Psychological Medicine

cambridge.org/psm

Empirically defining treatment response and remission in body dysmorphic disorder

Lorena Fernández de la Cruz[1,2], Jesper Enander[1], Christian Rück[1,2], Sabine Wilhelm[3], Katharine A. Phillips[4,5], Gail Steketee[6], Suraj Sarvode Mothi[3], Georgina Krebs[7,8], Laura Bowyer[9], Benedetta Monzani[7], David Veale[9,10] and David Mataix-Cols[1,2]

[1]Department of Clinical Neuroscience, Centre for Psychiatry Research, Karolinska Institutet, Stockholm, Sweden; [2]Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden; [3]Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; [4]New York-Presbyterian Hospital and Weill Cornell Medical College, New York, NY, USA; [5]Rhode Island Hospital and the Alpert Medical School of Brown University, Providence, RI, USA; [6]Boston University, School of Social Work, Boston, MA, USA; [7]The National and Specialist OCD, BDD and Related Disorders Team, Child and Adolescent Mental Health Services, South London and Maudsley NHS Foundation Trust, London, UK; [8]MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; [9]Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK and [10]Centre for Anxiety Disorders and Trauma, South London and Maudsley NHS Foundation Trust, London, UK

Original Article

Cite this article: Fernández de la Cruz L *et al* (2021). Empirically defining treatment response and remission in body dysmorphic disorder. *Psychological Medicine* **51**, 83–89. https://doi.org/10.1017/S0033291719003003

Received: 1 April 2019
Revised: 2 September 2019
Accepted: 1 October 2019
First published online: 30 October 2019

Key words:
Body dysmorphic disorder; clinical trials; signal detection; symptom remission; treatment response

Author for correspondence:
Lorena Fernández de la Cruz,
E-mail: lorena.fernandez.de.la.cruz@ki.se

Abstract

Background. The number of clinical trials in body dysmorphic disorder (BDD) has steadily increased in recent years. As the number of studies grows, it is important to define the most empirically useful definitions for response and remission in order to enhance field-wide consistency and comparisons of treatment outcomes across studies. In this study, we aim to operationally define treatment response and remission in BDD.

Method. We pooled data from three randomized controlled trials of cognitive-behavior therapy (CBT) for BDD (combined $n = 153$) conducted at four academic sites in Sweden, the USA, and England. Using signal detection methods, we examined the Yale-Brown Obsessive Compulsive Scale modified for BDD (BDD–YBOCS) score that most reliably identified patients who responded to CBT and those who achieved remission from BDD symptoms at the end of treatment.

Results. A BDD–YBOCS reduction ⩾30% was most predictive of treatment response as defined by the Clinical Global Impression (CGI) – Improvement scale (sensitivity 0.89, specificity 0.91, 91% correctly classified). At post-treatment, a BDD–YBOCS score ⩽16 was the best predictor of full or partial symptom remission (sensitivity 0.85, specificity 0.99, 97% correctly classified), defined by the CGI – Severity scale.

Conclusion. Based on these results, we propose conceptual and operational definitions of response and full or partial remission in BDD. A consensus regarding these constructs will improve the interpretation and comparison of future clinical trials, as well as improve communication among researchers, clinicians, and patients. Further research is needed, especially regarding definitions of full remission, recovery, and relapse.

Introduction

Body dysmorphic disorder (BDD) is characterized by excessive preoccupation with perceived defects in physical appearance that are not observable or appear only slight to others. This preoccupation leads to significant distress or significant impairment (usually both), avoidance behaviors (usually of social situations), and time-consuming repetitive behaviors (American Psychiatric Association, 2013). BDD is often a chronic disorder that persists if left untreated, and is associated with high rates of psychiatric hospitalization and suicidal behavior (Veale *et al.*, 1996; Phillips *et al.*, 2005). BDD is classified in the obsessive-compulsive and related disorders chapter in both the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) (American Psychiatric Association, 2013) and the International Classification of Diseases, 11th Revision (ICD-11) (Veale and Matsunaga, 2014). Recommended evidence-based treatments from randomized controlled trials (RCTs) include cognitive-behavior therapy (CBT) (Harrison *et al.*, 2016), clomipramine (Hollander *et al.*, 1999), and selective serotonin reuptake inhibitors (SSRIs) (Phillips *et al.*, 2002; Phillips *et al.*, 2016).

As the evidence for the clinical management of BDD continues to accumulate, an important goal for the future is to develop empirically grounded definitions of treatment response and remission for this condition. Generally, *treatment response* is defined as a meaningful improvement of symptoms compared to the beginning of treatment (Frank *et al.*, 1991;

© The Author(s) 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

CAMBRIDGE UNIVERSITY PRESS

Farris *et al.*, 2013; Mataix-Cols *et al.*, 2016). In turn, *remission* is defined as being illness-free after having been ill (Frank *et al.*, 1991; Farris *et al.*, 2013; Mataix-Cols *et al.*, 2016). Remission may be full (i.e. complete), in which there are no remaining symptoms, or partial, which reflects some remaining symptoms that are subthreshold for the diagnosis (Phillips *et al.*, 2006; Phillips *et al.*, 2013). While there seems to be a broad consensus regarding these conceptual definitions (Mataix-Cols *et al.*, 2016), operational definitions are generally more difficult to agree upon. In obsessive-compulsive disorder (OCD), a disorder closely related to BDD, the lack of agreement over these operational definitions has resulted in poor comparability among studies and treatment modalities, as varying definitions lead to different estimates of treatment efficacy and relapse risk. However, progress has recently been made with international consensus of OCD experts for defining such constructs (Mataix-Cols *et al.*, 2016).

In BDD, some initial progress in the empirical definition of treatment response has already been made. In the first report on the psychometric properties of the Yale-Brown Obsessive Compulsive Scale Modified for Body Dysmorphic Disorder (BDD-YBOCS), among 30 patients treated with the SRI fluvoxamine, Phillips and collaborators found that a cut-off of ⩾30% improvement on the BDD-YBOCS corresponded best to at least 'much improved' on the Clinical Global Impression – Improvement (CGI-I) scale (Phillips *et al.*, 1997). Subsequently, in a report of 63 BDD patients who were treated with various SSRIs, Phillips *et al.* (2014) again found that a cut-off of ⩾30% improvement on the BDD-YBOCS corresponded best to at least 'much improved' on the CGI-I; in addition, a score of ⩾50% best corresponded best to 'very much improved,' on the CGI-I. However, this previous work has been done only in pharmacotherapy studies, and it is important to replicate and extend this work to other samples and treatment modalities, such as CBT.

Various definitions of remission have also been proposed. In a prospective naturalistic observational study of the course of BDD (in which treatment was not assigned or controlled), Phillips *et al.* (2016) used the Longitudinal Interval Follow-Up Evaluation (LIFE) (Keller *et al.*, 1987), which employs the Psychiatric Status Rating (PSR) scale to indicate the global severity of DSM-defined disorders. The BDD-PSR is a 7-point scale that maps directly onto DSM-IV diagnostic criteria for BDD. Phillips *et al.* (2016) followed the LIFE convention, defining full remission as minimal or no BDD symptoms (scores of 1 or 2 on the BDD-PSR) during a period of at least eight consecutive weeks (Phillips *et al.*, 2006; Phillips *et al.*, 2013). They defined partial remission as a score of 3 or 4 on the BDD-PSR for at least eight consecutive weeks. This study did not report on the relationship between BDD-PSR scores and BDD-YBOCS scores, despite the fact that the BDD-YBOCS is considered to be the 'gold standard' outcome measure in this population. In another study, Veale *et al.* (2015) defined partial remission as a reduction in BDD-YBOCS score from above 24 at pre-treatment to between 12 and 24 at post-treatment, whereas full remission was defined as a reduction in from above 24 to below 12.

In summary, there has been consistency in definitions of treatment response, although data are limited to pharmacotherapy studies. Definitions of remission in BDD have been inconsistent, and no study to date has attempted to establish an empirically derived definition for this construct. As the number of future clinical trials in BDD grows, deriving operational definitions of treatment response and remission seems warranted. The use of inconsistent definitions of response and remission can lead to quite different estimates, which would, in turn, make comparison of different studies and treatment modalities challenging and hinder communication among researchers, clinicians, and patients.

In this study, we reanalyzed data from adults and adolescents with BDD who participated in various trials of CBT for BDD to empirically determine the optimal amount of symptom improvement to define treatment response and remission. Based on these results, we propose conceptual and operational definitions for these constructs for use in future clinical trials.

Method

Participants

We pooled data from 153 individuals who took part in three different RCTs of CBT for BDD published between 2014 and 2016; trials were carried out at four academic sites in Sweden ($n = 93$) (Enander *et al.*, 2016), the USA ($n = 31$) (Wilhelm *et al.*, 2014), and England ($n = 29$) (Mataix-Cols *et al.*, 2015). Only those individuals from the original studies who had pre- and post-treatment measurements in the outcomes of interest were included. The studies from Sweden and the USA contained adult samples, while the English study contained adolescents with BDD. Two of the RCTs offered face-to-face CBT (Wilhelm *et al.*, 2014; Mataix-Cols *et al.*, 2015) and one offered Internet-based CBT (Enander *et al.*, 2016).

Participants in the Enander *et al.* (2016) study received either CBT or supportive therapy – both delivered via the Internet – for twelve weeks. The patients in the Wilhelm *et al.* (2014) study were randomized to 22 face-to-face sessions of CBT over 24 weeks or a 12-week waitlist. For the current study, the measures collected at week 12 in Wilhelm *et al.* (2014) will be considered the end of the treatment scores because the controlled phase of the trial finished at that time-point. The adolescents in the Mataix-Cols *et al.* (2015) study were randomized to 14 sessions of face-to-face treatment delivered over four months or to a control condition of equivalent duration consisting of written psychoeducation materials and weekly telephone monitoring. In all three studies, the CBT protocol was heavily based on exposure with response prevention strategies. The Wilhelm *et al.* study also included substantial cognitive therapy and additional techniques such as perceptual retraining or mindfulness (2014).

The pooled sample was predominantly female ($n = 123$, 80%). Participants' ages ranged from 12 to 72 years (mean = 30.26, S.D. = 12.78). Participants in the American and English sites were predominantly White ($n = 51$, 85%). Race and/or ethnicity were not recorded in the Swedish study. Additional information about the individual studies can be found in the original reports (Wilhelm *et al.*, 2014; Mataix-Cols *et al.*, 2015; Enander *et al.*, 2016).

Measures

For all participants, the diagnosis of BDD was established according to the diagnostic criteria from the DSM-IV using the Structured Clinical Interview for DSM-IV (First *et al.*, 2001). All outcome measures were administered by independent assessors who were blind to the treatment condition (CBT or control/waitlist).

The primary outcome measure in all of the studies was the BDD-YBOCS. The BDD-YBOCS is a 12-item, semi-structured, rater-administered scale that assesses BDD symptom severity during the past week. The first five items assess obsessional

https://doi.org/10.1017/S0033291719003003 Published online by Cambridge University Press

preoccupations about perceived appearance defects (time preoccupied, interference in functioning and distress due to perceived appearance defects, attempt to resist preoccupations, and control over preoccupations). Items six to 10 assess BDD-related repetitive behaviors (e.g. excessive grooming, mirror checking) and mirror the content of the first five items (i.e. time spent performing the behaviors, interference in functioning due to the behaviors, distress experienced if the behaviors are prevented, attempt to resist behaviors, and control over the behaviors). Item 11 assesses insight regarding appearance beliefs, and item 12 assesses avoidance behaviors (e.g. occupational/academic or social activities). Scores for each item range from 0 to 4; the total score ranges from 0 to 48, with higher scores reflecting more severe symptoms. The BDD-YBOCS has shown good psychometric properties, including strong interrater and test-retest reliability, internal consistency, validity, and sensitivity to change (Phillips *et al*., 1997; Phillips *et al*., 2014).

The CGI is a clinician-rated scale that assesses overall improvement (CGI-I) and severity (CGI-S). The CGI-I consists of a single 7-point item scored from 1 (*'very much improved'*) to 7 (*'very much worse'*). Similarly, the CGI-S is a single 7-point item scored from 1 (*'normal, not at all ill'*) to 7 (*'extremely ill'*). The CGI has been widely used in psychiatric treatment trials and has shown good reliability and validity for a range of psychiatric disorders (Zaider *et al*., 2003; Kadouri *et al*., 2007). As in previous OCD and BDD studies, the CGI-I (Phillips *et al*., 1997; Phillips *et al*., 2014; Mataix-Cols *et al*., 2016) and the CGI-S (Mataix-Cols *et al*., 2016) were employed as the criteria against which we tested the change in BDD-YBOCS scores to define treatment response and remission, respectively. Specifically, CGI-I scores of 1 (*'very much improved'*) or 2 (*'much improved'*) were used to indicate treatment response and CGI-S scores of 1 (*'normal, not at all ill'*) or 2 (*'borderline mentally ill'*) corresponded to full or partial remission.

Analyses

Means and standard deviations (S.D.) of the predictor (BDD-YBOCS) and criterion (CGI) variables were calculated at pre- and post-treatment. For treatment response, the percent reduction from pre- to post-treatment on the BDD-YBOCS was coded by increments of five percent. For remission, raw scores at post-treatment on the BDD-YBOCS were coded by one point increments. Using signal detection analysis, we calculated the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), efficiency, and weighted κ statistic of the percent reduction on the BDD-YBOCS or the post-treatment BDD-YBOCS score in relation to the criterion outcome (CGI-I or CGI-S) to determine the cut-offs that best defined treatment response and remission, respectively. Sensitivity is defined as the probability that the BDD-YBOCS will correctly detect positive responses according to the criterion variable (i.e. CGI) [true positives/(true positives + false negatives)]. Specificity is the probability that the BDD-YBOCS will correctly detect negative responses according to the criterion [true negatives/(true negatives + false positives)]. The PPV test (or precision) is the probability that the criterion correctly identifies positive responses on the BDD-YBOCS [true positives/(true positives + false positives)] and the NPV test is the probability that the criterion test correctly identifies negative responses on the BDD-YBOCS [true negative/(true negatives + false negatives)]. Efficiency (or accuracy of detections) is the probability that the predictor and the criterion variables will agree [(true positives + true negatives)/total *n*]. Strength of the agreement was assessed according to the guidelines provided by Landis and Koch (1977) for κ coefficients (e.g. κ's from 0.61 to 0.80 indicate 'substantial' agreement; κ's from 0.81 to 1.00 indicate 'almost perfect' agreement).

Results

Change in BDD severity with treatment

In the treated group (i.e. CBT condition, $n = 78$), BDD-YBOCS mean scores dropped from pre-treatment (mean = 31.49, S.D. = 5.55) to post-treatment (mean = 21.68, S.D. = 8.70), reflecting a mean of 30.52% (S.D. = 26.22) reduction in BDD severity. As expected, this reduction was less in the control group ($n = 75$), in which baseline BDD-YBOCS mean scores changed from 30.80 (S.D. = 5.54) to 28.09 (S.D. = 6.65) at post-treatment (mean percentage of reduction = 7.58%, S.D. = 17.72). The percentage of change between the two groups was significantly larger in the CBT group (t-test = −6.36, $p < 0.001$).

Consistent with the BDD-YBOCS measurements, in the CBT group, mean CGI-S scores improved across time from the moderately-markedly ill range (mean = 4.50, S.D. = 0.73) to values closer to the mildly ill range (mean = 3.21, S.D. = 1.31). In contrast, in the control group, both pre- and post-treatment scores were in the moderately-markedly ill range (mean pre-treatment = 4.41, S.D. = 0.83; mean post-treatment = 4.22, S.D. = 0.96). The post-treatment CGI-S score was significantly different across groups (t-test = 5.43; $p < 0.001$). CGI-I scores in the CBT group ranged from much improved to minimally improved (mean = 2.63, S.D. = 1.11), whereas the control condition presented scores close to 'no change' (mean = 3.73, S.D. = 0.85). Again, this difference was significant between groups (t-test = 6.86; $p < 0.001$).

Treatment response

Thirty-two participants (42.67%) in the CBT group and six (7.79%) in the control group ($\chi^2 = 24.64$; $p < 0.001$) were treatment responders according to the predefined criterion measure (i.e. CGI-I scores of 1 or 2 (much or very much improved)). A 30% reduction in the BDD-YBOCS score was associated with the highest efficiency values in both the total sample (CBT + control) and the CBT group only (Table 1). The 30% cut-off also showed an optimal compromise between sensitivity and specificity in both the total sample and the CBT groups (0.89 and 0.91, and 0.94 and 0.81, respectively). κ values indicated substantial agreement between the predictor and the criterion variables (Table 1).

Full or partial remission

Twenty-three (30.67%) participants in the CBT group and two (2.56%) in the control group ($\chi^2 = 22.09$; $p < 0.0.001$) were in full or partial remission at the end of the treatment, according to the predefined criterion measure (i.e. CGI-S scores of 1 or 2). In the total sample, a BDD-YBOCS score ⩽15 was associated with the highest efficiency and the highest agreement value for full or partial remission (Table 2). At this cut-off, sensitivity was acceptable and specificity was high (0.72 and 0.99, respectively); in addition, 95% of those achieving remission and 95% of those not achieving remission were correctly identified as such. A BDD-YBOCS cut-off score of ⩽16 presented with similar efficiency and κ values and offered more balanced scores between sensitivity and specificity (0.80 and 0.97, respectively), with a slightly lower

https://doi.org/10.1017/S0033291719003003 Published online by Cambridge University Press

Table 1. Signal detection analysis predicting treatment response at various BDD-YBOCS percent reduction cut-off points using the CGI-I as the gold standard (scores 1: 'very much improved' to 2: 'much improved')

BDD-YBOCS percent reduction (%)	Sensitivity	Specificity	PPV	NPV	Efficiency	κ (0.5)
Total sample ($n = 152$)						
⩾5	1.00	0.45	0.38	1.00	0.58	0.29
⩾10	1.00	0.56	0.43	1.00	0.67	0.39
⩾15	1.00	0.70	0.53	1.00	0.78	0.54
⩾20	0.97	0.80	0.62	0.99	0.84	0.65
⩾25	0.95	0.85	0.68	0.98	0.87	0.70
⩾30	**0.89**	**0.91**	**0.77**	**0.96**	**0.91**	**0.77**
⩾35	0.71	0.92	0.75	0.90	0.87	0.64
⩾40	0.68	0.94	0.79	0.90	0.87	0.65
⩾45	0.60	0.97	0.88	0.88	0.88	0.65
⩾50	0.53	0.99	0.95	0.86	0.87	0.61
⩾55	0.34	1.00	1.00	0.82	0.83	0.44
⩾60	0.28	1.00	1.00	0.81	0.82	0.38
⩾65	0.24	1.00	1.00	0.80	0.81	0.32
⩾70	0.16	1.00	1.00	0.78	0.79	0.22
CBT condition ($n = 75$)						
⩾5	1.00	0.33	0.52	1.00	0.61	0.29
⩾10	1.00	0.44	0.60	1.00	0.68	0.40
⩾15	1.00	0.60	0.65	1.00	0.77	0.57
⩾20	1.00	0.67	0.70	1.00	0.81	0.64
⩾25	0.97	0.72	0.72	0.97	0.83	0.66
⩾30	**0.94**	**0.81**	**0.79**	**0.94**	**0.87**	**0.73**
⩾35	0.78	0.81	0.76	0.83	0.80	0.59
⩾40	0.75	0.86	0.80	0.82	0.81	0.61
⩾45	0.66	0.93	0.87	0.78	0.81	0.61
⩾50	0.56	0.98	0.95	0.75	0.80	0.57
⩾55	0.37	1.00	1.00	0.68	0.73	0.41
⩾60	0.31	1.00	1.00	0.66	0.71	0.34
⩾65	0.25	1.00	1.00	0.64	0.68	0.28
⩾70	0.16	1.00	1.00	0.61	0.64	0.17

BDD-YBOCS, Yale-Brown Obsessive Compulsive Scale modified for BDD; CBT, cognitive-behavior therapy; CGI-I, clinical global impression – improvement; NPV, negative predictive value; PPV, positive predictive value
Note: Optimal cut-offs are highlighted.

but still high score in precision (Table 2). When only the participant group who had received CBT was considered, the same cut-offs of 15–16 points appeared as the most efficient and presented the highest level of agreement with the criterion value (Table 2). Both values offered an acceptable level of sensitivity and excellent specificity (cut-off ⩽15 offered values of 0.70 and 0.98, and cut-off ⩽16 values of 0.78 and 0.94, respectively), as well as excellent PPV and NPV values (Table 2).

Discussion

In this study, we have calculated the optimal amount of symptom improvement needed to classify individuals with BDD as being responders to CBT, as well as the optimal amount of improvement to be classified as being in full or partial remission. Our findings on treatment response on the BDD-YBOCS are consistent with those of previous studies (Phillips *et al.*, 1997; Phillips *et al.*, 2014); no prior studies have examined the best BDD-YBOCS cut-off for remission. We propose the following conceptual and operational definitions of treatment response and remission to help guide future BDD research studies (also summarized in Table 3).

Treatment response

We conceptualize 'treatment response' as a clinically meaningful reduction in symptoms relative to baseline severity (Frank *et al.*,

https://doi.org/10.1017/S0033291719003003 Published online by Cambridge University Press

Table 2. Signal detection analysis predicting full or partial remission at various BDD-YBOCS cut-off points using the CGI-S as the gold standard (scores 1: 'normal, not at all ill' to 2: 'borderline mentally ill')

BDD-YBOCS cut-off	Sensitivity	Specificity	PPV	NPV	Efficiency	κ (0.5)
Total sample ($n = 153$)						
⩽4	0.08	1.00	1.00	0.85	0.85	0.13
⩽5	0.12	1.00	1.00	0.85	0.86	0.19
⩽6	0.16	1.00	1.00	0.86	0.86	0.24
⩽7	0.20	1.00	1.00	0.86	0.87	0.29
⩽8	0.20	1.00	1.00	0.86	0.87	0.29
⩽9	0.28	1.00	1.00	0.88	0.88	0.39
⩽10	0.32	1.00	1.00	0.88	0.89	0.44
⩽11	0.36	1.00	1.00	0.89	0.89	0.48
⩽12	0.52	1.00	1.00	0.91	0.92	0.64
⩽13	0.52	1.00	1.00	0.91	0.92	0.64
⩽14	0.68	0.99	0.94	0.94	0.94	0.76
⩽15	**0.72**	**0.99**	**0.95**	**0.95**	**0.95**	**0.79**
⩽16	**0.80**	**0.97**	**0.83**	**0.96**	**0.94**	**0.78**
⩽17	0.84	0.94	0.72	0.97	0.92	0.73
⩽18	0.88	0.91	0.65	0.97	0.90	0.69
⩽19	0.92	0.88	0.61	0.98	0.89	0.66
⩽20	1.00	0.84	0.56	1.00	0.87	0.64
⩽21	1.00	0.82	0.52	1.00	0.85	0.60
⩽22	1.00	0.77	0.45	1.00	0.80	0.52
⩽23	1.00	0.73	0.42	1.00	0.78	0.47
⩽24	1.00	0.66	0.36	1.00	0.71	0.38
CBT condition ($n = 75$)						
⩽4	0.09	1.00	1.00	0.71	0.72	0.12
⩽5	0.09	1.00	1.00	0.71	0.72	0.12
⩽6	0.13	1.00	1.00	0.72	0.73	0.17
⩽7	0.17	1.00	1.00	0.73	0.75	0.23
⩽8	0.17	1.00	1.00	0.73	0.75	0.23
⩽9	0.26	1.00	1.00	0.75	0.77	0.33
⩽10	0.30	1.00	1.00	0.76	0.79	0.38
⩽11	0.35	1.00	1.00	0.78	0.80	0.42
⩽12	0.52	1.00	1.00	0.82	0.85	0.60
⩽13	0.52	1.00	1.00	0.82	0.85	0.60
⩽14	0.65	0.98	0.94	0.86	0.88	0.69
⩽15	**0.70**	**0.98**	**0.94**	**0.88**	**0.89**	**0.73**
⩽16	**0.78**	**0.94**	**0.86**	**0.91**	**0.89**	**0.74**
⩽17	0.83	0.86	0.73	0.92	0.85	0.67
⩽18	0.87	0.81	0.67	0.93	0.83	0.62
⩽19	0.91	0.79	0.66	0.95	0.83	0.63
⩽20	1.00	0.73	0.62	1.00	0.81	0.62
⩽21	1.00	0.69	0.59	1.00	0.79	0.58
⩽22	1.00	0.63	0.55	1.00	0.75	0.52
⩽23	1.00	0.63	0.55	1.00	0.75	0.52
⩽24	1.00	0.56	0.50	1.00	0.69	0.44

BDD-YBOCS, Yale-Brown Obsessive Compulsive Scale modified for BDD; CBT, cognitive-behavior therapy; CGI-I, clinical global impression – improvement; NPV, negative predictive value; PPV, positive predictive value.
Note: Optimal cut-offs are highlighted.

https://doi.org/10.1017/S0033291719003003 Published online by Cambridge University Press

Table 3. Definitions and operationalization of treatment response and full or partial remission for BDD

	Conceptual definition	Operationalization
Treatment response	A clinically meaningful reduction in symptoms (time, distress, and interference associated with preoccupations about appearance, repetitive behaviors, and avoidance) relative to baseline severity in an individual who meets diagnostic criteria for BDD.	A ⩾30% reduction in BDD-YBOCS scores, lasting for *at least one week*.
Full or partial remission	Residual preoccupations about appearance, repetitive behaviors, and avoidance may be present but are minimally time consuming and cause no more than minimal distress or interference in the person's everyday life.	A score of ⩽16 on the BDD-YBOCS, lasting for *at least one week*.

BDD-YBOCS, Yale-Brown Obsessive Compulsive Scale modified for BDD.

1991; Mataix-Cols *et al.*, 2016). Based on our empirical results, and those of two prior studies by Phillips and collaborators (Phillips *et al.*, 1997; Phillips *et al.*, 2014), a BDD-YBOCS reduction ⩾30% seems to be the most clinically meaningful reduction to define response.

The BDD-YBOCS asks questions that refer to the *previous week*. Because some trial participants are not assessed at every session in CBT studies and for consistency with the corresponding OCD definitions (Mataix-Cols *et al.*, 2016), we propose that the operational definition of treatment response has a duration requirement of *at least one week*. This coincides with the primary end point employed in most clinical trials, that is, the post-treatment outcome evaluation.

Remission

Strictly speaking, patients should be classified as being 'in remission' when they no longer meet syndromal (i.e. threshold) criteria for the disorder and have no more than minimal symptoms (Farris *et al.*, 2013; Mataix-Cols *et al.*, 2016). We therefore suggest that, if feasible, diagnostic interviews should be used to determine that the person no longer meets diagnostic criteria for BDD at post-treatment. However, it is not common practice in BDD clinical trials to conduct structured diagnostic interviews at the end of treatment. If diagnostic interviews are unavailable, a BDD-YBOCS cut-off score could also be used to define full or partial remission at post-treatment. Clinically, these would be patients with more than very minimal or no symptoms but not meeting full DSM diagnostic criteria for BDD. Based on our results, the appropriate cut-off to define at least partial remission would be a score of either ⩽15 or ⩽16. However, because it has somewhat better balance between sensitivity and specificity, we recommend the use of the ⩽16 cut-off (rather than ⩽15). Further studies in larger samples are needed to further examine the best cut-off for full remission in BDD, which we did not examine separately in this study because few patients were considered fully remitted, according to the CGI-S scores (only seven patients scored 1, corresponding to 'normal, not at all ill').

Also, following the rationale outlined for treatment response, we propose that the duration requirement of full or partial remission in treatment studies be *at least one week* at the trial's end point. Although duration longer than a week would be desirable to define remission, it would be highly impractical and unfeasible to extend the primary endpoint of a clinical trial beyond the final post-treatment assessment. This recommendation is consistent with the BDD-YBOCS's assessment of symptom severity during the past week. Similar consensus was achieved for OCD (Mataix-Cols *et al.*, 2016).

We wish to emphasize that the definitions of remission in prospective naturalistic observational course of illness studies differ from definitions used in treatment studies in terms of required duration. This is the case for both BDD and other disorders. If participants are followed over years, as in course of illness studies, it makes sense to define remission in terms of months as opposed to one week. On the other hand, a definition of several months is not suitable for treatment studies, which may last for only a few weeks/months. The recommendations in this paper pertain only to treatment studies.

Limitations

The results of this study have to be interpreted in light of some limitations. Our sample was the result of the combination of samples from three different studies conducted in different countries that delivered treatments using different CBT modalities and included participants from different age groups. However, we lacked statistical power to examine different age groups and CBT modalities separately. Additionally, this study only used those data collected during the controlled phase of the RCTs as we wanted to take advantage of variables such as the blindness of the raters to the treatment condition. Most notably, our 'gold standard' measure of remission, the CGI-S, has limitations. For example, the score of 2 ('borderline mentally ill'), included in the definition of full or partial remission, is vague and imprecisely defined. Though the CGI-S was used as a secondary outcome measure in the included trials, it is possible that CGI-S scores reflect a broader set of symptoms other than BDD symptoms, leading to less precise scoring. Another limitation is that we did not examine full remission because so few patients fully remitted. Additionally, we did not address the issue of defining relapse or recovery in BDD; such definitions have previously been proposed and used in a few studies but require additional examination (Phillips *et al.*, 2006; Phillips *et al.*, 2013; Veale *et al.*, 2015; Phillips *et al.*, 2016). The impact of using previous or other alternative definitions of relapse or recovery requires empirical evaluation before recommendations can be made. Another limitation is that the BDD-YBOCS does not have normative data from the general population. Such data might assist in defining cut-offs, but the scale cannot be normed in this way because it is intended for use only in individuals who have been diagnosed with BDD.

Conclusions

We are hopeful that the proposed definitions for treatment response and remission (Table 3) will result in (1) improved design, interpretation, and comparison of clinical trials of various modalities (e.g. CBT *v.* medication trials), (2) improved communication

https://doi.org/10.1017/S0033291719003003 Published online by Cambridge University Press

of research findings between professionals and to the general public, (3) improved guidelines for evaluation of clinical efficacy of various treatments by regulatory agencies, and (4) development of improved treatment guidelines for clinical practice (Frank *et al.*, 1991). However, additional studies in more and larger samples are needed to further examine optimal cut-offs and definitions for remission (both full and partial), recovery, and relapse in treatment studies of BDD.

Acknowledgements. Dr Fernández de la Cruz is funded from the Swedish Research Council for Health, Working Life and Welfare (FORTE grant number 2015-00569). Dr Rück acknowledges funding from the regional agreement on medical training and clinical research (ALF) between the Stockholm County Council and Karolinska Institutet, the Swedish Research Council (reference number K2013-61X-22168-01-3), and the Swedish Society of Medicine (Söderströmska Königska sjukhemmet, grant No: SLS3B4451). Dr Wilhelm acknowledges funding from the National Institute of Mental Health (grant reference number R34 MH070490). Dr Krebs is funded by an MRC Clinical Research Training Fellowship (MR/N001400/1). Dr Veale is part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. Dr Mataix-Cols acknowledges funding from the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0110-21231).

Conflict of interest. Dr Fernández de la Cruz and Prof. Mataix-Cols receive royalties for contributing articles to UpToDate, Wolters Kluwer Health. In the past year, Dr Phillips has received research support and/or salary support from the Institute of Mental Health, the Norman Prince Neurosciences Institute/Brown Institute for Brain Science, Brown University Division of Biology and Medicine, and the National Institute of General Medical Sciences (salary and research support). She has received royalties from Oxford University Press, American Psychiatric Publishing, International Creative Management, Inc., UpToDate, and Guilford Press. She has received honoraria from Royal Pharma and the Merck Manual. The rest of authors report no conflicts of interest.

References

American Psychiatric Association (2013) *The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*). Washington, DC: American Psychiatric Association.

Enander J, Andersson E, Mataix-Cols D, Lichtenstein L, Alström K, Andersson G, Ljótsson B and Rück C (2016) Therapist guided internet based cognitive behavioural therapy for body dysmorphic disorder: single blind randomised controlled trial. *BMJ* **352**, i241.

Farris SG, McLean CP, Van Meter PE, Simpson HB and Foa EB (2013) Treatment response, symptom remission, and wellness in obsessive-compulsive disorder. *Journal of Clinical Psychiatry* **74**, 685–690.

First MB, Spitzer RL, Gibbon M and Williams JB (2001) *Structured Clinical Interview for DSM-IV-TR Axis I Disorders—Patient Edition* (*SCID-I/P. 2/2001 Revision*). New York: Biometrics Research Department, New York State Psychiatric Institute.

Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, Rush AJ and Weissman MM (1991) Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. *Archives of General Psychiatry* **48**, 851–855.

Harrison A, Fernández de la Cruz L, Enander J, Radua J and Mataix-Cols D (2016) Cognitive-behavioral therapy for body dysmorphic disorder: a systematic review and meta-analysis of randomized controlled trials. *Clinical Psychology Review* **48**, 43–51.

Hollander E, Allen A, Kwon J, Aronowitz B, Schmeidler J, Wong C and Simeon D (1999) Clomipramine vs desipramine crossover trial in body dysmorphic disorder – selective efficacy of a serotonin reuptake inhibitor in imagined ugliness. *Archives of General Psychiatry* **56**, 1033–1039.

Kadouri A, Corruble E and Falissard B (2007) The improved Clinical Global Impression Scale (iCGI): development and validation in depression. *BMC Psychiatry* **7**, 7.

Keller MB, Lavori PW, Friedman B, Nielsen E, Endicott J, McDonald-Scott P and Andreasen NC (1987) The longitudinal interval follow-up evaluation. A comprehensive method for assessing outcome in prospective longitudinal studies. *Archives of General Psychiatry* **44**, 540–548.

Landis JR and Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.

Mataix-Cols D, Fernández de la Cruz L, Isomura K, Anson M, Turner C, Monzani B, Cadman J, Bowyer L, Heyman I, Veale D and Krebs G (2015) A pilot randomized controlled trial of cognitive-behavioral therapy for adolescents with body dysmorphic disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* **54**, 895–904.

Mataix-Cols D, Fernández de la Cruz L, Nordsletten AE, Lenhard F, Isomura K and Simpson HB (2016) Towards an international expert consensus for defining treatment response, remission, recovery, and relapse for obsessive-compulsive disorder: a Delphi study. *World Psychiatry* **15**, 80–81.

Phillips KA, Hollander E, Rasmussen SA, Aronowitz BR, DeCaria C and Goodman WK (1997) A severity rating scale for body dysmorphic disorder: development, reliability, and validity of a modified version of the Yale-Brown Obsessive Compulsive Scale. *Psychopharmacology Bulletin* **33**, 17–22.

Phillips KA, Albertini RS and Rasmussen SA (2002) A randomized placebo-controlled trial of fluoxetine in body dysmorphic disorder. *Archives of General Psychiatry* **59**, 381–388.

Phillips KA, Coles ME, Menard W, Yen S, Fay C and Weisberg RB (2005) Suicidal ideation and suicide attempts in body dysmorphic disorder. *Journal of Clinical Psychiatry* **66**, 717–725.

Phillips KA, Pagano ME, Menard W and Stout RL (2006) A 12-month follow-up study of the course of body dysmorphic disorder. *American Journal of Psychiatry* **163**, 907–912.

Phillips KA, Menard W, Quinn E, Didie ER and Stout RL (2013) A 4-year prospective observational follow-up study of course and predictors of course in body dysmorphic disorder. *Psychological Medicine* **43**, 1109–1117.

Phillips KA, Hart AS and Menard W (2014) Psychometric evaluation of the Yale-Brown Obsessive-Compulsive Scale Modified for Body Dysmorphic Disorder (BDD-YBOCS). *Journal of Obsessive-Compulsive and Related Disorders* **3**, 205–208.

Phillips KA, Keshaviah A, Dougherty DD, Stout RL, Menard W and Wilhelm S (2016) Pharmacotherapy relapse prevention in body dysmorphic disorder: a double-blind, placebo-controlled trial. *American Journal of Psychiatry* **173**, 887–895.

Veale D and Matsunaga H (2014) Body dysmorphic disorder and olfactory reference disorder: proposals for ICD-11. *Revista Brasileira De Psiquiatria* **36**, S14–S20.

Veale D, Boocock A, Gournay K, Dryden W, Shah F, Willson R and Walburn J (1996) Body dysmorphic disorder. A survey of fifty cases. *British Journal of Psychiatry* **169**, 196–201.

Veale D, Miles S and Anson M (2015) Long-term outcome of cognitive behavior therapy for body dysmorphic disorder: a naturalistic case series of 1 to 4 years after a controlled trial. *Behavior Therapy* **46**, 775–785.

Wilhelm S, Phillips KA, Didie E, Buhlmann U, Greenberg JL, Fama JM, Keshaviah A and Steketee G (2014) Modular cognitive-behavioral therapy for body dysmorphic disorder: a randomized controlled trial. *Behavior Therapy* **45**, 314–327.

Zaider TI, Heimberg RG, Fresco DM, Schneier FR and Liebowitz MR (2003) Evaluation of the Clinical Global Impression Scale among individuals with social anxiety disorder. *Psychological Medicine* **33**, 611–622.

https://doi.org/10.1017/S0033291719003003 Published online by Cambridge University Press