Double trouble: Do symptom severity and duration interact to predicting treatment outcomes in adolescent depression?

Lorenzo Lorenzo-Luaces, PhD[*,a], **Natalie Rodriguez-Quintana, BA, MPH**[a], **Allen J. Bailey, BA**[a]

[a]Department of Psychological and Brain Sciences, Indiana University – Bloomington, Indiana

Abstract

Studies suggest that depression severity and duration interact to predict outcomes in depression treatment. To our knowledge, no study has explored this question in a sample with a placebo control, two therapies, and their combination nor with adolescents. We used data from the Treatment of Adolescent Depression Study (N=439), in which adolescent were randomized to placebo (PBO), cognitive-behavioral therapy (CBT), antidepressants medications (MEDs), or their combination (COMB). We explore the interaction between depression severity, chronicity, and treatments (vs. placebo) in predicting outcomes. There was interaction between severity and chronicity when comparing COMB and CBT with PBO, but not MEDs. In non-chronic depression, the effects of CBT were inversely related to severity to the point that CBT appeared iatrogenic with more severe depression. In chronic depression, the effects of CBT did not vary by severity, but the relative effects of COMB grew, being smallest in milder, more dysthymic-like depression, and largest in chronic-severe depression. These findings support calls to classify depression by severity and chronicity as well efforts to risk stratify patients to different intensity of care according to these variables.

Keywords

Personalized medicine; stepped care; depression; risk stratification; adolescents

---

A major depressive episode (MDE) is defined in the Diagnostic and Statistical Manual for Mental Disorders (DSM) by the two-week duration of five depressive symptoms (American Psychiatric Association (APA), 2013). The DSM-5 text states that the duration and severity of depressive presentations are what makes them pathological and distinct from normal sadness (see p. 168). Existing research correlates severity and duration to indices of pathology in depression such as: cognitive biases (Riso et al., 2003; Strunk, Lopez, & DeRubeis, 2006), genetic risk (Kendler, Gardner, & Prescott, 1999; Klein, 1990; Klein, Shankman, Lewinsohn, Rohde, & Seeley, 2004; Lux, Aggen, & Kendler, 2010; Lyons et al., 1998), and overall prognosis (Kessing, 2004; Penninx et al., 2011). Additionally, greater symptom severity and longer episode are among the most reliable predictors of treatment outcomes, both predicting worse outcomes in treatments for depression (Kessler et al.,

---

[*]Corresponding Author: Lorenzo Lorenzo-Luaces: 1101 E 10th St, Bloomington, IN 47405, lolorenz@indiana.edu.

2017). Recognizing the importance of severity and duration, Klein (2008) proposed that the DSM-5 adopt a two-dimensional classification system arraying patients on the basis of severity and duration (see also Ruscio, 2019), though this recommendation was not adopted.

Although patients with severe depression tend to experience worse *overall* outcomes when compared with patients mild depression, differences between active treatments and controls are often more pronounced the more severe depression is (Bower et al., 2013; Driessen, Cuijpers, Hollon, & Dekker, 2010; Fournier et al., 2010; Khan, Leventhal, Khan, & Brown, 2002; Kirsch et al., 2008). For example, in the Treatment for Adolescent Depression Study (TADS; N = 439), adolescents were randomly assigned to a placebo control, a cognitive-behavioral therapy (CBT) condition, antidepressant medications (MEDs), and their combination treatment (COMB). In a recent re-analysis of TADS, we constructed a multivariable prognostic index that predicted overall response as well as a relatively superior response to the combination of CBT and MEDs than to CBT or MEDs alone (Lorenzo-Luaces, Rodriguez-Quintana, Riley, & Weisz, 2019). Although the index was composed of multiple variables, the variables added very little beyond the effects of symptom severity. This is consistent with other studies that find that even in the presence of multiple predictive variables, severity is the most important one (Dinga et al., 2018; Shalev et al., 2019).

Evidence is more mixed regarding the role of duration as a moderator of outcomes. Fournier et al. (2009), for example, found that the 2-year duration of depression or dysthymia, together "chronic" depression, predicted poorer overall outcomes in both CBT and MEDs, relative to non-chronic depression. However, there were no differences between the two treatments across levels of chronicity (Fournier et al., 2009). By contrast, Cuijpers et al. (2010) reported that psychotherapy was less effective than selective serotonin reuptake inhibitors (SSRIs) for chronic depression (i.e., 2 years in adults) and that the combination of pharmacotherapy and psychotherapy was more effective than pharmacotherapy by itself (d = 0.23) and substantially more effective than psychotherapy by itself (d=0.45). However, these authors noted that all studies of SSRIs studied patients with dysthymia (i.e., mild but chronic depression) and suggested this could also account for the pattern of results in comparisons of combination treatment vs. psychotherapy or pharmacotherapy.

Treatment researchers are increasingly becoming interested in moving away from exploring single variables and towards the combination of multiple variables to assign patients to treatments for depression (Cloitre, Petkova, Su, & Weiss, 2016; Delgadillo, Huey, Bennett, & McMillan, 2017; DeRubeis et al., 2014; Kessler et al., 2017; Kraemer, 2013; Lorenzo-Luaces & DeRubeis, 2018; Lorenzo-Luaces, DeRubeis, van Straten, & Tiemens, 2017). The interaction of severity and duration, which should be of clinical interest, has proven to be useful in predicting treatment outcomes in several studies. In a large dataset (N=2,283) Nelson, Delucchi, and Schneider (2014) pooled comparisons of MEDs vs. placebos for late-life depression. They reported a large treatment effect (d=0.70) in depression that was both very chronic (i.e., >10 years) and severe. No statistical or clinically-significant differences emerged for patients with mild or acute depression. Using data from a trial comparing CBT to psychodynamic therapy (PDT; N = 341) Driessen et al. (2016) reported that the combination of MEDs and CBT was more effective (d=0.83) for severe depression that had less than a year's duration than the combination of PDT and MEDs. For severe depression

lasting over a year or more, combination treatment with psychodynamic therapy appeared more effective than CBT but this difference was small (d=−0.31) and not statistically significant. Hollon et al. (2014) tested the effects of severity and duration in a sample of patients (N = 452) with either chronic (2 years) or recurrent depression treated with MEDs or combination treatment with CBT. Combination treatment (81% recovery) was superior to MEDs (52%) if depression was non-chronic, recurrent, and severe. Otherwise, there were no differences between the treatments.

The studies by Nelson et al. (2013), Driessen et al. (2016), and Hollon et al. (2014), suggest that severity and chronicity interact to predict outcomes, though they are difficult to integrate. Not a single one of the studies has both a control condition and combination therapy, making it difficult to make establish conclusions about the efficacy of antidepressants and psychotherapy across the range of depression severity and chronicity. For example, in the Hollon et al. (2014) study there were no differences between combination treatment and MEDs for acute or non-severe depression. From this study, however, it is unclear whether the treatments were equally effective but superior to a credible control (e.g., placebo) or simply ineffective in mild-moderate or chronic depression. In the study by Nelson et al. (2013), while a treatment effect is found in chronic/severe depression, it is unclear whether psychotherapy or the combination of psychotherapy and medications would have had similar effects in chronic severe depression. Finally, while the interaction of severity and duration in predicting outcomes has been explored in adults and older adults, to our knowledge, this question has been unexplored in children and adolescents. To address the gaps in the literature exploring the interactions of symptom severity and duration in predicting outcomes, we explored this interaction in TADS. TADS is a unique sample in which to study predictors of response because it contains a placebo control, a CBT-only condition, a MEDs only condition, and combination treatment (COMB). Moreover, adolescents represent an interesting group in which to explore the interactions of severity and chronicity to predict outcomes because other variables (e.g., prior episodes, prior treatments) are less likely to play a confounding role.

Methods

TADS was a multicenter, randomized controlled trial that evaluated the efficacy of an acute (12-week) phase of: CBT, MEDs (fluoxetine), their combination, and a pill placebo as treatment for adolescents with depression. Several articles have discussed the rationale, design, and methods of the study (March et al., 2004; TADS Team, 2003, 2005). Participants consisted of 439 adolescents, age 12–17. They were recruited on the basis of a diagnosis of MDD on the basis of the Schedule for Affective Disorders and Schizophrenia for School-Aged Children-Present and Lifetime Version (Kaufman et al., 1997), and excluded on the basis of substance use, suicidality, psychosis, or intellectual disability. About half (54%) of participants were female, and 73.8% were Caucasian.

Treatment conditions

CBT (n=111).—Adolescents received a highly manualized and structured skills-based treatment for which the core assumption is that depression is maintained by lack of positive

reinforcement from the environment and negative thinking patterns (Curry, Wells, & Practice, 2005). Adolescents met with a therapist for 50–60 minutes. A thorough description of the CBT delivered in TADS can be found elsewhere (Curry, Wells, & Practice, 2005; March et al., 2004).

MEDs. (n=109).—Adolescents received fluoxetine. The initial dosage consisted of 10 mg/day and increased if necessary up to 40–60 mg/day. Youth met with a pharmacotherapist, who provided medication management and encouragement, weekly for 20–30 minutes.

Combination (COMB, n=107).—Adolescents received all the components of both CBT and medication.

Placebo (PBO, n=112).—Adolescents received a pill placebo instead of fluoxetine. Dosage and meeting structure were equivalent to the medication group.

Depression severity.—TADS used Clinical Global Impression (CGI) ratings as a categorical outcome (i.e., response/non-response). However, given that prior research has cast doubts on the reliability/validity of CGI ratings (Ruhé, Dekker, Peen, Holman, & De Jonghe, 2005), we used the self-reported Reynolds Adolescent Depression Scale (RADS) (Reynolds, 1987; Reynolds, 2004) as our outcome. The RADS is a 30-item self-reported depression severity scale that was completed at baseline, mid-treatment (i.e., week 6), and end of treatment (i.e., week 12). The week 12 (i.e., end of treatment) RADS was our main outcome. It showed the same pattern of results as the CGI: COMB, and MED proved superior to PBO whereas CBT was not. Prior research supports the validity and reliability of the RADS in measuring adolescent depression (Reynolds, 2004). Suggested cut-offs on the scale can be used to classify a none-minimal depressed mood (i.e., 75), mild (i.e., 76–81), moderate (i.e., 82–88), and severe (i.e., 89) depressive symptoms (Reynolds, 2004). In the current sample, the RADS showed excellent reliability (alpha=0.91).

Duration of current MDE.—An independent evaluator interviewed the adolescent and parent with the Schedule for Affective Disorders and Schizophrenia for School-Age Children (Kaufman et al., 1997). As part of this interview, they estimated the duration of the present episode (in weeks). As a continuous variable, duration was not normally distributed. To simplify the presentation of result, we coded duration as either chronic (1 year for adolescents, according to the DSM) or non-chronic (<1 year). All the results reported were also obtained when depression duration was used as a continuous variable transformed as suggested by Templeton (2011).

Missing data

At baseline, rates of missing data on co-variates we selected were low (all <12%, see table 1). Rates of missing outcome data were somewhat higher (i.e., 26% for missing RADS scores mid-treatment). Rate of missingness in mid-treatment RADS appeared to be somewhat lower in the MEDs (18%) than CBT (32%), COMB (28%), or placebo (24%) but this effect was not significant (x^2(3)=5.54, p=0.14). There were no other indications that missingness may differ between the treatments (ps>0.38). As would be expected given that

this was a randomized study, rates of missingness on the baseline co-variates did not vary between treatments (ps>0.24).

To address missing data, including outcome assessments, we used a random forests imputation procedure (see Lorenzo-Luaces et al., 2019) using random forests with the R package missForest (Stekhoven and Bühlmann, 2012). Multivariable imputation models are preferred to completers analyses (Janssen et al., 2009; Moons, Donders, Stijnen, & Harrell Jr, 2006; Sterne et al., 2009; van Kuijk, Viechtbauer, Peeters, & Smits, 2016) as well as last-observation-carried-forward (LOCF) imputations (Kenward & Molenberghs, 2009; Lachin, 2016) because they are more representative of the whole sample, have more power, and are less subject to bias due to the nature of missingness (Janssen et al., 2009; Moons et al., 2006; Sterne et al., 2009; van Kuijk et al., 2016). Before imputation, we winsorized variables that had outliers at three standard deviations over or under the mean. While it is sometimes assumed that study outcomes should not be imputed, this assumes that the relationship between the outcomes and other variables is 0 (Johnson & Young, 2011). The empirical literature on this topic suggests that omitting outcomes from imputation analyses actually can lead to biased statistical estimates at worst (Moons et al., 2006; Young & Johnson, 2010) or is inert at worst (Johnson & Young, 2011).

Analytic plan

Analyses were conducted using R (R Core Team, 2013). We computed bivariate correlations between severity (i.e., baseline RADS), chronicity, and end-of-treatment (i.e., week 12). outcomes. We first present the association between severity and chronicity in predicting end-of-treatment outcomes in the PBO condition, to characterize the effects of severity and chronicity absent active treatment. Our main analysis used a linear model with robust standard errors, with the R package 'estimatr' (Blair et al., 2018), to explore the three-way interaction of baseline depression severity, chronicity, and three dummy variables representing the three active treatments (i.e., CBT, ADM, and COMB) vs. placebo, as well as their lower-order effects. A question that often arises when presenting the results of statistical interactions is how to probe the nature of the interactions. Researchers often explore interactions at the mean and +/− 1 standard deviation, or by picking a specific point of interest (Hayes & Matthes, 2009). These approaches, however, are arbitrary. The Johnson-Neyman technique is an empirical alternative that can be used to obtain values of the moderator variable (e.g., symptom severity) at which the focal predictor (i.e., treatment) transitions from statistically significant predictor of outcomes (p 0.05) to not (p>0.05). While it is not possible to find a Johnson-Neyman significance region across three different statistical tests (i.e., for the three treatments) in a three-way interactions, we used the Johnson-Neyman technique, with the 'jtools' package , to explore values of severity at which the treatment contrast may no longer be significant in chronic and non-chronic depression.

We presented the predicted effects of treatment (vs. placebo) at different levels of severity for chronic vs. non-chronic depression, along with their 95% confidence intervals (CIs). Then, we selected the values suggested by the Johnson-Neyman technique to explore treatment differences at specific cut-offs. We calculate standardized mean differences

(SMDs) at the severity value suggested by the Johnson-Neyman technique, for chronic vs. non-chronic depression. Additionally, we wanted to rule out the possibility that our results were confounded by age (e.g., older adolescents have more time to be depressed) or baseline-functioning. Thus, we also explored the interactions between severity and age (which could confound duration) as well as between duration and baseline-functioning (which could confound severity) and which was measured with the Children's Global Assessment of Functioning (Shaffer et al., 1983).

Results

At the beginning of treatment, the baseline RADS score was 79.20 (SD=13.87). The average depressive episode lasted 71.59 weeks (SD=82.35, Me=40, IQR=20–100); 45% of adolescents had been depressed for a year or more. At the end of treatment, the mean RADS score was 62.31 (SD=15.50). Combination treatment was superior to placebos (B=−8.69, SE=2.04, t=−4.27, p<0.001) as were MEDs (B=−4.92, SE=2.03, t=−2.43, p=0.02) but CBT was not (B=0.86, SE=2.02, t=0.43, p=0.67). Participants with a chronic course (M=79.26, SD=14.40) did not report higher baseline symptoms than those with a non-chronic course (M=79.15, SD=13.46, t(406.64)=−0.08, p=0.93). Across the treatment groups, higher baseline severity predictor poorer end of treatment outcomes (r_{437}=0.45, 95% CI=0.37–0.52, p<0.001) but chronic status did not (r_{437}=0.03, 95% CI= −0.06–0.13, p=0.47).

Placebo condition

To illustrate the relations between chronicity and baseline severity in relation to end-of-treatment outcomes, we show in Figure 1 the relationship between these variables in the placebo condition. There was an interaction between severity and chronicity in the placebo condition, suggesting that the more severe and chronic depression was, the worse outcomes patients had (B=0.34, SE=0.12, 95% CI=0.11–0.58, t(108)=2.89, p=0.005). When considering the effects of severity on outcomes, baseline severity was a stronger predictor of outcomes in chronic depression, predicting poorer outcomes (r_{41}=0.68, 95% CI=0.48–0.81, p<0.001), than in non-chronic depression (r_{67}=0.37, 95% CI=0.15–0.56, p=0.002). Taken together, these data suggest that absent of treatment (i.e., in the placebo condition) the effects of severity and chronicity are multiplicative.

Treatments vs. placebo

There was an interaction between severity and chronicity in predicting end-of-treatment outcomes in combination treatment (B=−0.52, SE=0.25, t(423)=−2.03, p=0.04) and CBT (B=−0.45, SE=0.20, t(423)=−2.28, p=0.03) but the interaction between severity, chronicity, and MED was smaller and not statistically significant (B=−0.34, SE=0.24, t(423)=−1.47, p=0.14). We illustrate these effects in Figure 2, which plots the predicted effects of the treatments (relative to placebo), across the ranges of severity for chronic and non-chronic depression. Based on the Johnson-Neyman regions of significance in COMB (vs. PBO), we explored treatment contrasts between chronic and non-chronic depression according to baseline RADS scores of 70 ("minimal") and 88 ("severe"), values which are also close to the interquartile range.

Chronic depression.—The effects of CBT did not appear to vary across the spectrum of severity being relatively indistinguishable from placebo for patients with milder symptoms (B=0.71, SE=2.36, t=0.30, p=0.77, SMD=−0.05, 95% CI=−0.32–0.21) as well as more severe ones (B=−0.23, SE=2.67, t =−0.09, p=0.93, SMD=0.02, 95% CI=−0.25–0.28). However, the effects of COMB (vs. PBO) on chronic depression appeared to vary substantially by severity. They were much more pronounced in severe depression (B= −12.28, SE=3.55, t=−3.46, p<0.001, SMD=0.92, 95% CI=0.62–1.20) than for milder symptoms (B=−6.04, SE= 3.02, t=−2.00, p=0.046, SMD=0.44, 95% CI=0.17–0.71).

Non-chronic depression.—The pattern of treatment by severity interactions differed across the treatments. For CBT, there was an effect of severity such that the treatment was indistinguishable from placebo at milder levels (B=−0.65, SE=2.82, t=−.23, p =0.82, SMD=0.05, 95% CI=−0.22–0.31) but actually appeared iatrogenic at severe levels in non-chronic depression (B=6.59, SE=2.65, t=2.49, p=0.01, SMD=−0.48, 95% CI=−0.75–−0.20). In combination treatment, there was no effect of severity with the treatment producing medium-to-large effects in severe depression (B=−9.01, SE=2.55, t=−3.53, p<0.001, SMD=0.44, 95% CI=0.17–0.71) and milder symptoms (B=−5.94, SE= 3.17, t=−1.88, p=0.06, SMD=0.67, 95% CI= 0.39–0.94).

Sensitivity analyses

To explore whether the severity by chronicity by treatment condition interaction was accounted for by age, we added an interaction term of severity by age by treatments, and the lower order terms, as well as an interaction of baseline-functioning by duration by treatments, as well as the lower order terms. There was no interaction between age, severity, and treatment conditions in predicting outcomes (ps>0.64). The interactions between severity, chronicity, and CBT (B=−0.37, SE=0.20, t(407)=−1.83, p=0.07) as well as COMB (B=−0.44, SE=0.26, t(407)= −1.69, p=0.09) crossed the p<0.05 threshold but were similar in magnitude to what we reported previously. Interestingly, in this model, we also found evidence for interactions between baseline-functioning, chronicity, and CBT (B=−0.62, SE=0.34, t(407)=−1.82, p=0.07) and COMB (B=−0.98, SE=0.43, t(407)=−2.30, p=0.02). Because we did not hypothesize that this relationship existed, and the model contained nine three-way interactions as well as their lower-order effects, we did not explore it further. Additionally, to explore whether there were other variables that differed by the interaction between severity and chronicity, we regressed a number of variables on depression severity, chronicity, and their interaction. These variables included various measures of treatment expectancy, baseline anxiety, suicidality, psychosocial stress, trauma history, and IQ. The interaction between severity and chronicity did not predict any of these baseline variables (ps>0.12), suggesting that they are unlikely to confound the observed interaction between severity, chronicity, and treatment.

Discussion

As other studies (Driessen et al., 2016; Hollon et al., 2014; Nelson et al., 2013), we found a significant interaction between depression severity and duration in predicting outcomes, this time with combination therapy or CBT vs. placebo, in adolescents. These findings did not

appear to be accounted for by age or baseline-functioning, nor were the combinations of severity and chronicity predictive of other potential confounds. In the placebo condition of TADS, severity and chronicity interacted as one would expect such that outcomes were worse the more chronic and severe depression became. This is consistent with clinical intuition that a more prolonged and severe course of illness should predict negative outcomes as well as prior observations regarding the prognostic value of severity and chronicity (Lorenzo-Luaces, 2015). Consequently, the largest treatment-control difference, that between COMB and PBO, was observed in patients with chronic-severe depression.

We also found different treatment effects at the intersections of the chronicity and severity continua. The general pattern of findings reported in the trial was that COMB was superior to MEDs, which were superior to CBT and PBO, the latter two groups being comparable in outcomes. Two patient subgroups appear to be an exception to this pattern of findings. First, in non-chronic depression, CBT actually appeared iatrogenic at high levels of depression severity. Even though this finding might seem odd, it is not surprising given CBT's overall effect on the TADS study (March et al., 2004; TADS Team, 2003, 2005). In fact, our finding might help explain the original findings on why CBT produced lower response rates than expected in TADS (i.e., CBT performed very poorly in this patient subgroup). No such effect was seen in COMB, suggesting that the addition of medication to CBT helps potentiate its effects in subgroups that are at high risk of non-response or deterioration. In chronic depression, the only effective treatment at the lower end of severity (i.e., more "dysthymic" depression) was COMB, CBT, MEDs, and PBO were indistinguishable from each other. Although COMB was still the superior treatment in chronic non-severe depression, in this patient subgroup combination treatment had the lowest effects. As severity increases, the value of receiving COMB, but not CBT, also increases suggesting the MEDs and CBT may have a synergistic effect in chronic depression (Cuijpers et al., 2010).

Strengths and limitations

Several limitations of the study should be considered before interpreting the results. First, the TADS study was not powered to detect a three-way interaction. Thus, it remains possible that in a larger sample we would observe somewhat different relations. The three-way interaction between antidepressants (vs. placebo), symptom severity, and duration was not statistically significant at $p < .05$ but appeared similar in direction and magnitude to the interaction in the other conditions and may have been significant in a larger sample. Similarly, the slope of the effect of CBT across levels of severity seemed significant in chronic depression, though the confidence intervals crossed zero, which would suggest no effect. Larger samples may reveal an effect of CBT (vs. a control) for adolescents with chronic depression.

Despite issues with power, this study is large by depression treatment standards (Barth et al., 2016). Another feature of the study is noteworthy: the patients were all adolescents and most (86%) were in their first episode of depression. It could be the case that this pattern of results applies only to first-episode, early onset depressions. In addition to exploring variability in severity and duration as predictors of treatment response, future research should explore prior episodes and age of onset, both of which are etiologically-relevant variables. TADS is

unique as a depression treatment study in that it is one of the few studies containing CBT, MEDs, their combination, and neither (i.e., a placebo). Nonetheless, CBT did not outperform PBO, which has not been a consistent finding in adolescent depression research. Some authors have opined that the CBT delivered in TADS was too restrictive and prescriptive (see Hollon, Garber, & Shelton, 2005), limiting efficacy. Finally, it should be noted that the current study is also limited by the exclusion criteria which prohibited patients with substance use disorders and psychosis from entering the trial. These exclusions are common exclusions across studies of different treatments for depression (Lorenzo-Luaces, Johns, & Keefe, 2018a; Lorenzo-Luaces, Zimmerman, & Cuijpers, 2018b). Another common exclusion criterion that was used in this trial was for patients with even milder symptoms than we explored here. It is possible that with more variability in the sample's severity scores, the interaction effects we reported would be more dramatic.

Implications

The most interesting patterns in the data come from the contrast of the treatment conditions in non-chronic but severe depression as well as the treatment in non-severe but chronic depression. In chronic non-severe depression, CBT and MEDs did not separate from placebo and COMB had its smallest effect. This observation is consistent with prior studies that suggest combination treatment is superior to CBT and MEDs in dysthymia (i.e., chronic non-severe depression), with a more pronounced difference between COMB and psychotherapy (Cuijpers et al., 2010). Interestingly, in individuals with high severity non-chronic depression, CBT actually appeared iatrogenic. This is inconsistent with the findings by Hollon et al. (2014), who reported large effects of COMB vs. MEDs in severe non-chronic depression. It is difficult to reconcile their findings with ours because, by design, all patients in the Hollon et al. study who were non-chronic had recurrent depression.

It is possible that episodes of brief but severe depression in youth represent failures of emotion regulation. We explored variables that could be potential confounders of our results and failed to find a variable that accounted for the observed effect (e.g., stages of change, treatment expectancy). However, we did not have access to any variable measuring severe emotion dysregulation. Fournier et al. (2009) found that patients with predominant symptoms of "Cluster B" personality disorders had the poorest outcomes in CBT (e.g., 14% response), though these patients experienced relatively good outcomes with MEDs (67% response). This is consistent with Linehan's (1993) assertion that "traditional" CBT is ineffective for individuals with severe emotion dysregulation. Our findings suggest that the addition of medication in this patient subgroup may thus be necessary. Additionally, it is also possible that this subgroup of patients with non-chronic severe depression is highly sensitive to the quality of treatment, hence the large treatment effects across studies (Driessen et al., 2016; Hollon et al., 2014)

The interaction between severity, chronicity, and MEDs was not significant at $p<0.05$. Nonetheless, we observed that the largest effects of the medications were seen in patients with chronic and severe depression (SMD=0.46, 95% CI=0.19–0.73) with the smallest effects seen for chronic but non-severe depression (SMD = 0.01, 95% CI –0.25 – 0.27). This mimics the findings by Nelson et al. (2013) and suggests an interactive effect of severity and

chronicity in predicting positive treatment outcomes in MEDs, though these authors explored a very high threshold of chronicity (i.e., 10 years). In chronic and severe depression, the synergistic effects of adding CBT and MEDs appeared evident.

Our overall pattern of results adds to the literature suggesting that severity and chronicity interact to predict outcomes, especially contrasts between treatments that differ in potency (e.g., combination therapy vs. a monotherapy). These findings support the argument by Klein (2008) that depression severity and course are important determinants of outcomes in the treatment of depression and that a two-dimensional system may be more sensible than the current diagnostic approach. Perhaps most interestingly, at certain extremes of the severity by chronicity spectrum (e.g., chronic but mild depression, severe but non-chronic depression), different treatment patterns emerged. One could speculate about whether these extremes (i.e., more dysthymic-like depression vs. severe but non-chronic depression) represent entirely different illnesses, a notion that existed in ancient writing on depression (Shorter, 2007). Because the literature on depression chronicity and severity is mired by issues of restriction of range (Lorenzo-Luaces, 2018; Lorenzo-Luaces et al., 2018a), it is difficult to draw solid conclusions about the relationship between these variables and outcomes. In terms of treatment recommendations, however, several things appears clear. The superiority of combination treatment is especially marked in chronic and severe depression. Chronic but milder forms (e.g. dysthymia) appear the most difficult to treat, though here combination treatment has an advantage. In severe but non-chronic depression, the treatment differences are extreme with CBT singly appearing to be iatrogenic and combination treatment outperforming CBT, MEDs, and PBO.

Data from the TADS study suggest that, in adolescent depression, combination treatment appears to be the most effective treatment. Prior (see Hollon et al., 2005) and subsequent studies (Davey et al., 2019) have been more optimistic regarding the efficacy of CBT for adolescent depression. Our findings may be taken to suggest that the poor performance of CBT in TADS may have been driven by very negative outcomes in acute but non-severe depression. Overall, our findings suggest that severity and duration are two variables that can be used to guide treatment decisions in adolescent depression. Future research should explore whether specific symptoms dimensions account for this pattern of findings (Khazanov et al., 2020; McMakin et al., 2012).

Acknowledgements

References

American Psychiatric Association (APA). (2013). Diagnostic and Statistical Manual of Mental Disorders (Dsm 5). In: American Psychiatric Publishing.

Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, … Cuijpers P. (2016). Comparative Efficacy of Seven Psychotherapeutic Interventions for Patients with Depression: A Network Meta-Analysis. Focus, 14(2), 229–243. [PubMed: 31997951]

Blair G, Cooper J, Coppock A, Humphreys M, Sonnet L, Fultz N, & Blair MG (2018). Package 'Estimatr'. Stat, 7(1), 295–318.

Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, … Christensen H (2013). Influence of Initial Severity of Depression on Effectiveness of Low Intensity Interventions: Meta-Analysis of Individual Patient Data. Bmj, 346, f540. [PubMed: 23444423]

Cloitre M, Petkova E, Su Z, & Weiss BJ (2016). Patient Characteristics as a Moderator of Posttraumatic Stress Disorder Treatment Outcome: Combining Symptom Burden and Strengths. BJPsych open, 2(2), 101–106. [PubMed: 27703762]

Cuijpers P, van Straten A, Schuurmans J, van Oppen P, Hollon SD, & Andersson G (2010). Psychotherapy for Chronic Major Depression and Dysthymia: A Meta-Analysis. Clinical psychology review, 30(1), 51–62. [PubMed: 19781837]

Curry JF, Wells KCJC, & Practice B (2005). Striving for Effectiveness in the Treatment of Adolescent Depression: Cognitive Behavior Therapy for Multisite Community Intervention. 12(2), 177–185.

Davey CG, Chanen AM, Hetrick SE, Cotton SM, Ratheesh A, Amminger GP, … Harrison BJ (2019). The Addition of Fluoxetine to Cognitive Behavioural Therapy for Youth Depression (Yoda-C): A Randomised, Double-Blind, Placebo-Controlled, Multicentre Clinical Trial. The Lancet Psychiatry, Advanced online publication.

Delgadillo J, Huey D, Bennett H, & McMillan D (2017). Case Complexity as a Guide for Psychological Treatment Selection. Journal of consulting and clinical psychology, 85(9), 835. [PubMed: 28857592]

DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, & Lorenzo-Luaces L (2014). The Personalized Advantage Index: Translating Research on Prediction into Individualized Treatment Recommendations. A Demonstration. PloS one, 9(1), e83875. [PubMed: 24416178]

Dinga R, Marquand AF, Veltman DJ, Beekman AT, Schoevers RA, van Hemert AM, … Schmaal L (2018). Predicting the Naturalistic Course of Depression from a Wide Range of Clinical, Psychological, and Biological Data: A Machine Learning Approach. Translational psychiatry, 8(1), 241. [PubMed: 30397196]

Driessen E, Cuijpers P, Hollon SD, & Dekker JJ (2010). Does Pretreatment Severity Moderate the Efficacy of Psychological Treatment of Adult Outpatient Depression? A Meta-Analysis. Journal of consulting and clinical psychology, 78(5), 668. [PubMed: 20873902]

Driessen E, Smits N, Dekker J, Peen J, Don F, Kool S, … Van H (2016). Differential Efficacy of Cognitive Behavioral Therapy and Psychodynamic Therapy for Major Depression: A Study of Prescriptive Factors. Psychological medicine, 46(4), 731–744. [PubMed: 26750445]

Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, & Fawcett J (2010). Antidepressant Drug Effects and Depression Severity. JAMA: The Journal of the American Medical Association, 303(1), 47–53. [PubMed: 20051569]

Fournier JC, DeRubeis RJ, Shelton RC, Hollon SD, Amsterdam JD, & Gallop R (2009). Prediction of Response to Medication and Cognitive Therapy in the Treatment of Moderate to Severe Depression. Journal of consulting and clinical psychology, 77(4), 775. [PubMed: 19634969]

Hayes AF, & Matthes J (2009). Computational Procedures for Probing Interactions in Ols and Logistic Regression: Spss and Sas Implementations. Behavior research methods, 41(3), 924–936. [PubMed: 19587209]

Hollon SD, DeRubeis RJ, Fawcett J, Amsterdam JD, Shelton RC, Zajecka J, … 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(10), 1157–1164. [PubMed: 25142196]

Hollon SD, Garber J, & Shelton RC (2005). Treatment of Depression in Adolescents with Cognitive Behavior Therapy and Medications: A Commentary on the Tads Project. Cognitive and Behavioral Practice, 12(2), 149–155.

Janssen KJ, Vergouwe Y, Donders ART, Harrell FE, Chen Q, Grobbee DE, & Moons KG (2009). Dealing with Missing Predictor Values When Applying Clinical Prediction Models. Clinical chemistry, 55(5), 994–1001. [PubMed: 19282357]

Johnson DR, & Young R (2011). Toward Best Practices in Analyzing Datasets with Missing Data: Comparisons and Recommendations. Journal of Marriage and Family, 73(5), 926–945.

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, … Ryan N, J. J. o. t. A. A. o. C. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-Sads-Pl): Initial Reliability and Validity Data. Journal of the American Academy of Child Adolescent Psychiatry, 36(7), 980–988. [PubMed: 9204677]

Kendler KS, Gardner CO, & Prescott CA (1999). Clinical Characteristics of Major Depression That Predict Risk of Depression in Relatives. Archives of General Psychiatry, 56(4), 322–327. [PubMed: 10197826]

Kenward MG, & Molenberghs G (2009). Last Observation Carried Forward: A Crystal Ball? Journal of biopharmaceutical statistics, 19(5), 872–888. [PubMed: 20183449]

Kessing LV (2004). Severity of Depressive Episodes According to Icd-10: Prediction of Risk of Relapse and Suicide. The British Journal of Psychiatry, 184(2), 153–156. [PubMed: 14754828]

Kessler R, Van Loo H, Wardenaar K, Bossarte R, Brenner L, Ebert D, … Sampson N (2017). Using Patient Self-Reports to Study Heterogeneity of Treatment Effects in Major Depressive Disorder. Epidemiology and psychiatric sciences, 26(1), 22–36. [PubMed: 26810628]

Khan A, Leventhal RM, Khan SR, & Brown WA (2002). Severity of Depression and Response to Antidepressants and Placebo: An Analysis of the Food and Drug Administration Database. Journal of clinical psychopharmacology, 22(1), 40–45. [PubMed: 11799341]

Khazanov GK, Xu C, Dunn BD, Cohen ZD, DeRubeis RJ, & Hollon SD (2020). Distress and Anhedonia as Predictors of Depression Treatment Outcome: A Secondary Analysis of a Randomized Clinical Trial. Behaviour Research and Therapy, 125, 103507. [PubMed: 31896529]

Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, & Johnson BT (2008). Initial Severity and Antidepressant Benefits: A Meta-Analysis of Data Submitted to the Food and Drug Administration. PLoS medicine, 5(2), e45. [PubMed: 18303940]

Klein DN (1990). Symptom Criteria and Family History in Major Depression. American Journal of Psychiatry, 147(7), 850–854.

Klein DN (2008). Classification of Depressive Disorders in the Dsm-V: Proposal for a Two-Dimension System. Journal of abnormal psychology, 117(3), 552. [PubMed: 18729608]

Klein DN, Shankman SA, Lewinsohn PM, Rohde P, & Seeley JR (2004). Family Study of Chronic Depression in a Community Sample of Young Adults. American Journal of Psychiatry, 161(4), 646–653.

Kraemer HC (2013). Discovering, Comparing, and Combining Moderators of Treatment on Outcome after Randomized Clinical Trials: A Parametric Approach. Statistics in Medicine, 32(11), 1964–1973. [PubMed: 23303653]

Lachin JM (2016). Fallacies of Last Observation Carried Forward Analyses. Clinical trials, 13(2), 161–168. [PubMed: 26400875]

Lorenzo-Luaces L (2015). Heterogeneity in the Prognosis of Major Depression: From the Common Cold to a Highly Debilitating and Recurrent Illness. Epidemiology and psychiatric sciences, 24(6), 466–472. [PubMed: 26081748]

Lorenzo-Luaces L (2018). Representing the Heterogeneity of Depression in Treatment Research. Acta Psychiatrica Scandinavica, 138(4), 360. [PubMed: 30062768]

Lorenzo-Luaces L, & DeRubeis RJ (2018). Miles to Go before We Sleep: Advancing the Understanding of Psychotherapy by Modeling Complex Processes. Cognitive Therapy and Research, 42(2), 212–217.

Lorenzo-Luaces L, DeRubeis RJ, van Straten A, & Tiemens B (2017). A Prognostic Index (Pi) as a Moderator of Outcomes in the Treatment of Depression: A Proof of Concept Combining Multiple Variables to Inform Risk-Stratified Stepped Care Models. Journal of affective disorders, 213, 78–85. [PubMed: 28199892]

Lorenzo-Luaces L, Johns E, & Keefe JR (2018a). The Generalizability of Randomized Controlled Trials of Self-Guided Internet-Based Cognitive Behavioral Therapy for Depressive Symptoms: Systematic Review and Meta-Regression Analysis. Journal of medical Internet research, 20(11).

Lorenzo-Luaces L, Rodriguez-Quintana N, Riley T, & Weisz JR (2019). A Placebo Prognostic Index (Pi): Could It Inform Riskstratification in Treatment with Cognitive-Behavioral Therapy, Fluoxetine, or Their Combination? Manuscript submitted for publication.

Lorenzo-Luaces L, Zimmerman M, & Cuijpers P (2018b). Are Studies of Psychotherapies for Depression More or Less Generalizable Than Studies of Antidepressants? Journal of Affective Disorders, 234, 8–13. [PubMed: 29522947]

Lux V, Aggen S, & Kendler K (2010). The Dsm-Iv Definition of Severity of Major Depression: Inter-Relationship and Validity. Psychological Medicine, 40(10), 1691–1701. [PubMed: 20056021]

Lyons MJ, Eisen SA, Goldberg J, True W, Lin N, Meyer JM, … Tsuang MT. (1998). A Registry-Based Twin Study of Depression in Men. Archives of General Psychiatry, 55(5), 468–472. [PubMed: 9596050]

March J, Silva S, Petrycki S, Curry J, Wells K, Fairbank J, … Vitiello B (2004). Fluoxetine, Cognitive-Behavioral Therapy, and Their Combination for Adolescents with Depression: Treatment for Adolescents with Depression Study (Tads) Randomized Controlled Trial. Jama, 292(7), 807–820. [PubMed: 15315995]

McMakin DL, Olino TM, Porta G, Dietz LJ, Emslie G, Clarke G, … Birmaher B (2012). Anhedonia Predicts Poorer Recovery among Youth with Selective Serotonin Reuptake Inhibitor Treatment–Resistant Depression. Journal of the American Academy of Child & Adolescent Psychiatry, 51(4), 404–411. [PubMed: 22449646]

Moons KG, Donders RA, Stijnen T, & Harrell FE Jr (2006). Using the Outcome for Imputation of Missing Predictor Values Was Preferred. Journal of clinical epidemiology, 59(10), 1092–1101. [PubMed: 16980150]

Nelson JC, Delucchi KL, & Schneider LS (2013). Moderators of Outcome in Late-Life Depression: A Patient-Level Meta-Analysis. American Journal of Psychiatry, 170(6), 651–659.

Penninx BW, Nolen WA, Lamers F, Zitman FG, Smit JH, Spinhoven P, … der Meer K. v. (2011). Two-Year Course of Depressive and Anxiety Disorders: Results from the Netherlands Study of Depression and Anxiety (Nesda). Journal of Affective Disorders, 133(1), 76–85. [PubMed: 21496929]

R Core Team. (2013). R: A Language and Environment for Statistical Computing.

Reynolds WM (1987). Professional Manual for the Reynolds Adolescent Depression Scale.: Psychological Assessment Resources, Inc.

Reynolds WM (2004). Reynolds Adolescent Depression Scale – 2nd Edition. In Hilsenroth SM (Ed.), Comprehensive Handbook of Psychological Assessment (Vol. Volume 2: Personality assessment and psychopathology, pp. 224–236). New York: John Wiley & Sons.

Riso LP, Du Toit P, Blandino JA, Penna S, Dacey S, Duin JS, … Ulmer CS. (2003). Cognitive Aspects of Chronic Depression. Journal of Abnormal Psychology, 112(1), 72. [PubMed: 12653415]

Ruhé HG, Dekker JJ, Peen J, Holman R, & De Jonghe F (2005). Clinical Use of the Hamilton Depression Rating Scale: Is Increased Efficiency Possible? A Post Hoc Comparison of Hamilton Depression Rating Scale, Maier and Bech Subscales, Clinical Global Impression, and Symptom Checklist-90 Scores. Comprehensive Psychiatry, 46(6), 417–427. [PubMed: 16275208]

Ruscio AM J. A. r. o. c. p. (2019). Normal Versus Pathological Mood: Implications for Diagnosis. 15, 179–205.

Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, & Aluwahlia S (1983). A Children's Global Assessment Scale (Cgas). Archives of General psychiatry, 40(11), 1228–1231. [PubMed: 6639293]

Shalev AY, Gevonden M, Ratanatharathorn A, Laska E, van der Mei WF, Qi W, … Delahanty D (2019). Estimating the Risk of Ptsd in Recent Trauma Survivors: Results of the International Consortium to Predict Ptsd (Icpp). World psychiatry, 18(1), 77–87. [PubMed: 30600620]

Shorter E (2007). The Doctrine of the Two Depressions in Historical Perspective. Acta Psychiatrica Scandinavica, 115, 5–13.

Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, … Carpenter JR (2009). Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls. Bmj, 338, b2393. [PubMed: 19564179]

Strunk DR, Lopez H, & DeRubeis RJ (2006). Depressive Symptoms Are Associated with Unrealistic Negative Predictions of Future Life Events. Behaviour Research and Therapy, 44(6), 861–882. [PubMed: 16126162]

TADS Team. (2003). Treatment for Adolescents with Depression Study (Tads): Rationale, Design, and Methods. Journal of the American Academy of Child & Adolescent Psychiatry, 42(5), 531–542. [PubMed: 12707557]

TADS Team. (2005). The Treatment for Adolescents with Depression Study (Tads): Demographic and Clinical Characteristics. Journal of the American Academy of Child & Adolescent Psychiatry, 44(1), 28–40. [PubMed: 15608541]

van Kuijk SM, Viechtbauer W, Peeters LL, & Smits L (2016). Bias in Regression Coefficient Estimates When Assumptions for Handling Missing Data Are Violated: A Simulation Study. Epidemiology, Biostatistics and Public Health, 13(1).

Young R, & Johnson DR (2010). Imputing the Missing Y's: Implications for Survey Producers and Survey Users. Paper presented at the Proceedings of the AAPOR conference abstracts.
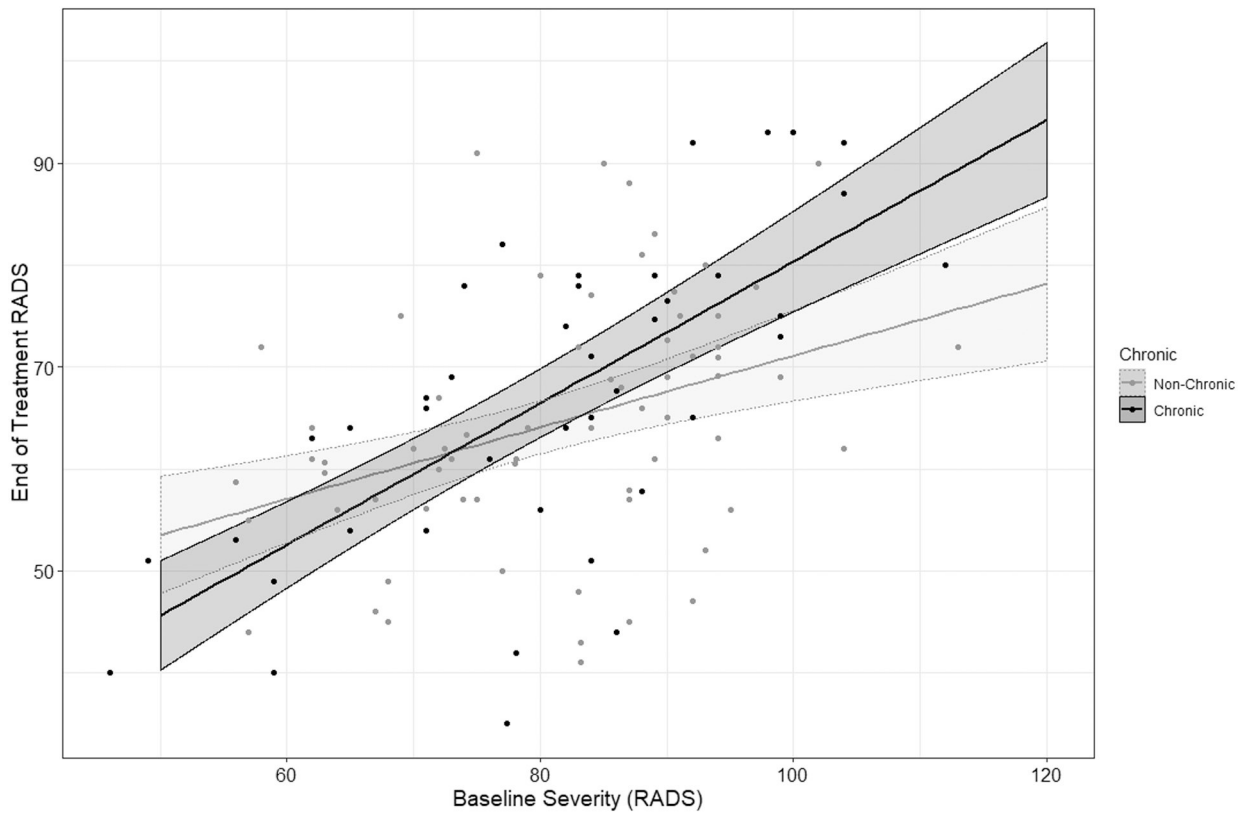
Figure 1. Estimated end of treatment (i.e., week 12) Reynolds Adolescent Depression Scale (RADS) for adolescents by baseline RADS severity for chronic and non-chronic depression
Note. Dark lines indicated predicted slope. Bands around the slope indicate 95% confidence intervals (CIs)
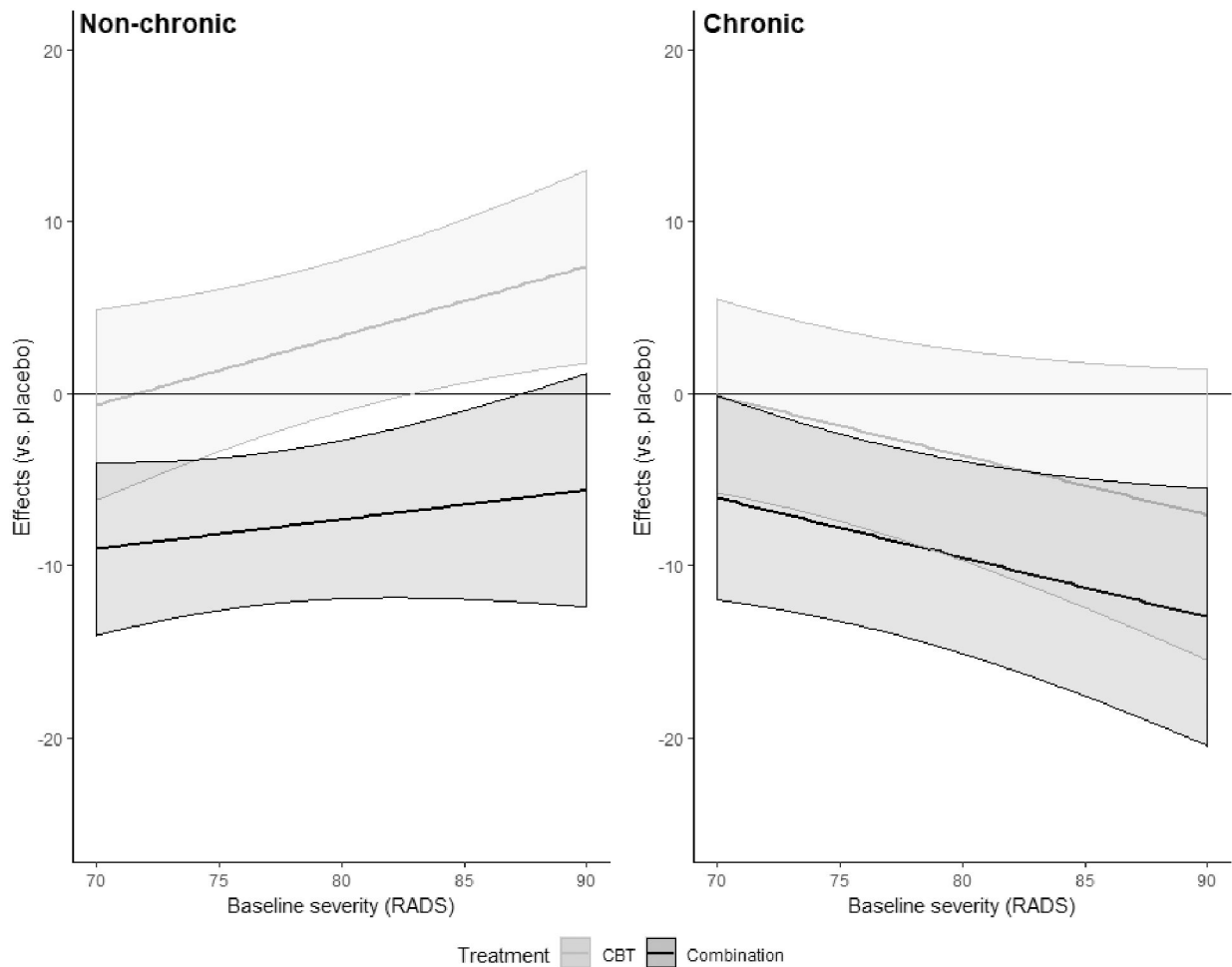
Figure 2. Predicted effects of cognitive-behavioral therapy (CBT) and combination of CBT and fluoxetine, vs. placebo, by chronicity and severity

Note. Figure indicates the effects of treatment vs. placebo, by severity, in non-chronic and chronic depression. Y-axis plots predicted difference between treatment conditions and the placebo conditions. Lower scores are better and indicate greater differences in end of treatment depression scores on the Reynolds Adolescent Depression Scale (RADS), favoring the active treatments. Bands indicate 95% confidence interval. The horizontal line indicates an effect of 0. When CIs cross zero, it indicates the treatment is no different than placebo at that level of severity.

Table 1.

Demographic characteristics of adolescents in the Treatment for Adolescent Depression Study (TADS; n= 439) with missing or imputed data

	Unimputed			Imputed		
	M	SD	% miss	M	SD	SMD
RADS						
Baseline	79.18	13.94	11.0%	79.28	14.44	−0.01
Mid-treatment	66.31	15.28	26.0%	66.82	16.72	−0.03
End of treatment	62.32	15.54	20.0%	62.10	16.54	0.01
Age	14.61	1.54	0.0%	14.61	1.54	0.00
Functioning (CGAS)	49.65	7.29	3.0%	49.64	7.37	0.00
Suicidality (SIQ)	23.81	21.46	6.0%	23.68	21.95	0.01
Treatment expectations (TEA)	2.37	0.64	9.0%	2.37	0.67	0.01
Action stage (SOC)	13.64	2.38	7.0%	13.65	2.47	0.00
Maternal BDI	12.71	9.29	11.0%	12.63	9.83	0.01
Somatic anxiety (MASC)	6.25	3.86	3.0%	6.25	3.90	0.00
Life stress (PLES)	3.73	3.46	10.0%	3.69	3.63	0.01
Vocabulary IQ (WISC)	10.61	2.45	4.0%	10.61	2.50	0.00
Sleep problems (CDRS)	3.71	1.52	0.0%	3.71	1.52	0.00
Sleep problems (RADS)	2.95	1.05	3.9%	2.95	1.07	0.00
Illness or disability (HON 6)	2.10	1.00	1.0%	2.10	1.00	0.00
Social relations (HON 10)	0.34	0.47	1.0%	0.34	0.47	0.00
	n	%	% miss	n	%	%diff
Female (vs. Male)	239	54.44%	0.0%	239	54.44%	0.00%
White (vs. non-White)	355	80.87%	0.0%	355	80.87%	0.00%
Parents live together (yes/no)	250	56.95%	0.2%	250	56.95%	0.00%
Serious trauma (yes/no)	122	27.79%	0.0%	122	27.79%	0.00%
Anxiety co-morbidity (yes/no)	113	25.74%	10.9%	113	25.74%	0.00%
Income			3.2%			
0 – $19,000	236	53.76%		244	55.58%	−1.82%
$20,000 – $39, 999	101	23.01%		103	23.46%	−0.46%
$40,000 and over	88	20.05%		92	20.96%	−0.91%

Note. M = mean, SD = standard deviation, % miss = percentage of missing data, SMD = standardized mean difference between imputed and unimputed data, % diff = absolute difference in percentage between imputed and unimputed data, Reynolds Adolescent Depression Scale (RADS), Children's Depression Rating Scale (CDRS), Children`s Global Assessment Scale (CGAS), Suicidal Ideation Questionnaire (SIQ), Average Treatment Expectancy across CBT, ADM, and COMBO, Stages of Change (SOC), Beck Depression Inventory-II (BDI), Mulitdimensional Anxiety Scale for Children (MASC), Pediatric Life Events Screen (PLES), Health of Nations Scale (HON)