

## **Extending the Use of Routine Outcome Monitoring: Predicting Long-term Outcomes in Cognitive Behavioral Therapy for Severe Health Anxiety**

**Objective:** Routine outcome monitoring (ROM) is a well-evidenced means of improving psychotherapy's effectiveness. However, it is unclear how meaningful ROM is for problems that span physical and mental health, such as severe health anxiety. Physical and mental health comorbidities are common amongst severe health anxiety sufferers and cognitive behavioral therapy (CBT) is a recommended treatment.

**Method:** Seventy-nine participants received CBT for severe health anxiety in a clinical trial. The Outcome Rating Scale (ORS: a ROM assessment of wellbeing) was completed at each session. Multilevel modelling assessed whether last-session ORS predicted health anxiety and other outcomes over 12-month follow-up. Similar models were developed using health anxiety as a comparative outcome-predictor. Outcome-improvements of treatment-responders with sudden gains were compared to those of non-sudden-gainers. **Results:** Last-session ORS scores predicted all outcomes up to 12 months later, with a comparable predictive effect to health anxiety. Sudden-gainers on the ORS reported significantly greater improvement in depression, functioning, and wellbeing, but no difference in health anxiety or other measures. **Conclusion:** The ORS may be a feasible, overall estimate of health, functioning, and quality of life in psychotherapy for severe health anxiety. Sudden gains on the ORS may be clinically meaningful with respect to some long-term outcomes.

**Keywords:** Routine outcome monitoring; outcome feedback; feedback-informed treatment; cognitive behavior therapy; health anxiety; comorbidity.

**Word count:** Abstract: 200; Main text: 6263

Routine outcome monitoring is an umbrella term used to describe methods in which an outcome assessment is given, typically at every session of psychotherapy, and resultant scores give therapists algorithm-based predictions of treatment outcome (Lambert, 2010).

This aims either to alert therapists to patients who are at risk of having a poor outcome or to help patients and therapists collaborate to enhance improvement. With appropriate training and supervision, psychotherapists can tailor therapy to the outcome trajectory of individual patients. As such, ROM becomes more than an assessment method and can be seen as an

intervention in its own right. Current evidence from Randomized Controlled Trials (RCTs) and subsequent meta-analyses suggest that ROM can have beneficial effects for both treatment efficacy and efficiency (Delgadillo et al., 2018; Lambert, Whipple, & Kleinstäuber, 2018; Shimokawa, Lambert, & Smart, 2010). The two best-evidenced ROM tools use general, overall assessments of mental health and wellbeing: The Outcome Questionnaire 45 items (OQ-45; Lambert, 2012) and the Outcome Rating Scale (ORS; Miller, Duncan, Sorrell, & Brown, 2005). Both systems aim to give an overview outcome assessment which can be used across conditions, identifying the “mental health vital signs” (Lambert et al., 2018, p. 520). They are designed for mixed, general populations and have typically been evaluated in routine practice settings.

The ORS is the central outcome measure in the Partners for Change Outcome Measurement System (PCOMS; Miller et al., 2005). A more nuanced understanding of PCOMS effects has emerged from recent meta-analyses (Lambert et al., 2018; Østergård, Randa, & Hougaard, 2018). The PCOMS system has been shown to have beneficial effects on outcome as compared to usual care and reduces the number of patients showing no improvement at the end of treatment. However, it does not reduce the number of patients dropping out or deteriorating. Furthermore, Østergård et al. (2018) found that PCOMS only had a beneficial treatment effect when outcome was measured with the ORS, but not when independent measures were used. The same was found in three studies using independent outcome measures in the nine-study sample investigated by Lambert, Whipple, and Kleinstäuber (in-press). This illustrates a central unaddressed issue within ROM research: The unclear relationship between ROM outcomes and other important independent outcomes.

### **Unaddressed issues in ROM**

The case is presented specifically for PCOMS, but in many RCTs assessing the effectiveness of ROM, there have been criticisms for using the ROM measure as the primary outcome

(Wampold, 2015). This means that the ORS is frequently used both as the ROM intervention and the outcome of the ROM intervention, meaning one cannot tell the independent value of ROM interventions. Even when independent assessments are used, they are typically general in nature with few studies including assessment of quality of life, functioning, or disorder-specific outcomes (Kendrick et al., 2014). Furthermore, ROM RCTs and meta-analyses have almost all been directly or indirectly conducted by researchers with an allegiance to the method investigated, which tends to inflate effectiveness (Luborsky et al., 1999). This is supported by evidence that RCTs carried out more independently have sometimes had less positive results (e.g. Rise, Eriksen, Grimstad, & Steinsbekk, 2016). Another element commonly missing from ROM research is independently-assessed, longer-term outcome follow-up (Kendrick et al., 2014). As ROM studies often use outcome data collected by therapists during treatment, there are fewer studies where outcome data are collected by independent or blinded assessors at fixed intervals after treatment (Østergård et al., 2018). The absence of independent longer-term follow-up or blinding is another factor likely to inflate the treatment effect, due to both assessor bias and the potential for treatment effects to diminish after completion (Wood et al., 2008).

### **Comorbidity and ROM**

From this type of evidence, it is difficult to assess the independent predictive ability of ROM measures on treatment outcomes for specific disorders, which usually lack relevant assessments and are typically conducted by those with an allegiance. This situation is especially problematic in mental health problems that commonly co-occur with physical health problems, as key outcomes would not be assessed by ROM measures. Comorbidity of this type leads to poorer health outcomes for patients and higher costs for services (Barnett et al., 2012; Naylor et al., 2016). Demonstrating that a commonly-used ROM tool is predictive of more specific independently-assessed physical and mental health outcomes would expand

our understanding of the value such ROM assessments hold. However, it is unclear whether the most widely-used ROM measures (such as the ORS) can effectively assess the types of outcome that are important where comorbid physical and mental health problems are present. Therefore, patients who appear to be improving on ROM measures may not improve in important unassessed areas to which the clinician is blind. Alternatively, the tailored care which is integrated within ROM systems may make ROM meaningful even for problems with physical and mental health sequelae, such as severe health anxiety.

### **Severe health anxiety and comorbidities**

Cognitive behavioral therapy (CBT) has proven effective for mental health problems that are frequently comorbid with chronic physical health problems; severe health anxiety being one of the most common examples (e.g. Tyrer et al., 2017). Severe health anxiety can be defined as an anxious preoccupation with having or contracting a serious health problem, which is not resolved by medical reassurance (Fink et al., 2004). It affects approximately 5.7% of the population across a lifetime and is associated with multimorbid, chronic physical and mental health problems (Patel et al., 2015; Sunderland, Newby, & Andrews, 2013). Symptoms of severe health anxiety are characterized by an interplay between physical and psychological difficulties. For example, persistent bodily checking is an identified factor maintaining severe health anxiety, wherein fear of bodily symptoms (such as pain) leads to recurrent body checking. However, repeatedly checking the body exacerbates symptoms such as pain; thus, pain is worsened and anxiety consequently maintained (Salkovskis & Warwick, 1986). Given this interaction, if severe health anxiety were to improve, one might expect concurrent improvement in general health too (Morriss et al., 2019). Yet this type of change may not be picked up in routine practice using ROM. Conversely, ROM may give a false indication of improvement when core symptoms remain. Therefore, it is unclear whether ROM tools can make a meaningful contribution to psychotherapy outcome assessment of specific disorders,

such as health anxiety, which often involve comorbid physical and mental health problems. Independent assessment of a range of outcomes over time would give a clearer understanding of the value ROM may hold for specific disorders, especially comorbid physical and mental health disorders.

### **Sudden gains and ROM**

A similar situation is apparent with sudden gains in ROM: Sudden gains have been identified from ROM data, but it is unclear whether they translate into improvements on other outcomes (Lambert, 2010). Sudden gains are defined as large, sustained, clinically meaningful improvements in symptoms, typically occurring early in psychological treatment (Tang & DeRubeis, 1999). Those experiencing sudden gains often have better long-term outcomes than those without (Aderka, Nickerson, Bøe, & Hofmann, 2012). This has been most clearly evidenced in CBT for depression (Tang, DeRubeis, Beberman, & Pham, 2005; Tang, DeRubeis, Hollon, Amsterdam, & Shelton, 2007). Beneficial effects of sudden gains have also been evidenced in severe health anxiety, but eligibility criteria restricted the likelihood of comorbidity and no outcome assessments were taken for other disorders (Hedman et al., 2014). Little crossover research has been conducted between sudden gains in disorder-specific outcomes and those occurring on ROM tools. As such, it is unclear whether sudden gains on ROM tools have any impact on long-term disorder-specific outcomes. Moreover, sudden gains have not been studied in populations with high rates of comorbidity. Thus, it is unclear whether the phenomenon still has clinical significance where multiple problems are present or what outcome measures are best for assessing sudden gains in this context.

This study assessed the predictive validity of a ROM assessment (the ORS) on long-term health anxiety severity measured independently up to 12 months after CBT for high service utilizers with severe health anxiety (Morriss et al., 2019). The study also assessed the predictive value of ROM on other anticipated effects of severe health anxiety: overall health,

functional impairment, quality of life, somatic symptoms, generalized anxiety, and depression. Lastly, this study assessed whether sudden gains in ROM have a significant impact on long-term trajectories of health anxiety and the other outcomes.

## **Method**

### *Design and participants*

The predictive value of the ORS for a range of independently-assessed health outcomes over time was assessed by an observational design. The study applied a secondary analysis to data from a single-blind, two-armed RCT evaluating the effectiveness of remotely delivered CBT for high service utilizers with severe health anxiety compared to treatment as usual (TAU) (Morriss et al., 2019). In the original RCT, 156 adult participants with severe health anxiety attending two or more urgent care appointments over the previous 12 months were equally randomized to CBT or TAU. Participants were recruited by treating healthcare professionals from primary and general hospital care (including emergency care, endocrinology, and neurology departments) with most participants seeking physical (rather than mental) health treatment. All 79<sup>1</sup> participants receiving CBT were offered weekly CBT adapted to severe health anxiety over video conferencing or the telephone. All those attending sessions were offered the use of ROM by their therapists with the ORS completed at each session. In parallel, a series of physical and mental health outcome assessments were taken at baseline, 3-, 6-, 9-, and 12-month follow-up by independent researchers blinded to treatment group. The study showed that, at 12-month follow-up, those offered CBT had significantly greater reductions in health anxiety, generalized anxiety, and depression, and greater improvements

---

<sup>1</sup> There were 78 participants randomized to CBT, and one participant was randomized to usual care but offered CBT in error: Their data is included in the analysis; therefore, the total sample is 79.

in general health. Furthermore, the intervention resulted in a net monetary benefit of £3,164 over usual care, due to the reduction in healthcare-use over the year amongst CBT recipients.

Fifty-six participants (71%) in the treatment arm attended two or more sessions of CBT and completed the ORS. All were included in the first stage of this study, to assess whether the ORS score at the final session predicted follow-up health outcomes. The predictive validity of the ORS was directly compared to 6-month follow-up assessment of health anxiety as a predictor of all secondary outcomes. The 57 participants who had responded to treatment at follow-up were included in a sensitivity assessment: Those who responded to treatment and experienced a sudden gain on the ORS were compared to the remaining treatment responders to see if sudden gains in routinely-monitored outcomes conferred a long-term clinical benefit.

### *Instruments*

#### *Routine Outcome Monitoring*

The ORS and associated PCOMS is one of the most widely used ROM systems, with approximately 45,000 therapists registered across 20 countries (Østergård et al., 2018). It is designed for use in all healthcare settings but has predominantly been evidenced in routine care psychotherapy populations (Lambert et al., 2018). It is designed with the primary aim of improving overall effectiveness for all patients by improving patient engagement in the treatment. No researchers in the current study or original RCT have any allegiance to PCOMS or other ROM systems.

The ORS uses four self-reported, brief visual analogue scales, which aim to assess the central targets of psychological treatment: individual wellbeing, change in relationships, social functioning, and overall wellbeing (Miller, Duncan, Brown, Sparks, & Claud, 2003). In normal use of the ORS, each scale is rated 0 to 10 and the scores are totaled out of 40.

Increasing scores indicate improved wellbeing. Scores below the clinical cut-off of 25 fall within the established clinical range (Prescott, Maeschalck, & Miller, 2017). The ORS has shown excellent internal consistency ( $\alpha = .93$ ); adequate test-retest reliability ( $r_s = .49 - .66$ ) and satisfactory convergent validity with the OQ-45 ( $r_s = .56 - .69$ ) (Miller et al., 2003).

### *Independent Assessments*

Due to high rates of comorbidity amongst those with severe health anxiety, a range of outcomes were assessed independently:

- Short form 14-item Health Anxiety Inventory (SHAI) (Salkovskis, Rimes, Warwick, & Clark, 2002).
- 7-item Generalized Anxiety Disorder assessing generalized anxiety (GAD-7) (Spitzer, Kroenke, Williams, & Löwe, 2006).
- 15-item Patient Health Questionnaire assessing somatic distress (PHQ-15) (Kroenke, Spitzer, & Williams, 2002).
- 9-item Patient Health Questionnaire assessing symptoms of depression (PHQ-9) (Kroenke, Spitzer, & Williams, 2001).
- 8-item Work and Social Adjustment Scale (WSAS) assessing functional impairment (Mundt, Marks, Shear, & Greist, 2002).
- 5-item 5-level EuroQol assessing quality of life (EQ5D-5L; Herdman et al., 2011).
- Visual analogue scale (VAS, Herdman et al., 2011) assessing general health.

The structured clinical interview for the Diagnostic and Statistical Manual of Mental Disorders IV (SCID; First, Spitzer, Gibbon, & Williams, 1998) was also administered at baseline. Long-term physical health conditions were recorded from baseline patient interviews.

The approach used mitigates common methods variance in some ways when associating the ORS with other outcomes. In most outcome prediction studies, both the

predictor and outcome measurements are completed with the same assessors, which is likely to inflate associations (Campbell & Fiske, 1959), whereas in the current study, the ORS uses a visual analogue scale completed with therapists to compare with questionnaire-based measures (with the exception of the VAS) completed with independent researchers. This reduces the likelihood of inflated associations and type-I errors.

### ***Procedure***

#### *Remote CBT*

Remote CBT was delivered via video conferencing or over the telephone if internet access was limited. A team of four therapists delivered treatment according to an established treatment protocol (Tyrer et al., 2017). Six to 12 CBT sessions were offered based on treatment response, with the option of up to three booster sessions (a median eight were attended by participants). This included time at the beginning of treatment designated to familiarizing patients with the delivery methods to be used and addressing any related concerns they might have about this. Patients were sent webcams and audio headsets if required to facilitate the use of remote technology. Patients were also sent a written guide for use of remote technology and a guide explaining how CBT would be used through remote technology to address health anxiety. Treatment used CBT principles to assess and test beliefs about health, illness, and associated issues that were likely to be causing distress. Behavioural strategies known to maintain severe health anxiety were identified and collaboratively reduced or stopped. Typically, treatment order included listing and prioritizing health fears; establishing the meaning of these fears to the individual; then collaboratively formulating ways in which these fears might be maintained by specific behavioral and cognitive strategies; experiments were then used to evaluate the impact of

changing these strategies (e.g. reducing bodily checking); changes in beliefs and behavior were reevaluated as a result.

#### *Administration of the ORS*

The ORS was completed by patients with their therapist at every session of CBT, unless patients refused or did not attend. Therapists used the ORS to tailor treatment by focusing on key problems of importance to the patient in each domain and setting goals associated with points on the scales. For example, if a patient highlighted social avoidance as a key problem that was reducing the score on the social functioning scale to 3, the therapist might establish what would be different if they were functioning 'at the level of a 4'. They could then work towards what the patient describes as a short-term goal. The ORS was completed on software which gave a computerized traffic light system ([www.myoutcomesukapp.com](http://www.myoutcomesukapp.com)), alerting therapists to patients who were at risk of non-improvement or deterioration, from algorithm predicted outcome trajectories (Miller et al., 2005). If patients were identified as at risk of a poor outcome, therapists initiated conversations to help identify unaddressed problems and collaboratively adjust the course of therapy. All trial therapists used ORS data to identify patients on their caseload at risk of poor outcome. They then attended monthly supervision focused on addressing issues with these cases in particular, using observation of recordings from previous sessions to guide decision-making.

#### *Method of analysis*

The baseline characteristics of included and excluded patients were compared using Mann-Whitney U tests for continuous variables and chi-squared for categorical variables. Pearson's correlation coefficients were calculated for associations among baseline outcome scores and Spearman's rho was used to calculate correlations among post-treatment scores (last session score for the ORS and 6-month follow-up for all other measures).

*Predictive modelling*

Due to the repeated measurement design, multilevel modelling was used to assess whether ORS total scores at the last session separately predicted each health outcome individually. Data were structured in a two-level model with repeated follow-up assessment time-point (“*time*” hereafter) at level 1, nested within patients at level 2. A preliminarily-tested third (“*therapist*”) level was not significant on any outcomes. Unconditional models were used to assess clustering within patients and linearity of trends before adding explanatory variables. Model fit was explored using -2 log likelihood and parameter estimation was made using maximum likelihood. Each parameter was added to the model in a stepped manner. As each parameter was added, change in -2 log likelihood and change in degrees of freedom were used to calculate whether a significant reduction in model deviance was caused using chi-squared. Table 1 illustrates the methods used for model development where health anxiety is the outcome.

[Table 1]

Fully adjusted models included the following fixed predictor variables for each outcome: *time*; *baseline outcome score* (controlling for baseline outcome severity); *first session ORS score* (controlling for initial ORS severity); *total number of mental and physical health diagnoses* (controlling for comorbidity); and *last session ORS score* (the main predictor of interest). Random intercepts for *time* were included in all models, but random slopes were not. This was either due to non-significance in fully adjusted models or non-significant change in -2 log likelihood when slope was added (using chi-squared). Below is an illustration of the equation tested using health anxiety:

$$y_{ij} = \beta_{0i} + \beta_1 \text{time}_{ij} + \beta_2 \text{Baseline SHAI}_{ij} + \beta_3 \text{First session ORS}_{ij} + \beta_4 \text{Total disorders}_{ij} + \beta_5 \text{Last session ORS}_{ij} + e_{ij} \quad (1)$$

$$\beta_{0i} = \beta_0 + \mu_i \quad (2)$$

Where  $y_{ij}$  is the SHAI score for the  $j$ th participant at the  $i$ th follow-up time point. As such  $j$  is participants (level 2 unit) and  $i$  is follow-up time order (level 1 within level 2 unit).  $\beta_{0i}$  indicates that estimates of the intercept for both the overall model and variability of level 1 intercepts were calculated.  $\beta_1$  through  $\beta_5$  were fixed level 2 effects, adjusting for differences between participants.  $e_{ij}$  is the residual or unexplained variance in the model.

The same methods were used to build models with 6-month health anxiety (SHAI) as the focal predictor-of-interest for each secondary outcome in turn. This is illustrated below, where  $y_{ij}$  is the generalized anxiety (GAD-7) score for the  $j$ th participant at the  $i$ th follow-up time:

$$y_{ij} = \beta_{0i} + \beta_1 \text{time}_{ij} + \beta_2 \text{Baseline GAD-7}_{ij} + \beta_3 \text{Baseline SHAI}_{ij} + \beta_4 \text{Total diagnoses}_{ij} + \beta_5 \text{6-month SHAI}_{ij} + e_{ij} \quad (3)$$

$$\beta_{0i} = \beta_0 + \mu_i \quad (4)$$

Fully adjusted models included the same fixed predictor variables for each outcome with SHAI in exchange for ORS: *time; baseline outcome score; baseline SHAI; total diagnoses; SHAI at 6 months* (the focal predictor). Random intercepts, but not slopes, were again included in all models.

Parallel models were constructed to compare the predictive validity of the ORS versus SHAI – facilitating evaluation of the predictive ability of the ORS against that of the primary treatment target and outcome variable of the trial (SHAI). This comparison was conducted using a pseudo- $R^2$ , which gives a measure of effect size in multilevel modelling, enabling a direct comparison between the ORS and SHAI (Raudenbush & Bryk, 2002). Using the model developed in Table 1 as an illustration, the pseudo- $R^2$  was calculated as the percentage of:  $((\sigma_5^2 - \sigma_6^2) / \sigma_1^2)$ , where  $\sigma_1^2$  is the unconditional model intercept variance (step 1),  $\sigma_5^2$  is intercept variance in the fully adjusted model *without* the focal predictor (step 5) and  $\sigma_6^2$  is the fully adjusted model's intercept variance (step 6). This shows the percentage reduction in

unexplained variance between participants when post treatment SHAI/last-session ORS are added to fully adjusted models (“full” below) (Hox, Moerbeek, & Van de Schoot, 2017):

$$Pseudo R^2 = \frac{(variance\ full\ without\ target) - (variance\ full\ with\ target)}{variance\ unconditional} \quad (5)$$

Analysis was conducted using SPSS 24.

### *Sudden gains modelling*

An adaptation of Tang and DeRubeis’ (1999) definition of sudden gains was applied to ORS scores, assessing large intersessional gains between session N and N + 1 if the following criteria were met:

1. Sudden gains should be large in absolute terms – specifically, the N + 1 gain needed to be at least 5 points, the established reliable change index on the ORS.
2. Sudden gains should be large in relative terms – specifically, the N + 1 gain needed to be at least a 25% improvement on session N.
3. Sudden gains should be sustained beyond random fluctuations – specifically, the mean ORS score of the three sessions prior to the gain should be at least a reliable change (5 points) lower than the mean ORS score for the three sessions following the sudden gain. This criterion was adapted to include sudden gains at the very beginning or end of treatment given that treatment in the current study was much briefer than in the study from which the criteria originate (16-20 versus 6-12 sessions). Indeed, only a sudden gain at session 3 could be identified if the minimum number of sessions were attended. This is particularly problematic when sudden gains identified at the first session have been found to positively influence outcome in their own right (Busch, Kanter, Landes, & Kohlenberg, 2006). Therefore, fewer than three sessions prior to the gain were allowed as long as the gain was sustained for the three ensuing sessions. Conversely, fewer than three sessions were allowed after a gain assuming

that the pre-gain score was stable for the previous three sessions. This aimed to include very early or late gains which would otherwise be missed.

In order to assess whether these adapted criteria were of clinical value, results were compared to those achieved using the more strict criteria from the original method: To be identified as a sudden gain, the mean ORS score of the three session scores prior to the gain were required to be significantly lower than the three sessions after the gain using an independent t-test, with an alpha of .05. Two sessions before and after were used if there were no more than two sessions before or after the gain, otherwise three were required.

Treatment responders who experienced sudden gains were compared to treatment responders who did not experience sudden gains. Treatment responders were defined as those experiencing  $\geq 4$  point reduction in health anxiety score at follow-up, as this is deemed the minimum clinically important change index on the SHAI (Clark & Oates, 2014). Multilevel modelling was again used to account for the hierarchical structure of level 1 *time* clustered within level 2 *patients*. Similar methods of model development as previously described were used. Random intercepts and fixed slopes for *time* were included in all models. Four fixed predictors were entered as covariates into models for each model: *time*, *baseline outcome score*, *total diagnoses* and *sudden gains* (a binary variable assessing the difference in growth curves over time of those identified as sudden gainers versus remaining treatment responders). As well as independently-assessed outcomes over 12-month follow-up, a model was developed which used session-by-session ORS scores as the ordered *time* parameter. This model evaluated whether sudden gainers had significantly improved ORS trajectories during treatment sessions.

Ethical approval was obtained from the National Research Ethics Service, London-Riverside Committee (reference 14/LO/1102).

## Results

### *Sample characteristics*

There was a greater proportion of females included in the assessment of the ORS' predictive validity than amongst those without available data ( $\chi^2 = 3.95, p = .047$ ). No other demographic or clinical characteristics were significantly different at baseline (Table 2). Average baseline scores in the overall sample fell within the clinical range on all assessments, including the ORS. Quality of life and general health were both lower than average. Diagnostic self-report assessments indicated that comorbid physical and mental health problems were very common in the sample, with a mean of 6 (range: 0–16) mental disorders and 1 (range: 0–3) chronic physical condition at baseline.

### *Outcome Correlations*

Significant weak-to-moderate correlations were observed between first-session ORS scores and baseline scores for other outcome measures in the expected directions, except for a lack of correlation with baseline health anxiety (absolute  $r_s = .055 - .502$ ; Table 3). Weak-to-moderate correlation was found between baseline health anxiety (SHAI) and generalized anxiety (GAD-7), but no correlations were observed with any other outcomes. No correlations between last-session ORS and other post-therapy outcome measures were of more than moderate magnitude (absolute  $r_{s,s} = .310 - .446$ ; Table 4). A similar pattern of significant weak-to-moderate correlations were observed between 6-month follow-up health anxiety and other post-treatment outcomes. As last-session ORS scores were not highly correlated with independently-assessed outcomes at pre- or post-therapy assessment-points, the ORS does not seem to be acting as a simple proxy for other outcome scores, so investigation of the predictive relationship is warranted.

[ Table 3 and 4]

***Predictive validity of the ORS***

In unconditional models for all outcomes 59% to 79% of the covariance was explained by clustering within patients, indicating that a two-level hierarchical structure was appropriate to account for this effect. Last-session ORS scores significantly predicted health anxiety over time. This meant that improvement on the ORS (higher scores) predicted lower severity of health anxiety (lower scores). The same was true of depression, generalized anxiety, somatic symptoms, and functional impairment. Significant positive associations were found between last-session ORS scores and quality of life. Therefore, improvement on the ORS predicted improvement in quality of life and the same was true of general health. All models controlled for baseline ORS, baseline outcome score, number of comorbidities, and changes over time (Table 5).

[Table 5]

***Comparison between the ORS and SHAI***

The same models were run with the SHAI as focal predictor, which demonstrated that it was also a significant predictor of all secondary outcomes over time when controlling for baseline severity. The percentage reduction in unexplained between-participant variance uniquely accounted for by the ORS and SHAI were compared for each model, to give an estimate of comparative effect size (Table 6). This analysis indicated that a similar proportion of outcome variance was explained by the ORS and the SHAI across outcomes (average 19% and 18% respectively). There were trends of larger predictive effects where there were large pre-post treatment effect sizes (with health anxiety, generalised anxiety and depression) compared to other outcomes showing smaller treatment effects.

[Table 6]

### *Sudden gains*

Thirty-two (56%) of the 57 treatment responders (72% of total treatment group) were identified as sudden gainers using the adapted criteria. Session 2 was the median session for sudden gains to occur (IQR 1–3). The average sudden gain led to a 53% increase in ORS score and accounted for 80% of the average sudden gainer's pre- to post-therapy improvement (Figure 1). Multilevel modelling was used to establish whether there was a significant difference in outcomes over time for sudden gainers versus non-sudden gaining treatment responders. This analysis showed that sudden gainers had significantly greater improvement in wellbeing over the course of therapy ( $B = 4.38$ ,  $SE B = 1.36$ ,  $p = .002$ ). Sudden gainers also reported significantly greater improvements in functional impairment (WSAS;  $B = -4.38$ ,  $SE B = 1.84$ ,  $p = .021$ ) and symptoms of depression (PHQ-9;  $B = -2.54$ ,  $SE B = 1.20$ ,  $p = .039$ ) over the 12-month follow-up period. There were trends for greater improvements across all outcomes, though no others reached significance. Figure 2 shows the percentage improvement achieved by sudden gainers as compared to other treatment responders using baseline scores as the percentage denominator. All models controlled for baseline ORS score, baseline outcome score, number of comorbidities and repeated measurement effects.

Using the strict original criteria for classifying sudden gains, 9 (16%) sudden gainers were identified and no differences were observed on any outcome (all  $ps \geq .320$ ). The allowance for gains at the earliest and latest points in therapy accounted for the vast majority of those not identified by the strict criteria. In the analysis using the adapted criteria, 22 sudden gains were identified before the third session and two further sudden gains at the final session.

## Discussion

### *Predictive validity of the ORS*

This study used longitudinal, independent, single-blinded assessments to evaluate the predictive validity of ORS ratings on a range of health outcomes from CBT for health anxiety. Findings suggest that ORS scores at the last treatment session predict a range of longer-term physical and mental health outcomes, irrespective of how many sessions were attended. Furthermore, the multilevel modelling design means that this relationship is less likely to be due to covariance across time, an artefact of baseline problem severity, or related to the number of comorbidities. This is a significant expansion in the use of the ORS as a common ROM outcome measure, because ORS outcome studies are often clinician-assessed with short follow-up periods. The assessments used in previous ORS studies have also been limited in what they can say about the relevance of ROM for specific disorders or its impact on general health (rather than just mental health) and quality of life.

This study indicates that those ending CBT for health anxiety with higher ORS scores relative to their baseline score are more likely to sustain long-term improvement in a range of health domains. Conversely, those who end treatment with lower ORS scores are unlikely to spontaneously improve further down the line. Therefore, therapists and other members of relevant patients' care-teams should be alert to the need for alternative interventions if psychotherapy offered was either not accepted or not effective. Consequently, ORS scores at the last treatment session may offer a broad, brief assessment of long-term health and quality of life after treatment. Given that the source data came from remote delivery of CBT via video conferencing or the telephone, this suggests that ROM may be a helpful predictor of outcome when using this modality.

Comparable post-treatment predictive abilities were achieved for the ORS and health anxiety assessments, which may be explained by an overlap in the construct of general

psychological distress. As health anxiety was both the primary outcome and the main treatment target for CBT interventions, this suggests that the more general ORS can have a cross-cutting predictive effect that is similar to more specific measures. The findings imply that a brief, general ROM assessment like the ORS may be useful to services if choosing between general and specific ROM measures. A general ROM assessment would also avoid the need to select disorder-specific measures for each patient. Importantly, and novel to ROM research, predictive validity was demonstrated across a wide range of outcomes – encompassing mental health, physical health, quality of life and functional impairment. Therefore, cross-domain outcome prognosis could be estimated, which could be particularly important for services offering psychological therapies in physical healthcare settings. Nonetheless, current evidence suggests that disorder-specific feedback can enhance outcomes (Clark et al., 2018; Nugter et al., 2017). Therefore, given the similar predictive validity and the established benefit of disorder-specific feedback, this paper supports the use of both general and specific ROM.

### ***Predictive validity of sudden gains***

Amongst treatment responders, sudden gains on the ORS were associated with greater improvements compared with non-sudden gainers. There were significantly greater improvements in ORS-rated wellbeing across therapy sessions and significant improvements in depression and functioning across 12-month follow-up. These findings add to the extensive literature emphasizing the importance of sudden gains for long-term benefits of psychotherapy (Aderka et al., 2012). This study also suggests that session-by-session ORS changes are somewhat sensitive to specific clinical changes in symptomatology – given that rapid, large, sustained ORS improvements correspond with greater long-term improvement in clinically important domains.

The inclusion of sudden gains from the first session using the ORS is an important adaptation for brief psychological treatments such as CBT for health anxiety, as compared to longer-treatments upon which established criteria have been based. It carries the risk of including sudden gains that are more likely to have occurred due to random fluctuations in score. However, in the current sample, the adapted criteria were more clinically meaningful than criteria closer to the original. This may be because fewer sudden gainers were identified using the original criteria than has been typical in previous studies. The ORS could be more susceptible to random fluctuations in score, which would make it less likely that a statistically significant pre-post gain difference would be identified. However, the inclusion of very early and very late sudden gains seemed to be the most influential factor in identifying the majority of sudden gainers excluded by the strict criteria. The initial sessions of psychological treatment are often considered to be primarily focused on assessment only, but this study and other sudden gains research suggest that clinically significant changes can occur very early (Busch et al., 2006). Within the treatment offered in the current study, only ‘problem identification and prioritization’ is intended to be covered in the first session, yet this contributed to significant improvements for some participants. This may be to some degree unique to this study, as most patients were unfamiliar with the concept of health anxiety at referral and so psychoeducation around this issue might have been an important therapeutic step.

No significant difference was identified between sudden gainers and the remainder of the sample in severity of health anxiety, generalized anxiety, general health, somatic symptoms, or quality of life – despite trends of greater improvement. Further research is required to establish the meaning of these results. It is possible that ORS sudden gains can only predict some specific outcomes but not others amongst patients with health anxiety. It is also possible that statistical power may have been lacking in this sample, given that there

were trends towards greater improvements amongst sudden gainers on all outcomes and the original RCT sample-size was not powered for this purpose. Taken together, the current results suggest that sudden gains on the ORS may indicate improvements in some domains, but, as previously identified, supplementing the ORS with disorder-specific assessment is likely to help understand changes of particular interest (Nugter et al., 2017).

A key strength of this study is the account it takes of multiple changes over time, as – through applying a multilevel modelling design – it does not rely on a specific measurement at a particular time-point. As such, the study accounts for trends during and after psychotherapy across health domains. This approach has not often been used in sudden gains research.

### ***Limitations***

Previous ROM studies have often been carried out in usual care settings. This adds two elements missing from the current study: Larger, mixed patient samples and large therapist samples. In particular, the limited variance amongst the four therapists in this study meant that the outcome variance often present in larger therapist samples was absent. However, the high rates of multimorbidity, including several mental health problems, suggest that findings may be relevant to other patient groups – including mixed usual care populations. This study aimed to assess the value of ROM within a patient population where a variety of physical and mental health changes would be important – it did not therefore aim to be comprehensive, rather it aimed to be illustrative of the value ROM might have for outcomes amongst people with difficulties which are hard to monitor using single outcomes. The analysis aimed to account for comorbidities by included the total number of disorders into models, but this was only a significant predictor of outcome in health anxiety and generalized anxiety. This may have been because total number of diagnoses was too crude an assessment of the impact multimorbidity can have.

The criteria used to identify sudden gains have been subject to debate, with no clear agreement on the standards to be used in different clinical populations. The Tang and DeRubeis (1999) criteria are the most commonly used and the adaptation to include very early gains aimed to fit with the general finding that very early changes can be significant. The proportion of patients identified as experiencing sudden gains in this study is broadly similar to those found in previous studies and the impact is comparable (Aderka et al., 2012). However, the null findings on the primary outcome (health anxiety) and some other measures suggests that there are important limitations to the value of ROM sudden gains identified in the way used. This is particularly important when reducing health anxiety was the primary focus and purpose of the psychotherapy in question. Yet, sudden gains on the ORS did not translate into differential gains in health anxiety.

Common methods variance was addressed by the study design in some ways, but the repeated completion of self-report measures could still have inflated outcome associations. This means that the relationship between ORS scores and follow-up outcomes could have been over-emphasized by the fact that similar types of assessment were completed by the same person at different time points.

In addition, this study does not give evidence for the effectiveness of ROM per se. It cannot establish whether use of ROM improved the therapy offered or not. However, it does inform the utility of ROM amongst patients with health anxiety using rigorous methodology.

### ***Future research and implications***

Future investigations of ROM systems should include a larger sample incorporating independent, blinded, multiple assessments during and for a sustained period after psychological treatment. The primary outcome should be assessed session-by-session alongside any ROM system to give a clearer understanding of their covariance over treatment. This would also inform how frequently sudden gains on ROM systems co-occur

with sudden gains on the primary outcome. It would also aid understanding of the possible additive effect of general and disorder-specific ROM.

Overall, this study suggests that general ROM assessment is specifically relevant to high users of health care with severe health anxiety – a patient population who rarely seek psychological therapy and a group that healthcare services have struggled to help generally (Tyrer, Eilenberg, Fink, Hedman, & Tyrer, 2016). This study also indicates that the ORS score at the final session of CBT for severe health anxiety has long-term prognostic value across a range of physical and mental health domains. In particular, this is one of the first studies to investigate the relationship between the ORS and quality of life, overall health, social functioning or work functioning. In practical terms, this gives more confidence for the use of the ORS in services offering psychological therapies for problems that span physical and mental health, such as severe health anxiety. This study also emphasizes the value of disorder-specific feedback alongside general outcome feedback. For example, during treatment a therapist may assume improvement in health anxiety when there is a large stable improvement on the ORS, but the findings presented suggest this assumption could be incorrect.

## References

- Aderka, I., 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.
- Barnett, K., Mercer, S. W., Norbury, M., Watt, G., Wyke, S., & Guthrie, B. (2012). Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet, 380*(9836), 37-43.
- Busch, A. M., Kanter, J. W., Landes, S. J., & Kohlenberg, R. J. (2006). Sudden gains and outcome: A broader temporal analysis of cognitive therapy for depression. *Behavior therapy, 37*(1), 61-68.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin, 56*(2), 81.
- Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *The Lancet, 391*(10121), 679-686.
- Clark, D. M., & Oates, M. (2014). *Improving Access to Psychological Therapies: Measuring improvement and recovery Adult Services. Version 2*. London: NHS England.
- Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., . . . Nevin, J. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: a multisite, open-label, cluster randomised controlled trial. *The Lancet Psychiatry, 5*(7), 564-572.
- Fink, P., Ørnbøl, E., Toft, T., Sparle, K. C., Frostholm, L., & Olesen, F. (2004). A new, empirically established hypochondriasis diagnosis. *American Journal of Psychiatry, 161*(9), 1680-1691.

- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1998). *Structured Clinical Interview for DSM-IV Axis I Disorders: Patient Edition (February 1996 Final)*, SCID-I/P: Biometrics Research Department, New York State Psychiatric Institute.
- Hedman, E., Lekander, M., Ljotsson, B., Lindefors, N., Ruck, C., Hofmann, S. G., . . . Schulz, S. M. (2014). Sudden gains in internet-based cognitive behaviour therapy for severe health anxiety. *Behaviour Research & Therapy*, *54*, 22-29.  
doi:10.1016/j.brat.2013.12.007
- Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., . . . Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research*, *20*(10), 1727-1736.
- Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.
- Kendrick, T., Moore, M., Gilbody, S., Churchill, R., Stuart, B., & El-Gohary, M. (2014). Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database of Systematic Reviews*(7).
- 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.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2002). The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. *Psychosomatic Medicine*, *64*(2), 258-266.
- Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring and feedback in clinical practice*. Washington, DC: American Psychological Association.
- Lambert, M. J. (2012). Helping clinicians to use and learn from research-based systems: The OQ-analyst. *Psychotherapy*, *49*(2), 109.

- Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy, 55*(4), 520.
- Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (in-press). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. In J. C. Norcross & L. M. J (Eds.), *Psychotherapy relationships that work* (3rd ed.). New York, NY: Oxford University Press.
- Luborsky, L., Diguier, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., . . . Schweizer, E. (1999). The researcher's own therapy allegiances: A “wild card” in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*(1), 95-106.
- Miller, S. D., Duncan, B., Brown, J., Sparks, J., & Claud, D. (2003). The outcome rating scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of brief Therapy, 2*(2), 91-100.
- Miller, S. D., Duncan, B. L., Sorrell, R., & Brown, G. S. (2005). The partners for change outcome management system. *Journal of Clinical Psychology, 61*(2), 199-208.
- Morriss, R., Patel, S., Malins, S., Guo, B., Higton, F., James, M., . . . Kaylor-Hughes, C. (2019). Clinical and economic outcomes of remotely delivered cognitive behaviour therapy versus treatment as usual for repeat unscheduled care users with severe health anxiety: a multicentre randomised controlled trial. *BMC medicine, 17*(1), 16.
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *The British Journal of Psychiatry, 180*(5), 461-464.
- Naylor, C., Das, P., Ross, S., Honeyman, M., Thompson, J., & Gilbert, H. (2016). *Bringing together physical and mental health*. London, UK: King's Fund.

- Nugter, M. A., Hermens, M. L., Robbers, S., Van Son, G., Theunissen, J., & Engelsbel, F. (2017). Use of outcome measurements in clinical practice: How specific should one be? *Psychotherapy research*, 1-13.
- Østergård, O. K., Randa, H., & Hougaard, E. (2018). The effect of using the Partners for Change Outcome Management System as feedback tool in psychotherapy—A systematic review and meta-analysis. *Psychotherapy research*, 1-18.
- Patel, S., Kai, J., Atha, C., Avery, A., Guo, B., James, M., . . . Morriss, R. (2015). Clinical characteristics of persistent frequent attenders in primary care: case-control study. *Family practice*, 32(6), 624-630.
- Prescott, D., Maeschalck, C., & Miller, S. D. (2017). *Feedback informed treatment in clinical practice: Reaching for excellence*. Washington, DC: American Psychological Association.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: : Sage Publications.
- Rise, M. B., Eriksen, L., Grimstad, H., & Steinsbekk, A. (2016). The long-term effect on mental health symptoms and patient activation of using patient feedback scales in mental health out-patient treatment. A randomised controlled trial. *Patient education and counseling*, 99(1), 164-168.
- Salkovskis, P. M., Rimes, K., Warwick, H., & Clark, D. (2002). The Health Anxiety Inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. *Psychological Medicine*, 32(1), 843-853.
- Salkovskis, P. M., & Warwick, H. (1986). Morbid preoccupations, health anxiety and reassurance: a cognitive-behavioural approach to hypochondriasis. *Behaviour research and therapy*, 24(5), 597-602.

- Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of consulting and clinical psychology, 78*(3), 298.
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097.
- Sunderland, M., Newby, J. M., & Andrews, G. (2013). Health anxiety in Australia: prevalence, comorbidity, disability and service use. *The British Journal of Psychiatry, 202*(1), 56-61.
- 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.
- 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), 168-172.
- Tang, T. Z., DeRubeis, R. J., Hollon, S. D., Amsterdam, J., & Shelton, R. (2007). Sudden gains in cognitive therapy of depression and depression relapse/recurrence. *Journal of Consulting & Clinical Psychology, 75*(3), 404-408. doi:10.1037/0022-006X.75.3.404
- Tyrer, P., Eilenberg, T., Fink, P., Hedman, E., & Tyrer, H. (2016). Health anxiety: the silent, disabling epidemic.
- Tyrer, P., Salkovskis, P. M., Tyrer, H., Wang, D., Crawford, M. J., Dupont, S., . . . Smith, G. (2017). Cognitive-behaviour therapy for health anxiety in medical patients (CHAMP): a randomised controlled trial with outcomes to 5 years. *Health Technology Assessment, 21*(50), 1-58.

Wampold, B. E. (2015). Routine outcome monitoring: Coming of age—With the usual developmental challenges.

Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Juni, P., Altman, D. G., . . . Sterne, J. A. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*, 336(7644), 601-605.

Table 1. Illustration of multilevel model development for last session ORS predicting health anxiety over time.

Step	Model	Deviance (-2LL)	Change in deviance, new dfs	Was parameter added statistically significant? <sup>a</sup>	Residual variance	Intercept variance	Slope variance
1	Unconditional	1271	-	-	16.53	34.39	-
2	Fixed effects and random intercept for time	1237	34, 1	yes	12.92	35.95	-
3	Random slope for time (removed)	1235	2, 1	no	11.93	31.35	0.65
4	Control variable 1: Fixed effect of baseline SHAI	1010	227, 1	yes	13.59	20.83	-
5	Control variable 2: Fixed effect of 1st session ORS	1004	6, 1	yes	13.64	17.55	-
6	Control variable 3: Number of diagnoses	1003	1, 1	no	13.65	17.26	-
7	Main predictor: Fixed effect of last session ORS	998	5, 1	yes	13.59	15.48	-

Note. Abbreviations: -2LL: -2 log likelihood; dfs: degrees of freedom; ORS: outcome rating scale; WAI-O: working alliance inventory – observer rated. <sup>a</sup>Statistical significance was assessed using chi-squared for the change in -2LL and change in dfs.

Table 2. Demographics and clinical characteristics of included and excluded participants.

<i>Demographics</i>	Included Sample N=56	Excluded Sample N=23
Females	44 (79%)	13 (57%)
Mean age (SD)	37 (17)	40 (19)
Ethnicity		
White British	42 (75%)	17 (74%)
Unemployed	6 (11%)	6 (26%)
<i>Clinical Characteristics</i>		
Baseline wellbeing - ORS (SD)	20.5 (6.5)	-
Baseline health anxiety - SHAI (SD)	27.2 (5.2)	27.7 (6.1)
Baseline depression – PHQ-9 (SD)	13.4 (6.4)	13.1 (6.4)
Baseline anxiety – GAD-7 (SD)	13.3 (5.4)	12.9 (5.6)
Baseline Somatic symptoms – PHQ-15 (SD)	13.9 (5.1)	14.0 (6.5)
Baseline functional impairment – WSAS (SD)	17.0 (11.1)	17.6 (11.8)
Baseline quality of life – EQ5D-5L (SD)	.624 (.245)	0.572 (.353)
Baseline general health – VAS (SD)	54.0 (21.5)	55.1 (23.5)
<i>Mean number of SCID Diagnoses (range)</i>	6 (0-16)	6 (0-13)
Generalized anxiety disorder	38 (68%)	15 (65%)
Hypochondriasis	35 (63%)	12 (52%)
Current depressive episode	34 (61%)	15 (65%)
Somatoform disorder	32 (57%)	13 (57%)
Panic disorder	32 (57%)	11 (48%)
<i>Mean chronic physical health problems (range)</i>	1 (0-6)	1 (0-3)
Chronic pain conditions	17 (30%)	4 (17%)
Bowel disorders (IBS, colitis, diverticulitis)	8 (14%)	1 (4%)
Chronic fatigue	6 (11%)	0
Functional neurological disorders	5 (9%)	0
Diabetes	3 (5%)	1 (2%)

*Note.* SD: standard deviation; ORS: outcome rating scale; SHAI Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalized anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale; SCID: structured clinical interview for DSM IV disorders.

Table 3. Pearson’s correlations between baseline outcome scores and wellbeing at the first treatment session.

	ORS	SHAI	EQ5D	GAD-7	WSAS	VAS	PHQ-9	PHQ-15
Wellbeing (ORS)	1	-.055	.496**	-.400**	-.336*	.293*	-.502**	-.340*
Health anxiety (SHAI)		1	.043	.389**	.044	-.161	.22	.096
Quality of Life (EQ5D)			1	-.336*	-.523**	.570**	-.488**	-.401**
Generalized anxiety (GAD-7)				1	.253	-.479**	.634**	.314*
Functional impairment (WSAS)					1	-.411**	.654**	.521**
General health (VAS)						1	-.430**	-.304*
Depression (PHQ-9)							1	.632**
Somatic symptoms (PHQ-15)								1

*Note.* \* $p < .05$  \*\* $p < .01$ .  $n = 56$ . Abbreviations: ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalized anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale.

Table 4. Spearman's correlations between 6-month follow-up outcome scores and wellbeing at the last treatment session.

	ORS	SHAI	EQ5D	GAD-7	WSAS	VAS	PHQ-9	PHQ-15
Wellbeing (ORS)	1	-.435**	.388*	-.365*	-.446**	.354*	-.433**	-.310*
<i>n</i>	56	46	37	45	45	46	45	46
Health anxiety (SHAI)		1	-.449**	.520**	.286	-.340*	.482**	.333*
<i>n</i>		46	37	45	45	46	45	46
Quality of Life (EQ5D)			1	-.351*	-.786**	.591**	-.721**	-.762**
<i>n</i>			37	36	36	37	36	37
Generalized anxiety (GAD-7)				1	.364*	-.445**	.739**	.278
<i>n</i>				45	45	45	45	45
Functional impairment (WSAS)					1	-.493**	.660**	.587**
<i>n</i>					45	45	45	45
General health (VAS)						1	-.666**	-.556**
<i>n</i>						46	45	46
Depression (PHQ-9)							1	.648**
<i>n</i>							45	45
Somatic symptoms (PHQ-15)								1
<i>n</i>								46

Note. \* $p < .05$  \*\* $p < .01$ . Abbreviations: ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalized anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale.

Table 5. Predictive validity of patient wellbeing at the last treatment session on follow-up outcomes over 3, 6, 9 and 12-month follow-ups, controlling for baseline severity and changes over time.

<b>Health anxiety (SHAI)</b>										
Fixed effects						Random effects				
Parameter	B	SE	<i>p</i>	95% CI		Parameter	Variance	SE	Z	<i>p</i>
				Lower Bound	Upper Bound					
Intercept	22.53	5.01	<.001	12.47	32.59	Residual variance	13.59	1.75	7.75	<.001
Assessment time point	-1.53	0.26	<.001	-2.05	-1.01	Intercept variance	15.48	3.93	3.94	<.001
Baseline health anxiety (SHAI)	0.22	0.13	.107	-0.05	0.49					
session 1 wellbeing (ORS)	-0.08	0.10	.391	-0.28	0.11					
Total diagnoses	0.44	0.20	.028	0.05	0.83					
<i>Last session wellbeing (ORS)</i>	<i>-0.34</i>	<i>0.09</i>	<i>&lt;.001</i>	<i>-0.51</i>	<i>-0.16</i>					
<b>Somatic symptoms (PHQ-15)</b>										
Fixed effects						Random effects				
Parameter	B	SE	<i>p</i>	95% CI		Parameter	Variance	SE	Z	<i>p</i>
				Lower Bound	Upper Bound					
Intercept	13.21	3.55	.001	6.06	20.36	Residual variance	7.84	1.05	7.44	<.001
Assessment time point	-0.76	0.20	<.001	-1.16	-0.36	Intercept variance	9.72	2.65	3.67	<.001
Baseline somatic symptoms (PHQ-15)	0.52	0.10	<.001	0.31	0.73					
session 1 wellbeing (ORS)	-0.11	0.08	.184	-0.27	0.05					
Total diagnoses	0.004	0.14	.978	-0.28	0.29					
<i>Last session wellbeing (ORS)</i>	<i>-0.19</i>	<i>0.07</i>	<i>.008</i>	<i>-0.33</i>	<i>-0.05</i>					

<b>Generalised anxiety (GAD-7)</b>										
Fixed effects						Random effects				
Parameter	B	SE	p	95% CI		Parameter	Variance	SE	Z	p
				Lower Bound	Upper Bound					
Intercept	16.27	2.74	<.001	10.75	21.78	Residual variance	10.63	1.43	7.45	<.001
Assessment time point	-0.90	0.24	<.001	-1.36	-0.43	Intercept variance	5.12	1.82	2.82	0.005
Baseline generalized anxiety (GAD-7)	0.15	0.09	.087	-0.02	0.33					
session 1 wellbeing (ORS)	-0.11	0.07	.129	-0.26	0.03					
Total diagnoses	0.30	0.12	.019	0.05	0.55					
<i>Last session wellbeing (ORS)</i>	<i>-0.33</i>	<i>0.06</i>	<i>&lt;.001</i>	<i>-0.44</i>	<i>-0.21</i>					

<b>Depression (PHQ-9)</b>										
Fixed effects						Random effects				
Parameter	B	SE	p	95% CI		Parameter	Variance	SE	Z	p
				Lower Bound	Upper Bound					
Intercept	17.83	3.89	<.001	10.00	25.65	Residual variance	10.43	1.41	7.41	<.001
Assessment time point	-0.44	0.24	.067	-0.90	0.03	Intercept variance	10.81	3.00	3.61	<.001
Baseline depression (PHQ-9)	0.32	0.10	.002	0.12	0.52					
session 1 wellbeing (ORS)	-0.13	0.10	.190	-0.32	0.07					
Total diagnoses	0.001	0.15	.994	-0.31	0.31					
<i>Last session wellbeing (ORS)</i>	<i>-0.34</i>	<i>0.07</i>	<i>&lt;.001</i>	<i>-0.49</i>	<i>-0.19</i>					

<b>General health (VAS)</b>										
Fixed effects						Random effects				
Parameter	B	SE	p	95% CI		Parameter	Variance	SE	Z	p
				Lower Bound	Upper Bound					
Intercept	13.21	3.55	.001	6.06	20.36	Residual variance	7.84	1.05	7.44	<.001
Assessment time point	-0.76	0.20	<.001	-1.16	-0.36	Intercept variance	9.72	2.65	3.67	<.001
Baseline general health (VAS)	0.52	0.10	<.001	0.31	0.73					
session 1 wellbeing (ORS)	-0.11	0.08	.184	-0.27	0.05					
Total diagnoses	0.00	0.14	.978	-0.28	0.29					
<i>Last session wellbeing (ORS)</i>	<i>-0.19</i>	<i>0.07</i>	<i>.008</i>	<i>-0.33</i>	<i>-0.05</i>					
<b>Quality of life (EQ-5D)</b>										
Fixed effects						Random effects				
Parameter	B	SE	p	95% CI		Parameter	Variance	SE	Z	p
				Lower Bound	Upper Bound					
Intercept	-0.018	0.135	.896	-0.289	0.253	Residual variance	0.015	0.002	7.27	<.001
Assessment time point	0.019	0.009	.040	0.001	0.036	Intercept variance	0.025	0.006	4.10	<.001
Baseline quality of life (EQ-5D)	0.643	0.120	<.001	0.401	0.885					
session 1 wellbeing (ORS)	-0.002	0.004	.633	-0.011	0.007					
Total diagnoses	0.002	0.007	.817	-0.012	0.016					
<i>Last session wellbeing (ORS)</i>	<i>0.008</i>	<i>0.003</i>	<i>.021</i>	<i>0.001</i>	<i>0.015</i>					

Functional impairment (WSAS)										
Fixed effects						Random effects				
Parameter	B	SE	p	95% CI		Parameter	Variance	SE	Z	p
				Lower Bound	Upper Bound					
Intercept	17.30	6.43	0.010	4.38	30.23	Residual variance	30.65	4.14	7.40	<.001
Assessment time point	-1.01	0.41	0.014	-1.82	-0.20	Intercept variance	32.82	9.06	3.62	<.001
Baseline functional impairment (WSAS)	0.59	0.09	<.001	0.40	0.77					
Session 1 wellbeing (ORS)	-0.06	0.15	0.716	-0.36	0.25					
Total diagnoses	0.02	0.27	0.926	-0.51	0.56					
<i>Last session wellbeing (ORS)</i>	<i>-0.46</i>	<i>0.13</i>	<i>0.001</i>	<i>-0.73</i>	<i>-0.20</i>					

*Note.* Abbreviations: ; B: unstandardized  $\beta$  coefficient; SE: standard error; p: significance at an alpha of .05; CI: confidence interval;; Z: Walt’s z test for significance of variance; ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalized anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale.

Table 6. Pre-post treatment effect sizes and predictive effect sizes for post-treatment wellbeing and health anxiety.

Outcome	Pre-post treatment effect size (Cohen's <i>d</i> ) <sup>a</sup>	% Reduction in unexplained variance unique to last session ORS	% Reduction in unexplained variance unique to 6-month SHAI
Health anxiety (SHAI)	1.9	30	-
Generalised anxiety (GAD-7)	1.4	35	29
Wellbeing (ORS)	1.3	-	11
Depression (PHQ-9)	0.8	24	27
General health (VAS)	0.7	15	19
Somatic Symptoms (PHQ-15)	0.6	11	16
Functional impairment (WSAS)	0.6	9	10
Quality of life (EQ-5D)	0.2	10	15

*Note.* <sup>a</sup>Cohen's *d* effect sizes descriptors: 0.2: small; 0.5: Medium  $\geq$ 0.8: Large.

Abbreviations: ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalized anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale.

Figure 1. The average sudden gain. The averaged Outcome Rating Scale (ORS) scores are shown for the sessions around the pre-gain session (N).

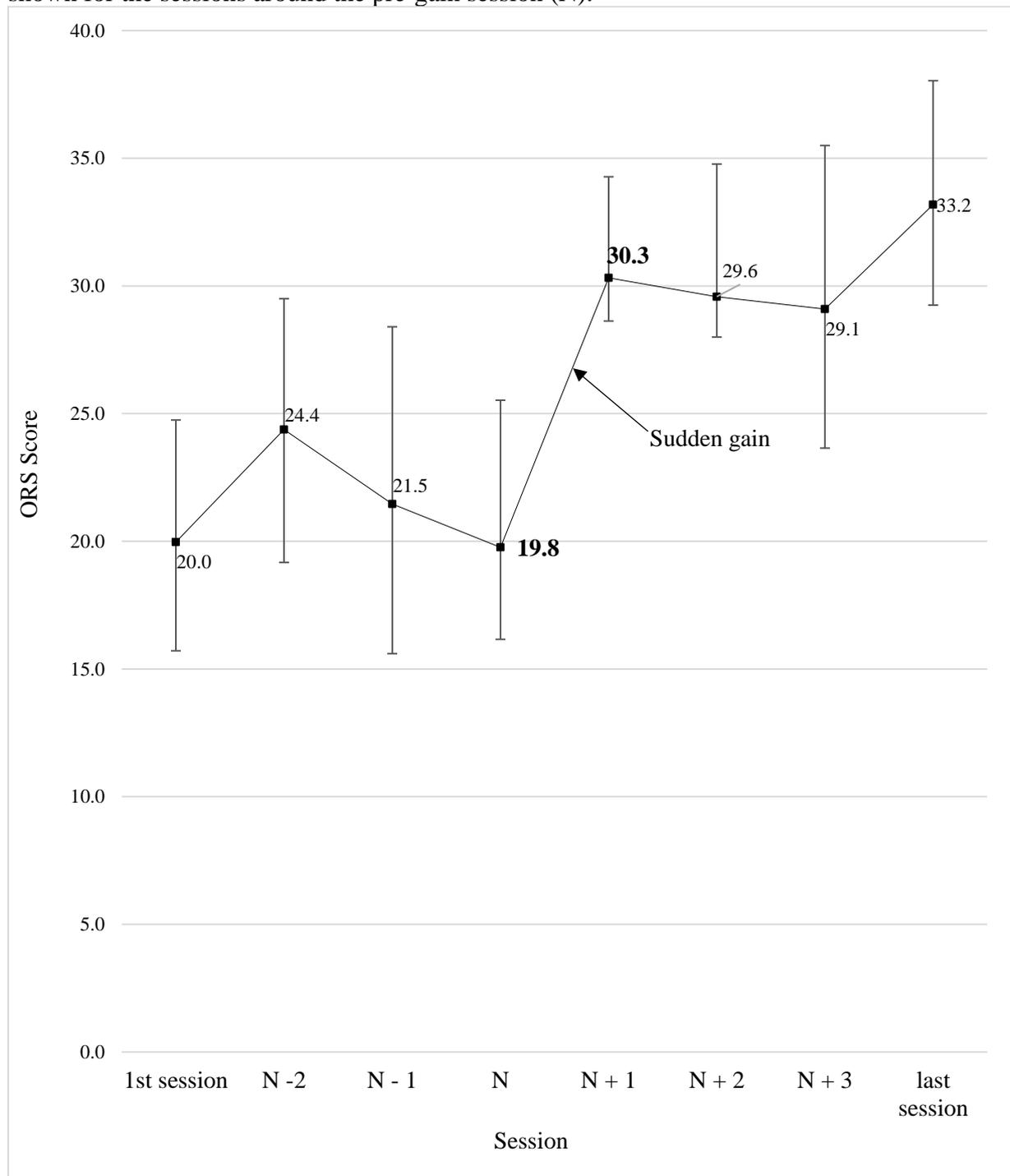
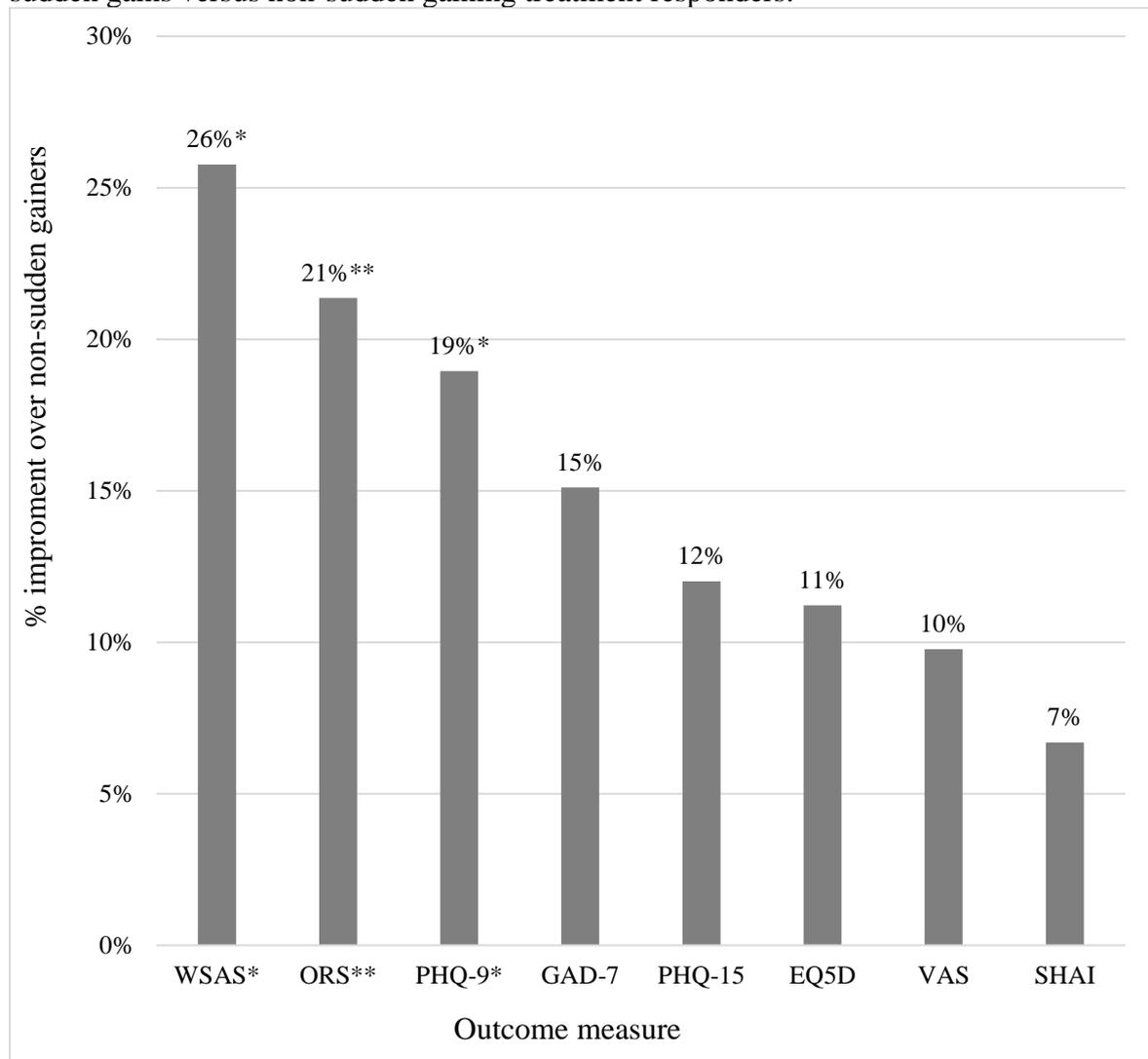


Figure 2. Percentage greater outcome improvement achieved by treatment responders with sudden gains versus non-sudden gaining treatment responders.



Note. \*p < .05 \*\*p < .01. Abbreviations: ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalized anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale.