ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Behavior Therapy xxx (2021) xxx–xxx

Behavior Therapy

www.elsevier.com/locate/bt

Mapping Female Patients' Judgments of Satisfaction to Hypothetical Changes in Depression Symptom Severity

Thomas T. Kim
Colin Xu
Robert J. DeRubeis
University of Pennsylvania

Within mental health, approaches to determine whether a patient experienced "meaningful" change from treatment have predominantly involved imposing thresholds on three types of metrics derived from assessments of symptom severity: end score (posttreatment score), absolute change (pre- minus posttreatment score), and proportion of change. However, none of these approaches have considered input from the consumer. This study examined correspondences between various reductions from pre- to posttreatment symptom severity levels and patients' judgments of satisfaction with change. Former or currently depressed patients were asked to provide judgments of their satisfaction reflected in vignettes that used descriptions from the Hamilton Rating Scale for Depression. Judgments from 108 female participants were fit using four metrics: end score, absolute change, proportion of change, and the combination of end score and absolute change. Akaike information criteria (AICs) and Akaike weights were used to determine the best-fitting model. Cutoffs were calculated for the five levels of satisfaction with change. Proportion of change best accounted for variation in the patients' ratings. For "slightly . . . ," "somewhat . . . ," "moderately . . . ," and "very . . . ," the proportions of reduction that corresponded with each of these ratings of satisfaction were, respectively: 17%, 39%, 62%, and 84%. Our a priori level of satisfaction (between "somewhat" and "moderately") corresponded to a 50% reduction in pretreatment severity. This study may provide services some insight into their female patients' satisfaction with change from treatment for depression using only the proportion of reduction from pretreatment severity. A similar procedure could be applied to other diagnostic groups, as well as other constructs that attend to the patient's perspective.

Keywords: clinically meaningful change; Reliable Change Index; clinical significance

TREATMENTS FOR MENTAL HEALTH DISORDERS are often evaluated, or compared with alternative treatments or control conditions, using continuous measures of symptom severity (e.g., Hamilton Rating Scale for Depression; Hamilton, 1960). However, there are contexts, such as the construction of quality control metrics in large-scale mental health services (Clark et al., 2009), in which dichotomous judgments about individuals' responses to treatment (e.g., the patients did or did not meet remission criteria) are desired. Dichotomous judgments that derive from pre- and posttreatment measures of symptom severity are also commonly used in systematic reviews (Howlin et al., 2009; Leichsenring & Klein, 2014; Ost et al., 2015).

In psychotherapy research, the construct "clinical significance" has been proposed to distinguish reductions in symptoms that are considered "meaningful" from reductions that fall short of a given clinical significance criteria. The implication is that a clinically significant change makes a real difference in the everyday life of the patient (McGlinchey et al., 2008). The most popular approach to defining clinical significance was first described in 1984 by Jacobson et al. They argued that a patient's improvement in therapy should be considered clinically significant if and only if the patient began treatment in the dysfunctional range on a symptom measure and ended with a

Address correspondence to Thomas T. Kim, Levin 456, 425 South University Avenue, University of Pennsylvania, Philadelphia, PA. e-mail: thomastk@sas.upenn.edu.

Please cite this article as: Kim, Xu and DeRubeis, Mapping Female Patients' Judgments of Satisfaction to Hypothetical Changes in Depression Symptom Severity, Behavior Therapy, https://doi.org/10.1016/j.beth.2021.10.003

score closer to the mean of a functional population. Jacobson and Truax (1991) later noted that a small reduction in symptom score could meet their definition if the patient began therapy with a score only somewhat closer to the dysfunctional range than to the functional range. To address this limitation, they proposed that in addition to criteria that relied only on the posttreatment score, a criterion based on a difference score, or the Reliable Change Index, should be applied. They recommended the use of psychometric information to construct the index. Over the past three decades, the term "reliable and clinically significant change" and metrics that derive from it have been widely adopted in the psychotherapy outcome literature (Arch et al., 2012; Clark et al., 2006; Franklin et al., 2000; Giesen-Bloo et al., 2006).

An alternative means of defining "meaningful" change, commonly used in evaluations of psychiatric medications, is to set a minimum requirement for the proportion of reduction in a symptom score from pre- to posttreatment. For example, the American College of Neuropsychopharmacology Task Force (Rush et al., 2006) recommended that for evaluations of treatments for depression to qualify as a "response," the symptom score should be reduced by 50% or more, relative to the pretreatment score (Hiller et al., 2012; Prien et al., 1991).

Both of these commonly used approaches to defining "meaningful" change are applied using the patient's pre- and posttreatment severity scores, but the way in which these scores are combined have not been empirically derived, using the judgments of patients or clinicians (Kazdin, 1999; Zimmerman et al., 2006b). A related construct, proposed by McGlothlin and Lewis (2014), is the "minimal clinically important difference" (MCID), defined as "the smallest benefit of value to patients." Button et al. (2015) used patients' judgments to derive an MCID index, in terms of the magnitude of change between two successive scores on the Beck Depression Inventory–II (BDI-II; Beck et al., 1996). Button et al. compared a difference metric (the simple difference between scores) with one that reflected proportion of change (number of point reductions divided by the pretreatment score). They found that the proportion of change metric tracked patients' judgments more closely than did the difference metric. Their general approach can be applied to identify the optimal rules for corresponding reductions in symptom scores from the start to the end of therapy to patients' evaluations of these changes.

The Current Study

We asked individuals who are or had been in treatment for depression to evaluate descriptions of improvements in depressive symptoms, presented as vignettes, that depicted a variety of pre- and posttreatment levels of symptom severity. Vignettes were created to represent specific pairs of scores, based on the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960), before and after treatment. We then evaluated which of four metrics best fit participants' judgments of satisfaction with change: end score (posttreatment score), absolute change (the difference between pre- and posttreatment scores), proportion of change, or the combination of end score and absolute change. The metric that fit patients' judgments best was then used to estimate the cutoff values that corresponded to each of five levels of satisfaction with change.

Material and Methods

PROCEDURE

Participants completed an online survey in which they were presented with four vignettes, each of which used language from the HRSD to describe the symptoms experienced by a hypothetical patient prior to the initiation of treatment, as well as the changes in those symptoms over the course of a 4-month treatment for depression. The intake scores reflected in the vignettes ranged from mild to very severe, as defined by the American Psychiatric Association (2000), at intervals of three points, with 13 (*mild*) being the lowest, and 25 (*very severe*) the highest. The posttreatment score reflected in each vignette was always lower than the pretreatment score, with the largest difference being 25 (pretreatment score of 25 to posttreatment score of 0), and the smallest being 2 (16 to 14). There were 62 combinations of pre- and posttreatment scores represented in the stimulus set.

VIGNETTES

We aimed to set the specific symptom levels for each of the pre- and posttreatment HRSD scores represented in the vignettes so as to reflect a typical symptom profile for that HRSD total score. To achieve this, using HRSD item and total scores from 616 patients who had received treatment in one of two randomized clinical trials (DeRubeis et al., 2005; Hollon et al., 2014), we estimated the mean score on each symptom for each of the total HRSD scores represented in the vignettes. For example, data from 141 patients whose post-

Please cite this article as: Kim, Xu and DeRubeis, Mapping Female Patients' Judgments of Satisfaction to Hypothetical Changes in Depression Symptom Severity, Behavior Therapy, https://doi.org/10.1016/j.beth.2021.10.003

treatment HRSD total score was 13 were used to construct the descriptions in vignettes that reflected this score. The average scores on Items 1 (depressed mood), 2 (guilt), and 3 (suicide) were 2.01, 1.31, and 0.67, respectively. These estimates, as well as those for the other HRSD items, were rounded to determine the descriptions of the symptoms in the vignette, with the constraint that the item scores represented must sum to 13. The same procedure was followed for each of the other HRSD total scores. Descriptive anchors from Hamilton's (1960) version of the HRSD were then adapted for the survey. For example, Item 1 (depressed mood) was represented by these phrases: 0 = "She no longer reports feelings of sadness, hopelessness, or worthlessness"; 1 = "She reports mild feelings of sadness and hopelessness, intermittently throughout the week"; 2 = "She reports moderate feelings of sadness and hopelessness, more days than not"; and 3 = "She reports persistent and strong feelings of sadness, hopelessness, and worthlessness, with occasional crying." For simplicity, insomnia descriptions comprised one sentence that covered initial, middle, and delayed insomnia. Likewise, we collapsed the somatic items (Items 13, 14, and 15) into one sentence, and did the same for the weight and appetite items (11 and 16). Participants were not given the numeric values that corresponded with each symptom level. Table 1 provides an example of a vignette, in which the pretreatment score represented was 19 and the posttreatment score was 10. After reading a vignette, with the instruction to imagine that one has experienced the changes represented in it, the participant indicated an answer to the following question, "Considering how you were feeling before treatment, how *satisfied* would you be with the amount of change you experienced from treatment?" The response options were *not at all*, *slightly*, *somewhat*, *moderately*, *very*, and *extremely*.

Vignettes were varied on two dimensions: pretreatment HRSD score and amount of symptom change. The pretreatment HRSD score varied from very severe (25), severe (22), moderately severe (19), moderate (16), and mild (13). The amount of symptom change varied from minimal (2–5 points of improvement), more than minimal (6–8 points of improvement), moderate (9–14 points of improvement), and substantial reduction (14–25 points of reduction). The ranges of symptom improvement were determined in discussions with multiple clinicians. In order to reduce participant burden, each participant encountered a set of four vignettes (rather than all 62 vignettes), where each set contained a vignette of a minimal reduction of symptom severity, more than a minimal reduction, a moderate reduction, and a substantial reduction. Vignettes were presented in randomized order. The name of the individual in the vignette was randomized to one of the four following names: Jack, Claire, Sophie, and Dan. These names were chosen from a list of the "most popular names in the United Kingdom."

Table 1
Hypothetical Patient With a Pretreatment HRSD Total Score of 19 to a Posttreatment HRSD Total Score of 10

Before starting treatment, Claire ...	After 4 months of treatment, Claire ...
... experienced persistent and strong feelings of sadness, hopelessness, and worthlessness with occasional crying.	... occasionally experiences mild feelings of sadness and hopelessness throughout the week.
... often experienced feelings of guilt, which were at times hard to control.	... experiences occasional thoughts that she has let other people down.
... sometimes thought that life was not worth living, but did not think about suicide.	... is no longer bothered by thoughts of death or suicide.
... was frequently having difficulty with falling and staying asleep. It often took her 30 minutes to fall asleep most nights, and she frequently woke up an hour before she needed to.	... experiences some difficulty with falling asleep one or two nights each week and is often restless during the night.
... was spending very little time on things she used to enjoy, was turning down social opportunities, and was less productive at work because of a lack of interest and motivation.	... experiences decreased interest than normal in work and hobbies, and often needs to push herself to do things.
... showed normal expression of affect and speech.	... *still* shows normal expression of affect and speech.
... experienced nervousness and anxiety frequently and often worried about minor matters.	... experiences occasional periods of tension and irritability.
... noticed that her appetite decreased slightly—however, she did not lose weight.	... *still* reports a slight decrease in appetite and still has not lost weight.
... experienced extreme fatigue and tiredness almost every day, and had no interest in sex.	... experiences occasional periods of fatigue and a decrease in libido.

Note. HRSD = Hamilton Rating Scale for Depression.

Please cite this article as: Kim, Xu and DeRubeis, Mapping Female Patients' Judgments of Satisfaction to Hypothetical Changes in Depression Symptom Severity, Behavior Therapy, https://doi.org/10.1016/j.beth.2021.10.003

PARTICIPANTS

The Institutional Review Board approved all procedures. The study was posted on MQ's Take Part in Research platform, where UK citizens can participate in mental health research through online surveys or interviews. Their platform tends to attract—through word of mouth, their social channels, or email—a pool of participants who are interested in mental health, including former or current patients. MQ is a registered charity in England that focuses on transforming mental health through research. At the end of the survey, participants had the option to provide comments and to enter a raffle to win a £50 Amazon gift card.

At the beginning of their encounter with the survey, participants were presented with a consent form and assured that all responses would be anonymous. Only participants who answered that they had received treatment for depression were asked to continue the survey. Participants self-reported information about age, gender, whether or not they were currently depressed, and their history of mental health treatment. This included the type of treatment received and number of treatment sessions they recalled receiving. When asked to report the type of treatment they received, participants were presented with a multiple-choice questionnaire of the most common face-to-face therapies in the United Kingdom.

Note that while participants were initially recruited irrespective of their gender, a vast majority of the respondents were female (93.9%). Therefore, we decided to report results only from the female participants.

ANALYSIS

The dependent variable for all models was the participant's satisfaction rating. We created four mixed-effects models using the lmer package in R (Bates et al., 2007) with participant entered as a random effect. The four models were, with fixed-effect predictors: (a) end score, (b) absolute change, (c) proportion of change, and (d) end score and absolute change (as separate predictors).

We computed the Akaike information criteria (AIC) to characterize the goodness of fit for each of the models. The AIC is a popular method for comparing the adequacy of multiple models that are non-nested (Wagenmakers & Farrell, 2004), where the smaller the AIC, the better the model fit. Although significance tests cannot be applied to such comparisons (Wagenmakers & Farrell, 2004), an informative procedure is to calculate a set of Akaike weights for the candidate models. These weights are used to estimate each model's "selection uncertainty" (Symonds & Moussalli, 2011). One can then estimate the probability that a given model, among those examined, provides the best approximation.

The model with the highest probability was then queried to identify cutoffs corresponding to mean satisfaction ratings of 2 (*slightly satisfied*), 3 (*somewhat satisfied*), 4 (*moderately satisfied*), and 5 (*very satisfied*). A cutoff was also calculated for 3.5, the midpoint of our scale, which was the level of satisfaction we specified a priori as reflecting a dichotomous judgment of "meaningful" change prior to estimating the fit of the four models. Note that cutoffs for satisfaction ratings of 1 (*not at all satisfied*) and 6 (*extremely satisfied*) were not calculated due to the high likelihood of floor and ceiling effects.

Results

SATISFACTION WITH CHANGE FROM TREATMENT

A total of 385 responses were obtained from 108 female participants from the United Kingdom. The average number of responses for the 62 combinations of vignettes was 3.6, with a range of 2–20 responses (1 combination had 2 responses, 3 combinations had 3 responses, 20 combinations had 4 responses, and the rest of the combinations had more than 4 responses). Demographics of these participants can be found in Table 2. The AICs associated with each model were, in order of best to worst goodness of fit: proportion of change = 1101, end score and absolute change (in the same model) = 1115, end score only = 1165, and absolute change only = 1216. Akaike weights, in order of highest to lowest probability, were proportion of change = 0.998, end score and absolute change = 0.002, end score ≅ 0.00, and absolute change = ≅ 0.00. Akaike weights suggested that there is a 99.8% chance that the model with proportion of change best fit the data, among the four models we examined.

DEFINING THE CUTOFFS

Table 3 presents the proportion of change values that corresponded to mean satisfaction ratings of 2, 3, 3.5, 4, and 5. Our a priori-defined level of satisfaction, 3.5, corresponded with a 50% reduction in pretreatment score. Therefore, if this rule was used to define a service's criteria of "meaningful" change, a patient with a pretreatment HRSD score of 13 would need a posttreatment score less than or equal to 6 to meet the criteria.

Please cite this article as: Kim, Xu and DeRubeis, Mapping Female Patients' Judgments of Satisfaction to Hypothetical Changes in Depression Symptom Severity, Behavior Therapy, https://doi.org/10.1016/j.beth.2021.10.003

Table 2
Demographics of 108 Female Participants From the United Kingdom

Characteristics	
Age	37.1 ± 12.9
Self-reported currently depressed?	81.5%
Number of treatment sessions[a]	17
Received both antidepressants and psychotherapy	80.6%
Received only antidepressants	16.7%
Received only psychotherapy	2.8%
If received psychotherapy, what type?	
Cognitive-behavioral therapy	59.3%
Psychodynamic therapy	16.7%
Behavioral activation	2.8%

Note. Statistics reported are in percentages (*n*/*N*) for categorical variables and mean ± standard deviation for continuous variables. A total of 385 responses were obtained from the 108 participants.

[a] The median is reported for "Number of treatment sessions."

Table 3
Correspondence Between Model Estimates of Proportion of Change and the Levels of Satisfaction

Satisfaction rating	Proportion of change
2.0	17%
3.0	39%
3.5	50%
4.0	62%
5.0	84%

Note. Satisfaction ratings reflect these descriptions: 2.0 = *slightly*; 3.0 = *somewhat*; 4.0 = *moderately*; 5.0 = *very*; 3.5, the midpoint of our scale, was selected as the a priori-defined level of satisfaction.

Discussion

This is the first study—of which we are aware—that solicited judgments from former or current patients to distinguish degrees of "meaningful" change from a hypothetical course of treatment for depression. However, since our sample comprised primarily female participants, we analyzed data only from the female participants. It was perhaps unsurprising that the model with end score and the model with absolute change performed poorly. As Jacobson and Truax (1991) noted, the application of just a posttreatment threshold is prone to misclassify patients who not only experience a small reduction in symptoms but also start treatment with scores close to the posttreatment threshold. Similarly, the application of just an absolute change threshold would misclassify patients who experience a large reduction in symptoms, but end treatment with substantial symptomology. Thus, the model that included both absolute change and end score performed better than the single predictor models of absolute change and end score. However, the model that used the pre- and posttreatment scores to form a "proportion of change" predictor performed best of all.

Cutoffs, using the model with proportion of change, were calculated for each of these levels of satisfaction: "slightly ... ," "somewhat ... ," "moderately ... ," "very ... ," and our a priori-defined level. Since symptom severity is commonly assessed at pre- and posttreatment by providers and administrators (Clark et al., 2009), these cutoffs may provide services some insight into their female patients' satisfaction with change (either for an individual patient or an overall estimate) from treatment for depression. Additionally, researchers chose metrics, such as end score and absolute change, for their ease of applicability—these cutoffs retained this advantage.

Most psychopharmacological studies define "response" as improvement of at least 25–50% (Hiller et al., 2012). Furthermore, with depressive disorders there is widespread expert agreement that response requires at least a 50% reduction of pretreatment symptom severity for antidepressant treatment (Hiller et al., 2012). Note that neither the basis for this threshold (i.e., proportion of change), nor the recommended value (i.e., 50%), was determined with explicit input from stakeholders. Our study found a similar proportion of change rule for our a priori-defined level of satisfaction with change, suggesting that female patients may experience a satisfaction level at least between "somewhat" and "moderately" for those meeting the "response" criterion. Furthermore, an advantage of the "50% reduction" criterion for clinicians and other stakeholders is its straightforward application (Hiller et al., 2012).

Our findings echo those of Button et al. (2015), who found that the proportion of change metric, compared to absolute change, best fit patients' judgments of improvement on the BDI-II. They

Please cite this article as: Kim, Xu and DeRubeis, Mapping Female Patients' Judgments of Satisfaction to Hypothetical Changes in Depression Symptom Severity, Behavior Therapy, https://doi.org/10.1016/j.beth.2021.10.003

estimated an MCID of 17.5% reduction in scores from pretreatment severity. This value is very close to the estimate we obtained that corresponded with judgments of “slightly satisfied” with change.

A similar procedure, in principle, could be implemented in relation to changes during treatment for diagnostic groups other than major depressive disorder (MDD). However, the specific metric that would reflect satisfaction with change from treatment would need to be determined for other clinical disorders. For example, in posttraumatic stress disorder, a threshold higher than 50% might be appropriate for defining “meaningful” change (cf. Norton & Price, 2007). On the other hand, thresholds lower than 50% might be more suitable in studies of treatments for patients with chronic psychotic conditions.

LIMITATIONS AND FUTURE DIRECTIONS

The present study employed a moderately sized sample of female participants, all of whom stated that they had received a course of treatment for depression. Our sample may not be fully representative of all female individuals who received treatment for depression, as it was a convenience sample of unpaid volunteers who signed up via the MQ’s Take Part in Research platform. Furthermore, a review by MQ in 2020 of the visitors participating in their research surveys found that 91.8% identified as White, 3.8% as mixed/multiple ethnic groups, 3.1% as Asian, and 0.88% as Black. Future studies should assess the generalizability of this study’s findings to not only non-White racial and ethnic groups, but also other minority groups (e.g., those who identify as nonbinary gender). Another potential limitation is that participants self-reported their mental health and treatment history; a future study should replicate our methods with a sample of participants with verified mental health histories. Our study reported results only from female participants, and while the majority of patients treated in UK’s Improving Access to Psychological Therapies program are female (65%; Clark, 2018), future studies should obtain ratings of satisfaction with change from male participants.

While studies employing hypothetical vignettes can provide an interpretation of the real world (Hughes, 1998) and responses to vignettes can closely resemble responses to real-life experiences (if the vignettes appear to be both relevant and real to participants; Finch, 1987; Rahman, 1996), it may be worthwhile to query patients about their own experience of symptom changes from a course of treatment for depression. However, much larger samples would be needed for such an inquiry to ensure sufficient density of coverage of the range of pre- and posttreatment severity levels. This procedure may also be susceptible to recall bias—patients would have to consider their changes in symptoms dating back an extended period of time. Therefore, it may be worthwhile to provide patients with their pretreatment symptom history.

Future work might also focus on changes assessed with other depression measures, such as the Patient Health Questionnaire–9 (PHQ-9; Kroenke et al., 2001) or the BDI-II (Beck et al., 1996), which are used widely in clinics and in large mental health care systems (Clark et al., 2009).

The most common approaches to defining “meaningful” change from a course of treatment have largely excluded the patient’s perspective (Kazdin, 1999). While the present study examined the correspondence of satisfaction with change from a course of treatment with that of symptom change, future studies should apply a similar procedure for other constructs that attend to the patient’s perspective, such as level of functioning or quality of life (Zimmerman et al., 2006a). Future studies should also obtain clinicians’ judgments of changes in symptom severity, as their judgments might correspond to a different metric, as well as cutoffs, compared to patients’ judgments.

While previous methods to define whether a patient experienced “meaningful” change primarily imposed cutoffs on four metrics (end score, absolute change, proportion of change, or the combination of both end score and absolute change), a subsequent study should assess the ability of other metrics (e.g., a combination of both end score and proportion of change) to fit judgments of satisfaction with change from treatment.

Conflict of Interest Statement

The authors declare that there are no conflicts of interest.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.beth.2021.10.003.

References

American Psychiatric Association, Task Force for the Handbook of Psychiatric Measures & Rush, A. J. (2000). *Handbook of psychiatric measures*. American Psychiatric Association.

Arch, J. J., Eifert, G. H., Davies, C., Plumb Vilardaga, J. C., Rose, R. D., & Craske, M. G. (2012). Randomized clinical trial of cognitive behavioral therapy (CBT) versus acceptance and commitment therapy (ACT) for mixed anxiety disorders. *Journal of Consulting and Clinical Psychology, 80*, 750–765. https://doi.org/10.1037/a0028310.

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. J. R. (2007). The lme4 package. *Journal of Statistical Software, 2*, 74.

Please cite this article as: Kim, Xu and DeRubeis, Mapping Female Patients’ Judgments of Satisfaction to Hypothetical Changes in Depression Symptom Severity, Behavior Therapy, https://doi.org/10.1016/j.beth.2021.10.003

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. (1996). Comparison of Beck Depression Inventories–-IA and –II in psychiatric outpatients. *Journal of Personality Assessment, 67*, 588–597. https://doi.org/10.1207/s15327752jpa6703_13.

Button, K. S., Kounali, D., Thomas, L., Wiles, N. J., Peters, T. J., Welton, N. J., Ades, A. E., & Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory–II according to the patient's perspective. *Psychological Medicine, 45*, 3269–3279. https://doi.org/10.1017/S0033291715001270.

Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology, 14*, 159–183. https://doi.org/10.1146/annurev-clinpsy-050817-084833.

Clark, D. M., Ehlers, A., Hackmann, A., McManus, F., Fennell, M., Grey, N., Waddington, L., & Wild, J. (2006). Cognitive therapy versus exposure and applied relaxation in social phobia: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 74*, 568–578. https://doi.org/10.1037/0022-006X.74.3.568.

Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy, 47*, 910–920. https://doi.org/10.1016/j.brat.2009.07.010.

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., O'Reardon, J. P., Lovett, M. L., Gladis, M. M., Brown, L. L., & Gallop, R. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry, 62*, 409–416. https://doi.org/10.1001/archpsyc.62.4.409.

Finch, J. (1987). The vignette technique in survey research. *Sociology, 21*, 105–114. https://doi.org/10.1177/0038038587021001008.

Franklin, M. E., Abramowitz, J. S., Kozak, M. J., Levitt, J. T., & Foa, E. B. (2000). Effectiveness of exposure and ritual prevention for obsessive-compulsive disorder: Randomized compared with nonrandomized samples. *Journal of Consulting and Clinical Psychology, 68*, 594–602.

Giesen-Bloo, J., van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., van Asselt, T., Kremers, I., Nadort, M., & Arntz, A. (2006). Outpatient psychotherapy for borderline personality disorder: Randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry, 63*, 649–658. https://doi.org/10.1001/archpsyc.63.6.649.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*, 56–62. https://doi.org/10.1136/jnnp.23.1.56.

Hiller, W., Schindler, A. C., & Lambert, M. J. (2012). Defining response and remission in psychotherapy research: A comparison of the RCI and the method of percent improvement. *Psychotherapy Research, 22*, 1–11. https://doi.org/10.1080/10503307.2011.616237.

Hollon, S. D., DeRubeis, R. J., Fawcett, J., Amsterdam, J. D., Shelton, R. C., Zajecka, J., Young, P. R., & Gallop, R. (2014). Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder: A randomized clinical trial. *JAMA Psychiatry, 71*, 1157–1164. https://doi.org/10.1001/jamapsychiatry.2014.1054.

Howlin, P., Magiati, I., & Charman, T. (2009). Systematic review of early intensive behavioral interventions for children with autism. *American Journal on Intellectual and Devlopmental Disabilities, 114*, 23–41. https://doi.org/10.1352/2009.114:23;nd41.

Hughes, R. (1998). Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociology of Health and Illness, 20*, 381–400. https://doi.org/10.1111/1467-9566.00107.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336–352. https://doi.org/10.1016/S0005-7894(84)80002-7.

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

Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology, 67*, 332–339. https://doi.org/10.1037//0022-006x.67.3.332.

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*, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Leichsenring, F., & Klein, S. (2014). Evidence for psychodynamic psychotherapy in specific mental disorders: A systematic review. *Psychoanalytic Psychotherapy, 28*, 4–32. https://doi.org/10.1080/02668734.2013.865428.

McGlinchey, J. B., Zimmerman, M., & Atkins, D. C. (2008). Clinical significance and remission in treating major depressive disorder: Parallels between related outcome constructs. *Harvard Review of Psychiatry, 16*, 25–34. https://doi.org/10.1080/10673220701885815.

McGlothlin, A. E., & Lewis, R. J. (2014). Minimal clinically important difference: Defining what really matters to patients. *JAMA, 312*, 1342–1343. https://doi.org/10.1001/jama.2014.13128.

Norton, P. J., & Price, E. C. (2007). A meta-analytic review of adult cognitive-behavioral treatment outcome across the anxiety disorders. *Journal of Nervous and Mental Disease, 195*, 521–531. https://doi.org/10.1097/01.nmd.0000253843.70149.9a.

Ost, L. G., Havnen, A., Hansen, B., & Kvale, G. (2015). Cognitive behavioral treatments of obsessive-compulsive disorder. A systematic review and meta-analysis of studies published 1993–2014. *Clinical Psychology Review, 40*, 156–169. https://doi.org/10.1016/j.cpr.2015.06.003.

Prien, R. F., Carpenter, L. L., & Kupfer, D. J. (1991). The definition and operational criteria for treatment outcome of major depressive disorder. A review of the current research literature. *Archives of General Psychiatry, 48*, 796–800. https://doi.org/10.1001/archpsyc.1991.01810330020003.

Rahman, N. (1996). Caregivers' sensitivity to conflict: The use of the vignette methodology. *Journal of Elder Abuse and Neglect, 8*, 35–47. https://doi.org/10.1300/J084v08n01_02.

Rush, A. J., Kraemer, H. C., Sackeim, H. A., Fava, M., Trivedi, M. H., Frank, E., Ninan, P. T., Thase, M. E., Gelenberg, A. J., Kupfer, D. J., Regier, D. A., Rosenbaum, J. F., Ray, O., Schatzberg, A. F., & Force, A. T. (2006). Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology, 31*, 1841–1853. https://doi.org/10.1038/sj.npp.1301131.

Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information crite-

Please cite this article as: Kim, Xu and DeRubeis, Mapping Female Patients' Judgments of Satisfaction to Hypothetical Changes in Depression Symptom Severity, Behavior Therapy, https://doi.org/10.1016/j.beth.2021.10.003

rion. *Behavioral Ecology and Sociobiology, 65*, 13–21. https://doi.org/10.1007/s00265-010-1037-6.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin and Review, 11*, 192–196. https://doi.org/10.3758/BF03206482.

Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Attiullah, N., & Boerescu, D. (2006a). How should remission from depression be defined? The depressed patient's perspective. *American Journal of Psychiatry, 163*, 148–150. https://doi.org/10.1176/appi.ajp.163.1.148.

Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Boerescu, D., & Attiullah, N. (2006b). Discordance between self-reported symptom severity and psychosocial functioning ratings in depressed outpatients: Implications for how remission from depression should be defined. *Psychiatry Research, 141*, 185–191. https://doi.org/10.1016/j.psychres.2005.05.016.

Received: February 1, 2021
Accepted: October 12, 2021
Available online: xxxx

Please cite this article as: Kim, Xu and DeRubeis, Mapping Female Patients' Judgments of Satisfaction to Hypothetical Changes in Depression Symptom Severity, Behavior Therapy, https://doi.org/10.1016/j.beth.2021.10.003