coefficients in longitudinal studies in which the retest interval corresponds to the measurement schedule used in the study

– internal consistency coefficients for composite scales in which these indices are appropriate for understanding the nature of the instruments being used in the study

Report the basic demographic characteristics of other samples if reporting reliability or validity coefficients from those samples, such as those described in test manuals or in norming information for the instrument.

Conditions and Design

• State whether conditions were manipulated or naturally observed. Report the type of design as per the JARS–Quant tables:

– experimental manipulation with participants randomized

– experimental manipulation without randomization

– clinical trial with randomization

– clinical trial without randomization

– nonexperimental design (i.e., no experimental manipulation): observational design, epidemiological design, natural history, and so forth (single-group designs or multiplegroup comparisons)

– longitudinal design

– *N*-of-1 studies

– replications

• Report the common name given to designs not currently covered in JARS–Quant.

Data Diagnostics

Describe planned data diagnostics, including

– criteria for post-data-collection exclusion of participants, if any

– criteria for deciding when to infer missing data and methods used for imputation of missing data

– definition and processing of statistical outliers

– analyses of data distributions

– data transformations to be used, if any

Describe the analytic strategy for inferential statistics and protection against experimentwise error for

– primary hypotheses

– secondary hypotheses

– exploratory hypotheses

## Results

### Participant Flow

Report the flow of participants, including

– total number of participants in each group at each stage of the study

– flow of participants through each stage of the study (include figure depicting flow, when possible; see the JARS–Quant Participant Flowchart)

Recruitment

Provide dates defining the periods of recruitment and repeated measures or follow-up.

Statistics and Data Analysis

• Provide information detailing the statistical and data-analytic methods used, including

– missing data

› frequency or percentages of missing data

› empirical evidence and/or theoretical arguments for the causes of data that are missing—for example, missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR)

› methods actually used for addressing missing data, if any

– descriptions of each primary and secondary outcome, including the total sample and each subgroup, that includes the number of cases, cell means, standard deviations, and other measures that characterize the data used

– inferential statistics, including

› results of all inferential tests conducted, including exact *p* values if null hypothesis significance testing (NHST) methods were used, and reporting the minimally sufficient set of statistics (e.g., *df*s, mean square [*MS*] effect, *MS* error) needed to construct the tests

› effect-size estimates and confidence intervals on estimates that correspond to each inferential test conducted, when possible

› clear differentiation between primary hypotheses and their tests–estimates, secondary hypotheses and their tests–estimates, and exploratory hypotheses and their test–estimates

– complex data analyses—for example, structural equation modeling analyses (see also Table 7), hierarchical linear models, factor analysis, multivariate analyses, and so forth, including

› details of the models estimated

› associated variance–covariance (or correlation) matrix or matrices

› identification of the statistical software used to run the analyses (e.g., SAS PROC GLM or the particular R package)

– estimation problems (e.g., failure to converge, bad solution spaces), regression diagnostics, or analytic anomalies that were detected and solutions to those problems.

– other data analyses performed, including adjusted analyses, if performed, indicating those that were planned and those that were not planned (though not necessarily in the level of detail of primary analyses).

Report any problems with statistical assumptions and/or data distributions that could affect the validity of findings.

## Discussion

Support of Original Hypotheses

Provide a statement of support or nonsupport for all hypotheses, whether primary or secondary, including

– distinction by primary and secondary hypotheses

– discussion of the implications of exploratory analyses in terms of both substantive findings and error rates that may be uncontrolled

Similarity of Results

Discuss similarities and differences between reported results and work of others.

Interpretation

• Provide an interpretation of the results, taking into account

– sources of potential bias and threats to internal and statistical validity

– imprecision of measurement protocols

– overall number of tests or overlap among tests

– adequacy of sample sizes and sampling validity

Generalizability

• Discuss generalizability (external validity) of the findings, taking into account – target population (sampling validity)

– other contextual issues (setting, measurement, time; ecological validity)

Implications

• Discuss implications for future research, program, or policy.

## Appendix C- Patient Health Questionnaire- 9 Item (PHQ-9)

Name _____ Date _____

| Over the *last 2 weeks*, how often have you been bothered by any of the following problems? | Not at all | Several days | More than half the days | Nearly every day |
|---|---|---|---|---|
| 1. Little interest or pleasure in doing things | 0 | 1 | 2 | 3 |
| 2. Feeling down, depressed, or hopeless | 0 | 1 | 2 | 3 |
| 3. Trouble falling or staying asleep, or sleeping too much | 0 | 1 | 2 | 3 |
| 4. Feeling tired or having little energy | 0 | 1 | 2 | 3 |
| 5. Poor appetite or overeating | 0 | 1 | 2 | 3 |
| 6. Feeling bad about yourself—or that you are a failure or have let yourself or your family down | 0 | 1 | 2 | 3 |
| 7. Trouble concentrating on things, such as reading the newspaper or watching television | 0 | 1 | 2 | 3 |
| 8. Moving or speaking so slowly that other people could have noticed? Or the opposite—being so fidgety or restless that you have been moving around a lot more than usual | 0 | 1 | 2 | 3 |
| 9. Thoughts that you would be better off dead or of hurting yourself in some way | 0 | 1 | 2 | 3 |

(For office coding: Total Score ____ = ____ + ____ + ____)

## Appendix D- Collaboration Statement

Collaboration statement

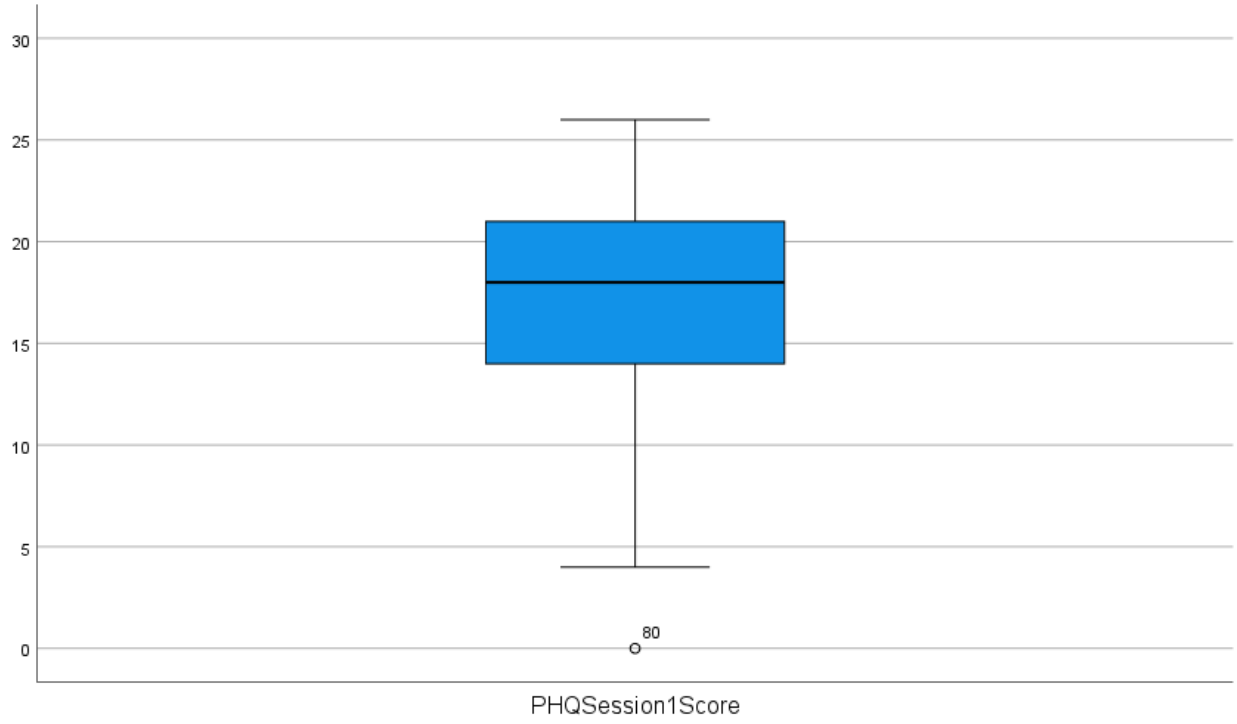
This document details the contributions of this thesis that were undertaken jointly by myself and peer CN. These contributions were undertaken equally. All other work in this thesis was undertaken independently.

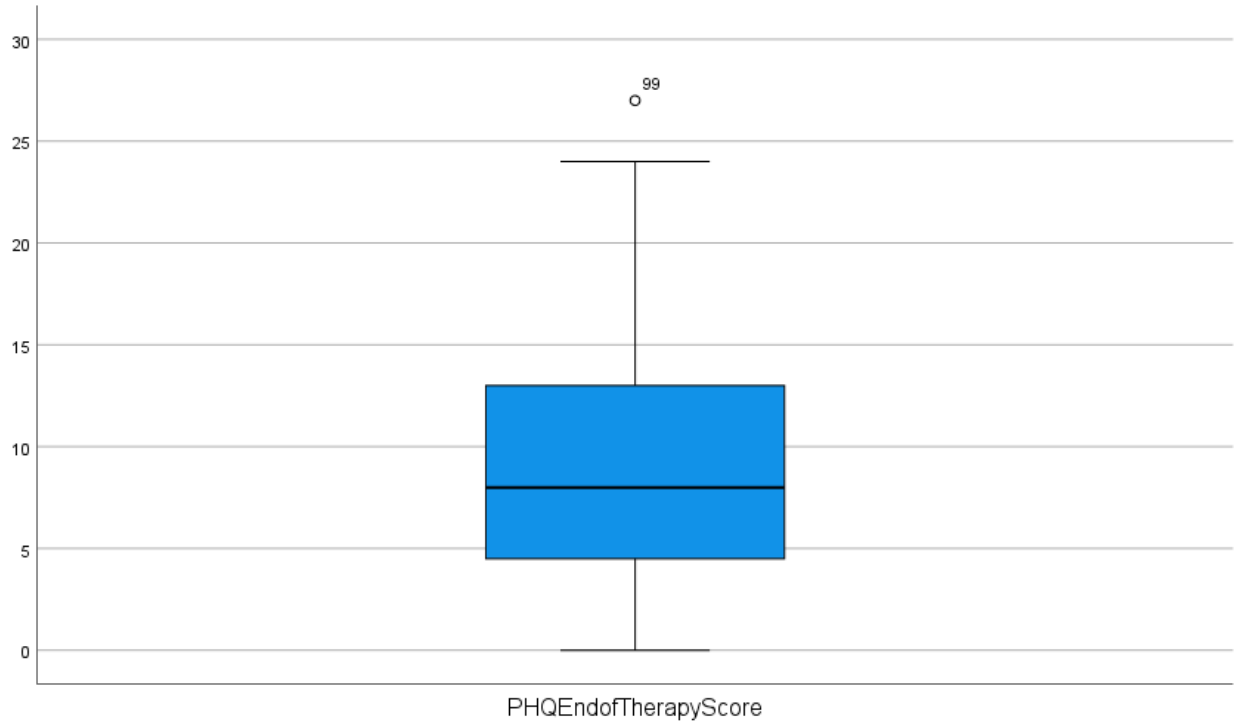
Work conducted in collaboration:

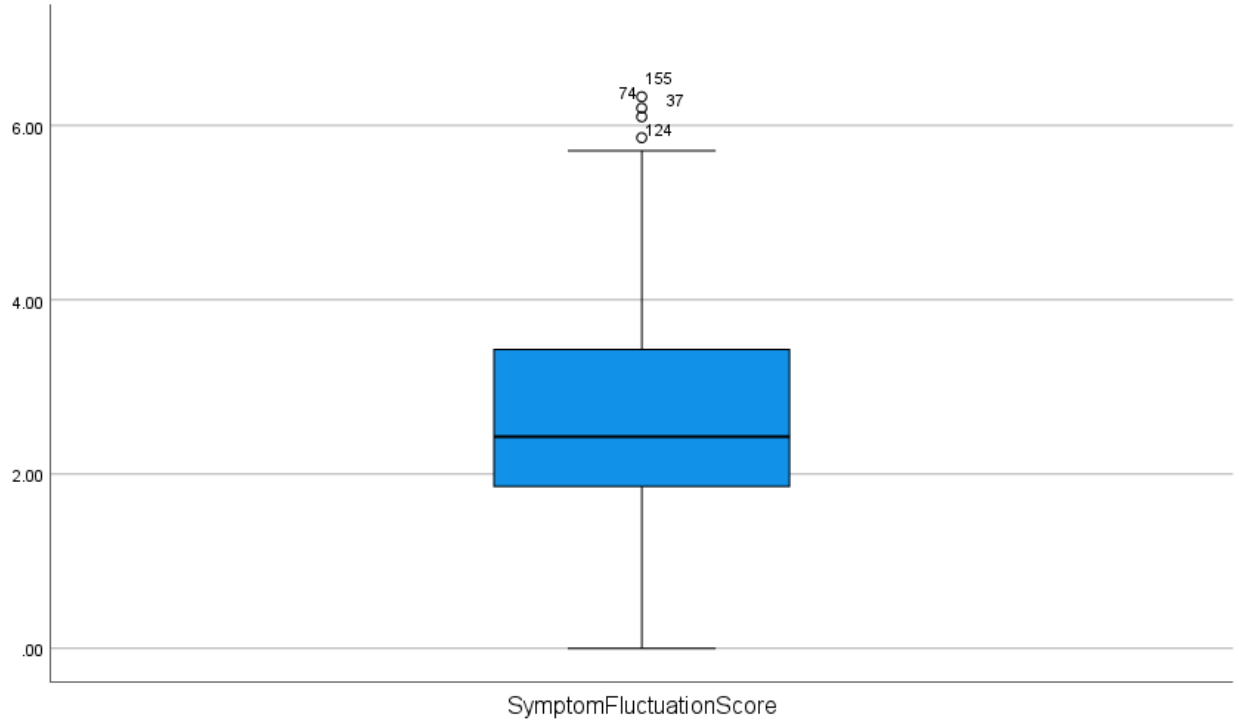
- Identification of sudden gains using the Kelly et al. (2005) criteria.
- Descriptive statistics of sudden gains found using the Kelly criteria.

## Appendix E- Box Plots to Identify Outliers

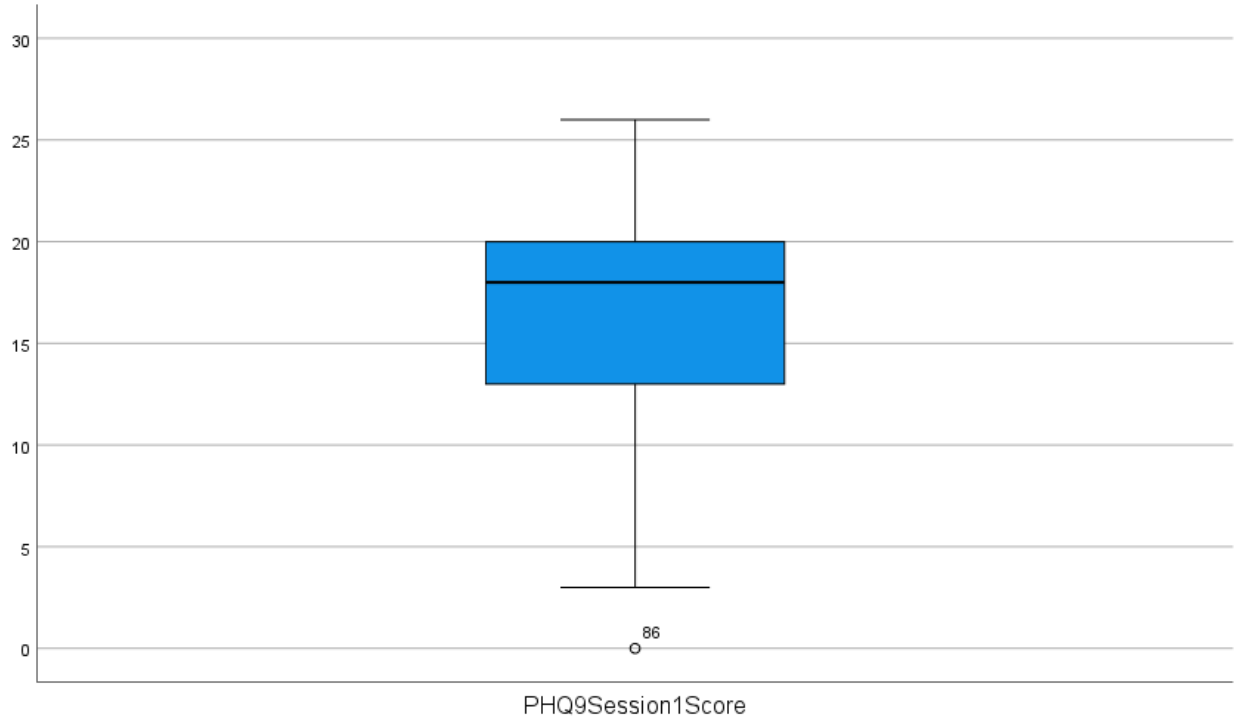
### *Original Criteria*

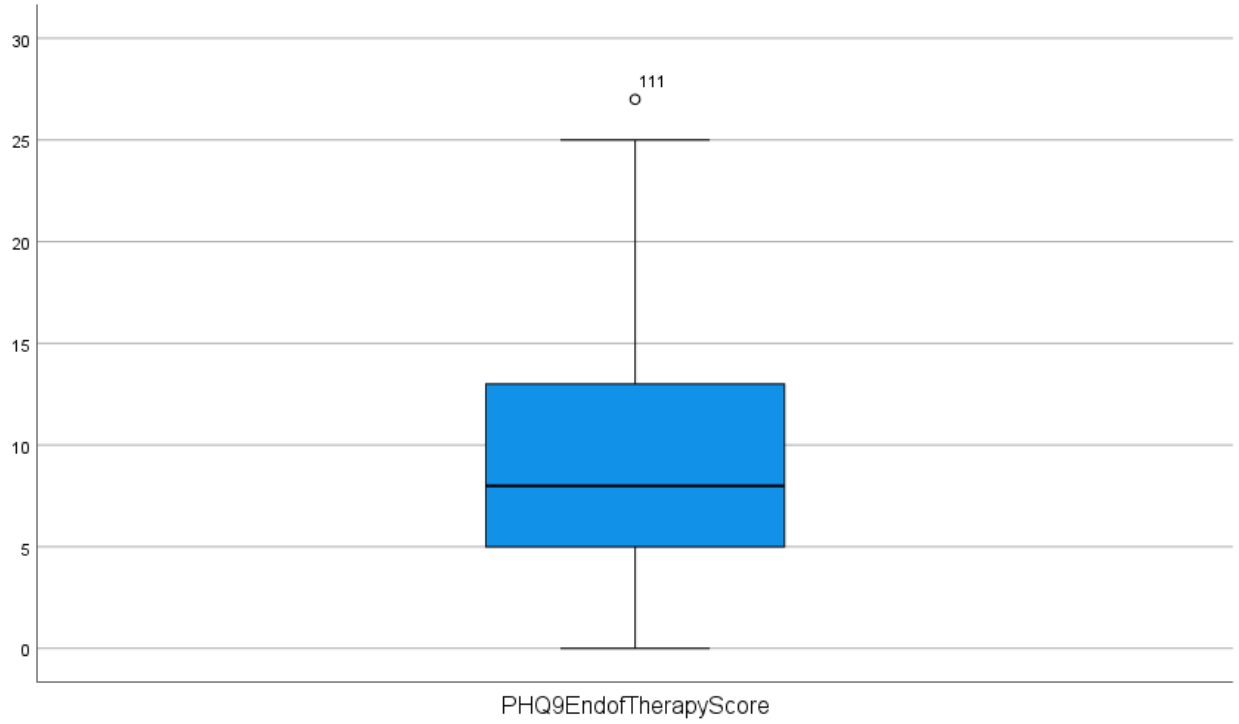


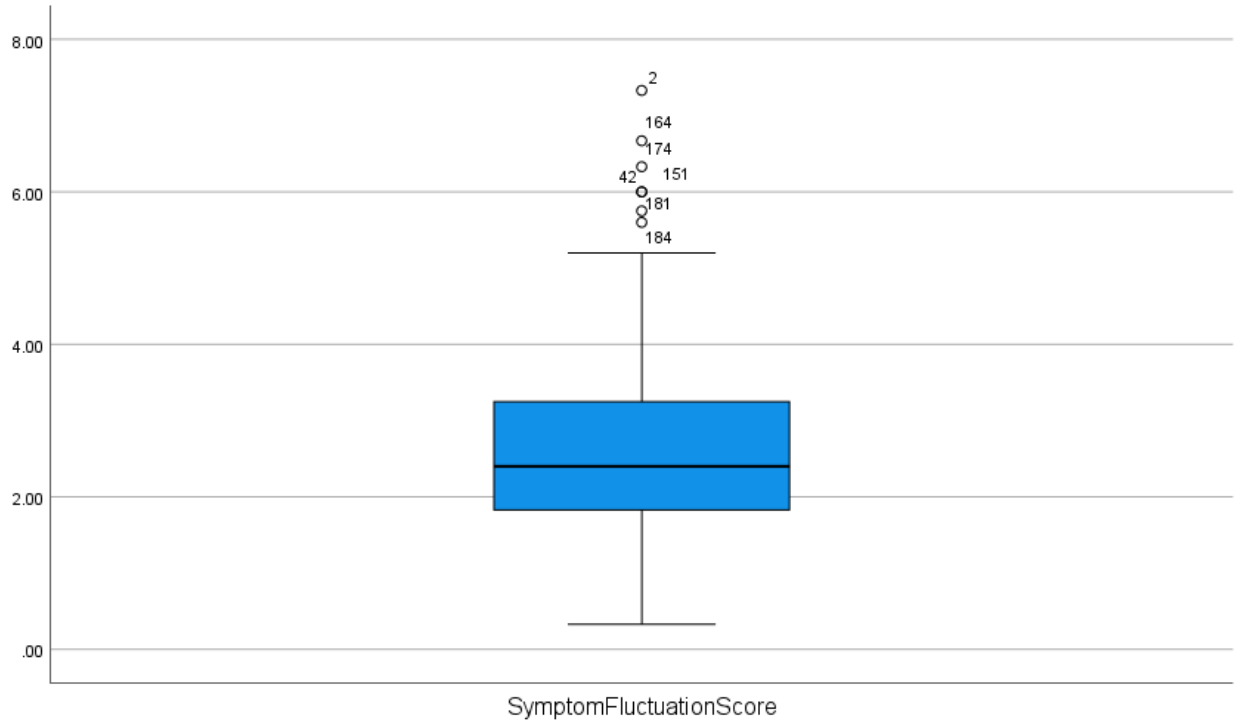






***Altered Criteria***




30
25
20
15
10
5
0
111
PHQ9EndofTherapyScore

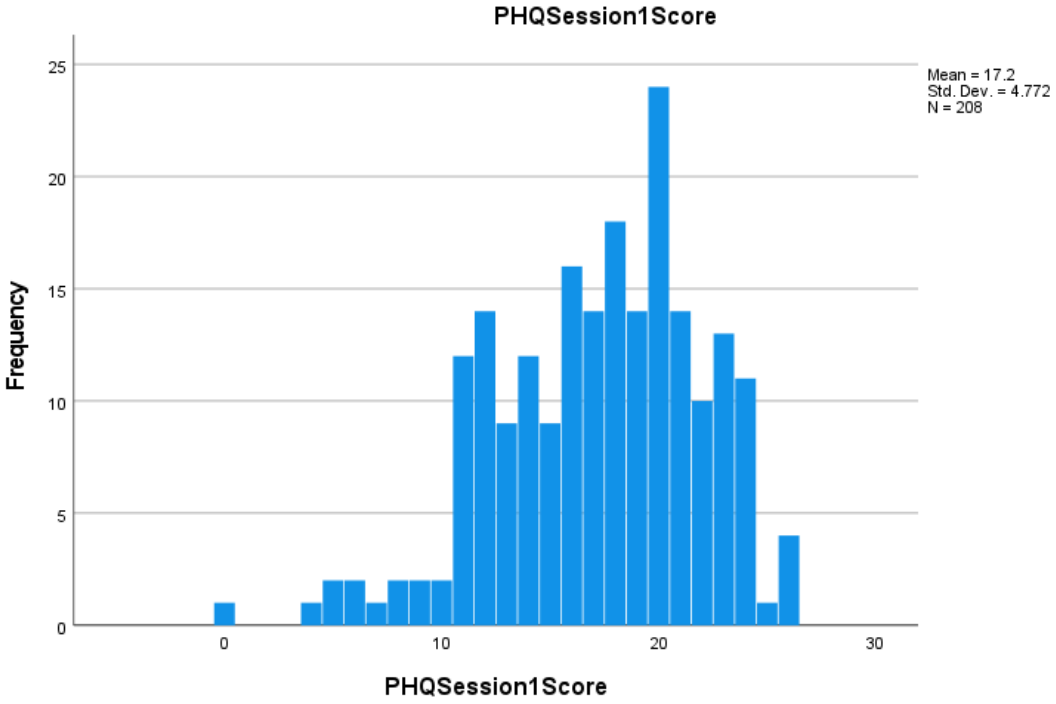

8.00
6.00
4.00
2.00
.00
2
164
174
42
151
181
184
SymptomFluctuationScore

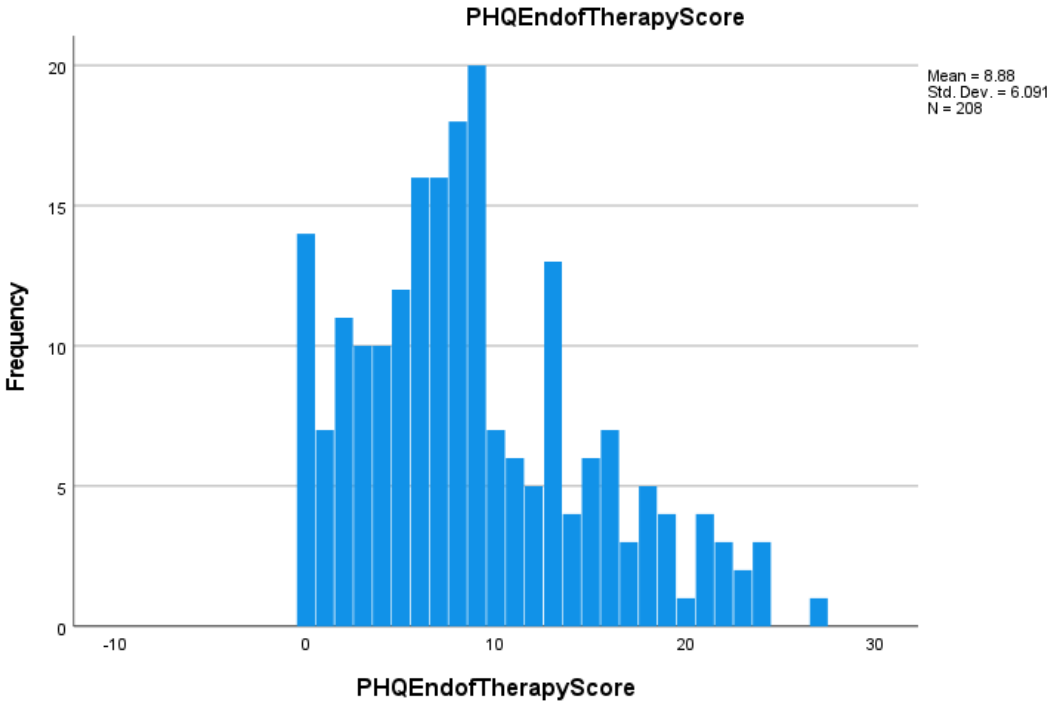
## Appendix F- Histograms to Assess Assumptions of Normality

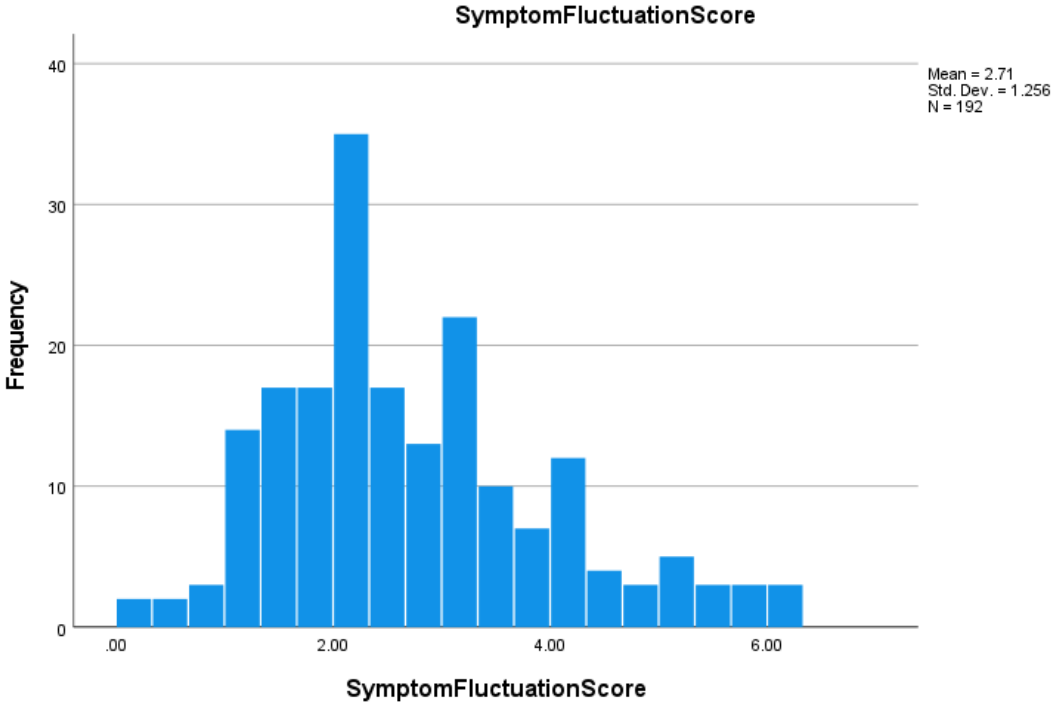
### *Original Criteria*

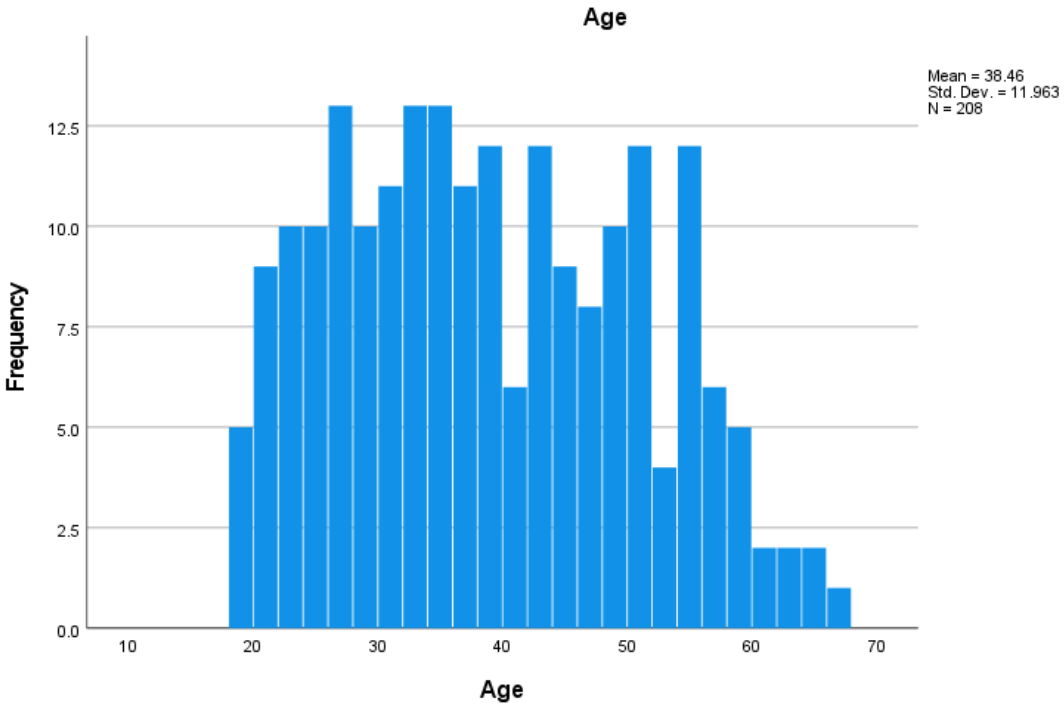
**Statistics**

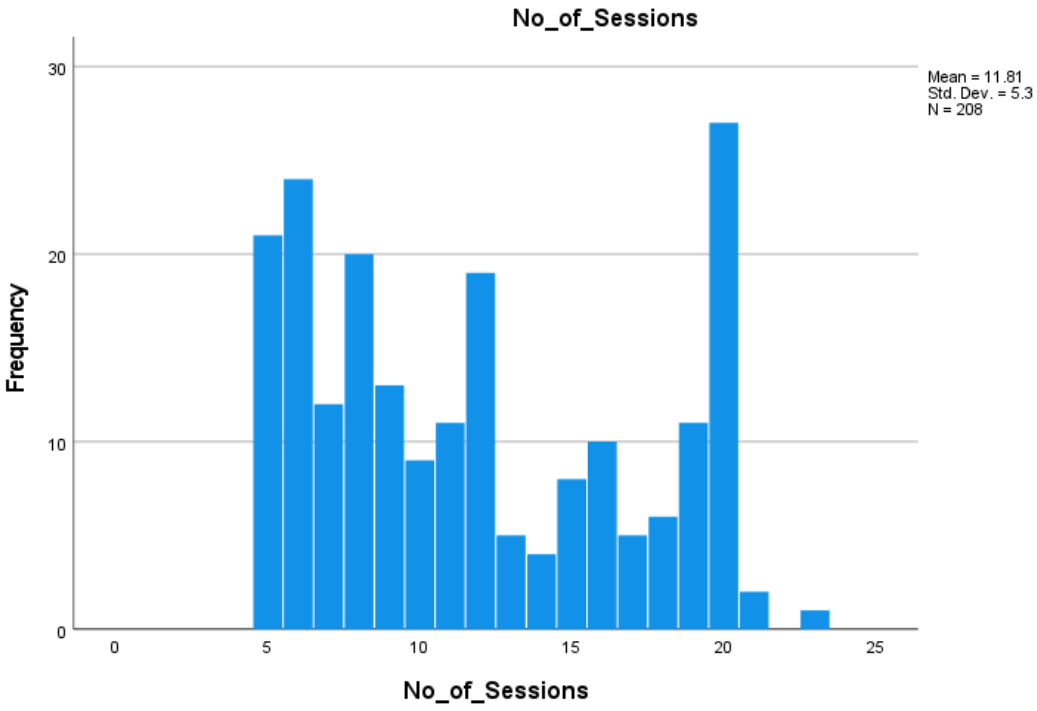
| | | PHQSession1Score | PHQEndofTherapyScore | SymptomFluctuationScore | Age | No_of_Sessions |
|---|---|---|---|---|---|---|
| N | Valid | 208 | 208 | 192 | 208 | 208 |
| | Missing | 0 | 0 | 16 | 0 | 0 |
| Skewness | | -.565 | .674 | .734 | .230 | .356 |
| Std. Error of Skewness | | .169 | .169 | .175 | .169 | .169 |
| Kurtosis | | .229 | -.087 | .280 | -.937 | -1.271 |
| Std. Error of Kurtosis | | .336 | .336 | .349 | .336 | .336 |




PHQEndofTherapyScore
Mean = 8.88
Std. Dev. = 6.091
N = 208
Frequency
20
15
10
5
0
-10
0
10
20
30
PHQEndofTherapyScore


SymptomFluctuationScore
Mean = 2.71
Std. Dev. = 1.256
N = 192
Frequency
40
30
20
10
0
.00
2.00
4.00
6.00
SymptomFluctuationScore

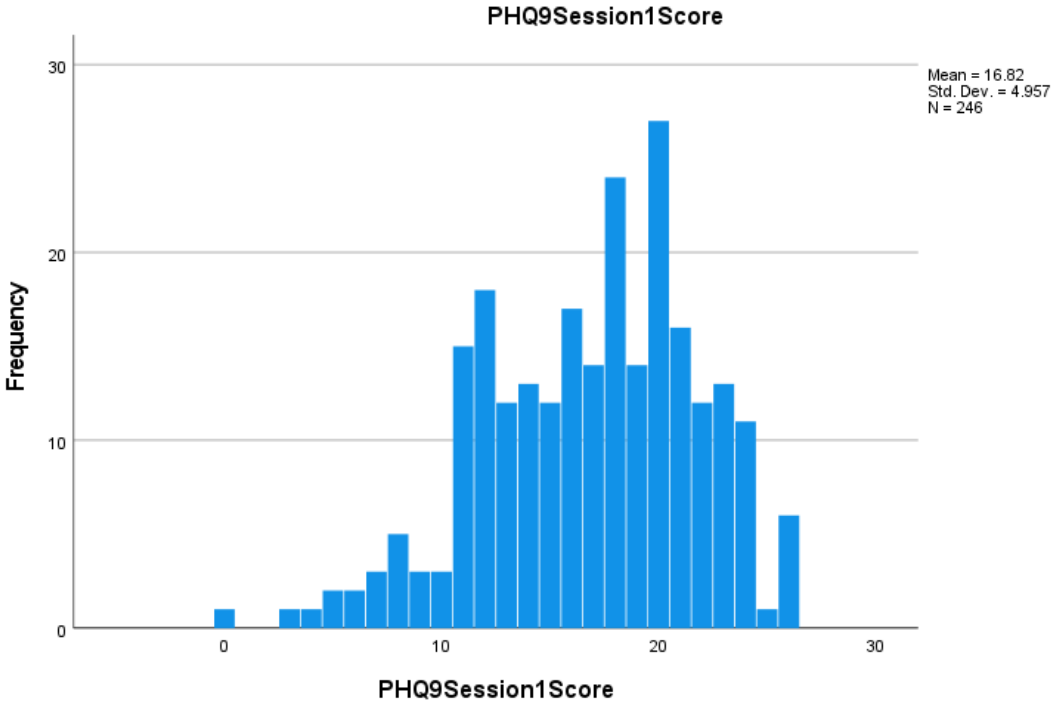

Age
Mean = 38.46
Std. Dev. = 11.963
N = 208
Frequency
12.5
10.0
7.5
5.0
2.5
0.0
10
20
30
40
50
60
70
Age

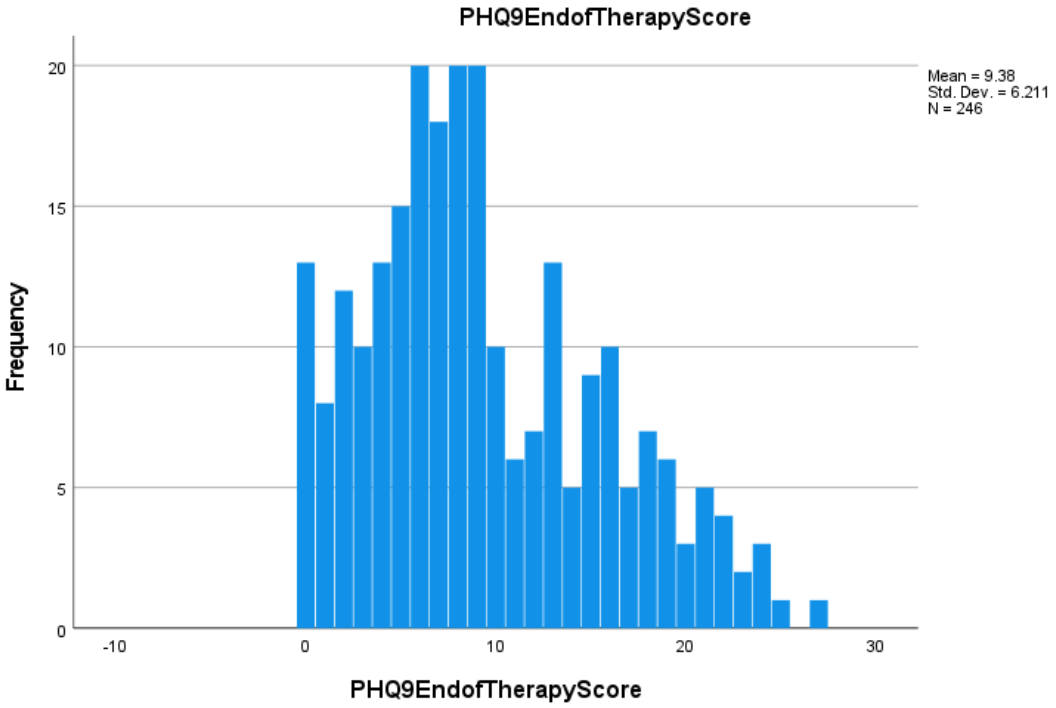

No_of_Sessions
Mean = 11.81
Std. Dev. = 5.3
N = 208
Frequency
30
20
10
0
0
5
10
15
20
25
No_of_Sessions

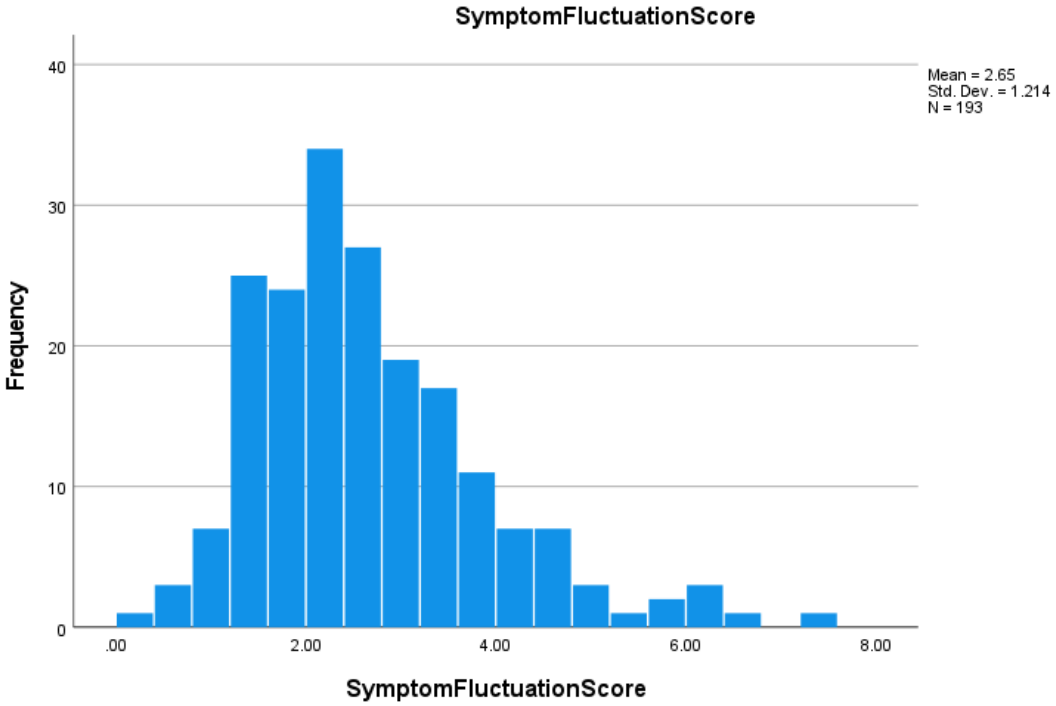
***Altered Criteria***

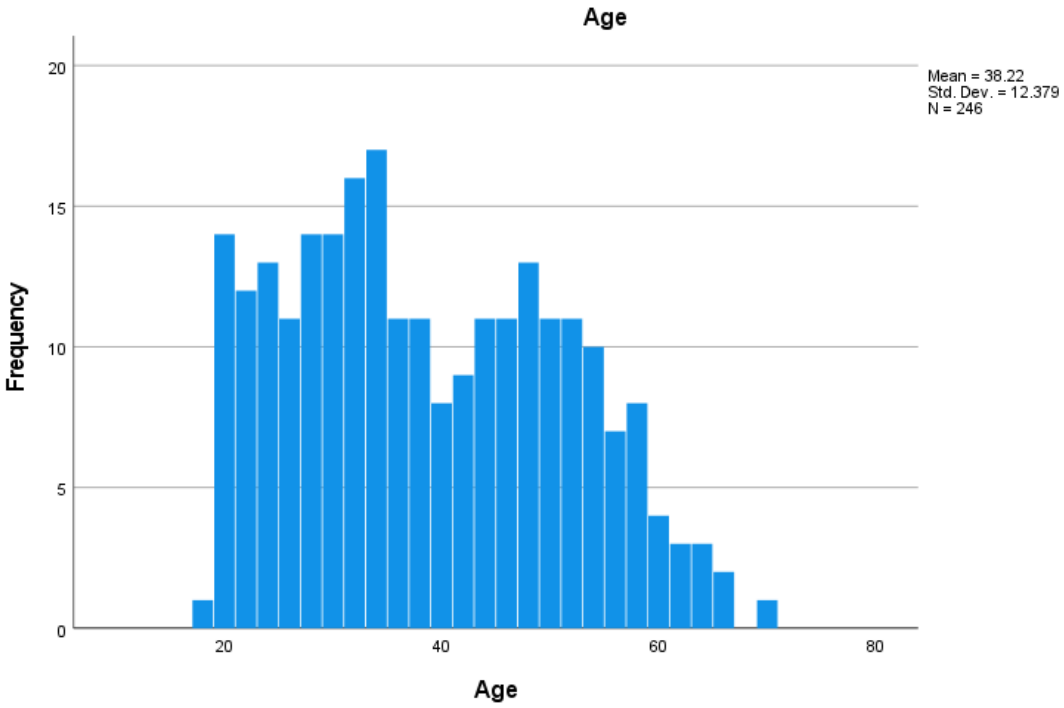
**Statistics**

| | | PHQ9Session1Score | PHQ9EndofTherapyScore | SymptomFluctuationScore | Age |
|---|---|---|---|---|---|
| N | Valid | 246 | 246 | 193 | 246 |
| | Missing | 0 | 0 | 53 | 0 |
| Skewness | | -.467 | .581 | 1.065 | .277 |
| Std. Error of Skewness | | .155 | .155 | .175 | .155 |
| Kurtosis | | -.070 | -.382 | 1.508 | -.939 |
| Std. Error of Kurtosis | | .309 | .309 | .348 | .309 |




PHQ9EndofTherapyScore
Mean = 9.38
Std. Dev. = 6.211
N = 246
Frequency
PHQ9EndofTherapyScore


SymptomFluctuationScore
Mean = 2.65
Std. Dev. = 1.214
N = 193
Frequency
SymptomFluctuationScore


Age
Mean = 38.22
Std. Dev. = 12.379
N = 246
Frequency
20
15
10
5
0
20
40
60
80
Age

**Appendix G- Assumption of Multicollinearity**

***Logistic regression- Original and Altered Criteria***

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. | Tolerance | VIF |
| 1 | (Constant) | .015 | .148 | | .099 | .921 | | |
| | PHQSession1Score | .015 | .007 | .170 | 2.362 | .019 | .986 | 1.014 |
| | SymptomFluctuationScore | .021 | .025 | .061 | .852 | .395 | .985 | 1.015 |
| | Employment_Condensed | -.042 | .025 | -.119 | -1.665 | .098 | .999 | 1.001 |

a. Dependent Variable: SG_Tang

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. | Tolerance | VIF |
| 1 | (Constant) | .621 | .148 | | 4.192 | <.001 | | |
| | PHQ9Session1Score | -.020 | .007 | -.195 | -2.774 | .006 | 1.000 | 1.000 |
| | SymptomFluctuationScore | .069 | .029 | .167 | 2.378 | .018 | 1.000 | 1.000 |

a. Dependent Variable: SG_Kelly

## *Linear regression- Original and Altered Criteria*

**Excluded Variables[a]**

| Model | | Beta In | t | Sig. | Partial Correlation | Collinearity Statistics Tolerance |
|---|---|---|---|---|---|---|
| 1 | SG_Tang | -.272[b] | -4.032 | <.001 | -.292 | .976 |
| | SymptomFluctuationScore | -.237[b] | -3.487 | <.001 | -.255 | .983 |
| | Gender | .140[b] | 2.033 | .044 | .152 | 1.000 |
| 2 | SymptomFluctuationScore | -.217[c] | -3.313 | .001 | -.244 | .978 |
| | Gender | .148[c] | 2.239 | .026 | .167 | .999 |
| 3 | Gender | .131[d] | 2.031 | .044 | .153 | .993 |

a. Dependent Variable: PHQEndofTherapyScore

b. Predictors in the Model: (Constant), PHQSession1Score

c. Predictors in the Model: (Constant), PHQSession1Score, SG_Tang

d. Predictors in the Model: (Constant), PHQSession1Score, SG_Tang, SymptomFluctuationScore

**Excluded Variables[a]**

| Model | | Beta In | t | Sig. | Partial Correlation | Collinearity Statistics Tolerance |
|---|---|---|---|---|---|---|
| 1 | SymptomFluctuationScore | -.294[b] | -4.588 | <.001 | -.317 | .999 |
| | Ethnicity_Condensed | .178[b] | 2.690 | .008 | .192 | .999 |
| | SG_Kelly | -.009[b] | -.127 | .899 | -.009 | .953 |
| 2 | Ethnicity_Condensed | .172[c] | 2.729 | .007 | .195 | .999 |
| | SG_Kelly | .047[c] | .704 | .483 | .051 | .922 |
| 3 | SG_Kelly | .043[d] | .651 | .516 | .048 | .922 |

a. Dependent Variable: PHQ9EndofTherapyScore

b. Predictors in the Model: (Constant), PHQ9Session1Score

c. Predictors in the Model: (Constant), PHQ9Session1Score, SymptomFluctuationScore

d. Predictors in the Model: (Constant), PHQ9Session1Score, SymptomFluctuationScore, Ethnicity_Condensed

**Appendix H- Linearity of the Logit**

***Original Criteria***

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | PHQSession1Score | .705 | .913 | .596 | 1 | .440 | 2.024 |
| | SymptomFluctuation Score | -.708 | .923 | .588 | 1 | .443 | .493 |
| | Employment_Condensed | .281 | .395 | .508 | 1 | .476 | 1.325 |
| | Ln(PHQ1) by PHQSession1Score | -.161 | .240 | .453 | 1 | .501 | .851 |
| | Ln(INTRAINDIVIDUAL VARIABILITY) by SymptomFluctuation Score | .414 | .437 | .897 | 1 | .343 | 1.513 |
| | Employment_Condensed by Ln(Employ) | -.643 | .482 | 1.781 | 1 | .182 | .525 |
| | Constant | -4.257 | 4.102 | 1.077 | 1 | .299 | .014 |

a. Variable(s) entered on step 1: PHQSession1Score, SymptomFluctuationScore, Employment_Condensed, Ln(PHQ1) * PHQSession1Score , Ln(INTRAINDIVIDUAL VARIABILITY) * SymptomFluctuationScore , Employment_Condensed * Ln(Employ) .

***Altered Criteria***

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | PHQ9Session1Score | -.056 | .667 | .007 | 1 | .933 | .946 |
| | SymptomFluctuation Score | 1.936 | .954 | 4.116 | 1 | .042 | 6.928 |
| | LnPHQ1 by PHQ9Session1Score | -.012 | .177 | .004 | 1 | .947 | .988 |
| | LnIV by SymptomFluctuation Score | -.764 | .442 | 2.989 | 1 | .084 | .466 |
| | Constant | -1.571 | 3.006 | .273 | 1 | .601 | .208 |

a. Variable(s) entered on step 1: PHQ9Session1Score, SymptomFluctuationScore, LnPHQ1 * PHQ9Session1Score , LnIV * SymptomFluctuationScore .