Time-Varying Moderation of Treatment Outcomes by Illness Duration and Comorbid Depression in Generalized Anxiety Disorder

Michelle G. Newman, Ki Eun Shin, and Stephanie T. Lanza
Pennsylvania State University

Objective: To extend the sparse literature on moderators, we used time-varying effect modeling (TVEM; Tan, Shiyko, Li, Li, & Dierker, 2012) to examine how depressive symptoms and duration of generalized anxiety disorder (GAD) moderated effects of 3 treatments for GAD (applied relaxation [AR], cognitive– behavioral therapy [CBT], and nondirective therapy [ND]) over time using intensive repeated measures.

Method: In a secondary analysis of Borkovec and Costello (1993), 66 GAD clients were randomly assigned to AR (n = 23), CBT (n = 23), or ND (n = 20). Clients received 12 therapy sessions over 6 weeks, and after 2 weeks of posttreatment assessment, had 2 additional weekly fading sessions. They completed thrice daily anxiety ratings during this 10-week period. GAD duration (Anxiety Disorders Interview Schedule-Revised) and depressive symptoms (Hamilton Depression Rating Scale) were assessed at baseline. Results: Longer GAD duration predicted less anxiety reduction in CBT and ND relative to AR. These effects were pronounced in the later phase of treatment, suggesting benefits of focused relaxation practice for clients with longer duration. Higher depression predicted better response to CBT than AR and ND. The moderation effects were also more noticeable in the later phase. In multilevel analyses, a similar moderation pattern held at 1-year follow-up on clinician-rated measures.

Conclusion: GAD clients with long-standing symptoms may benefit more from repeatedly practicing fewer skills than learning multiple skills. On the other hand, clients with comorbid depression may respond better to CBT than AR, perhaps because CBT includes cognitive interventions that can generalize to depression.

What is the public health significance of this article? This study suggests benefits of repeated practice of fewer coping skills for GAD patients with longer illness duration, and exposure to diverse skills, especially cognitive techniques, for GAD patients with higher depressive symptoms.

Keywords: generalized anxiety disorder, applied relaxation, cognitive–behavioral therapy, personalized treatment, time-varying effect modeling
overall bodily tension. On the other hand, CBT for GAD can include not only AR components but also cognitive therapy and self-control desensitization (SCD; Goldfried, 1971). In cognitive therapy, clients learn to identify biases in their thoughts, weigh evidence for and against such thoughts, and replace them with more logical thoughts (i.e., cognitive restructuring). The goal of cognitive therapy is to learn to think flexibly and diffuse catastrophic cognitions (Newman, 2000). In self-control desensitization, clients first relax, then imagine themselves in a worrisome situation. Once they elicit anxiety, they continue to imagine themselves in the situation but use coping techniques to reduce their anxiety (Newman, 2000).

One way to enhance treatment outcomes in GAD is to tailor treatments to individual clients (Cuijpers, Ebert, Acarturk, Andersson, & Crisstea, 2016). Such personalized interventions require both selecting and implementing techniques in a way that maximizes efficacy for a specific client. The first step toward achieving that aim is identifying moderators of outcomes in existing treatments (e.g., Paul, 1967). When AR and CBT for GAD were compared without accounting for potential moderation, they were equally efficacious (Borkovec & Costello, 1993; Dugas et al., 2010). However, prior controlled trials used random assignment to equalize baseline characteristics across treatments. Any null findings from these studies might be masking patterns of specificity in which some clients responded better to one treatment than the other because of baseline individual differences. Thus, examining moderation of treatment effects is necessary to inform how to individualize treatments in GAD.

The existing literature on moderation of GAD treatment outcomes is sparse. Only a few moderators have been examined, including duration of the GAD diagnosis (Newman & Fisher, 2013), worry severity (Westra et al., 2009), attachment styles (Newman, Castonguay, Jacobson, & Moore, 2015), and types of interpersonal problems (Newman, Jacobson, Erickson, & Fisher, 2017). Surprisingly, none of the moderators compared AR and CBT. Two studies (Newman et al., 2015; Westra et al., 2009) examined moderation effects in additive designs, comparing CBT with augmented CBT. Newman et al. (2015) compared CBT plus supportive listening and CBT plus interpersonal and emotional processing therapy. Patients with higher dismissive attachment responded better to CBT plus interpersonal and emotional processing therapy than CBT plus supportive listening, whereas those with higher angry attachment responded better to the latter. Westra et al. (2009) compared regular CBT with CBT plus motivational interviewing and found that higher worry severity predicted greater symptom improvement in augmented CBT. The remaining studies used dismantling designs, comparing CBT with its components. Both Newman et al. (2017) and Newman and Fisher (2013) compared CBT with cognitive therapy and self-control desensitization.

In the former study, patients with greater intrusiveness or dominance responded better to behavioral therapy (self-control desensitization) than cognitive therapy or CBT. In the latter study, longer GAD duration predicted better outcomes from single-component interventions (cognitive therapy, self-control desensitization) than CBT.

These prior findings were based on single studies and await replication and extension. One innovative way to replicate and extend past findings is to examine time-varying moderation effects using intensive longitudinal data (i.e., many repeated assessments over time). Typical moderation analyses focus on identifying characteristics linked to better treatment outcomes, on average, across time. However, examining how individual differences effect change over the course of different treatments can elucidate not only which treatment works for whom but also when moderation effects emerge, strengthen, or weaken during treatment. Therefore, such approach can facilitate more nuanced understanding of treatment moderation effects and provide insights on potential mechanisms of the moderation. Of the existing analytic tools for examining change over time in intensive longitudinal data, time-varying effect modeling (TVEM; Tan, Shiyko, Li, Li, & Dierker, 2012) is especially suited for this purpose. TVEM uses a spline-based approach to model regression intercepts and coefficients as smooth, flexible functions of time rather than fixed estimates. By specifying interaction terms to be time-varying, dynamic moderation effects can be examined. Past studies (e.g., Shiyko, Burkhalter, Li, & Park, 2014; Wright, Hallquist, Swartz, Frank, & Cyranoowski, 2014) have shown that TVEM can be more sensitive to detecting group differences than more traditional methods such as multilevel modeling.

TVEM has been applied to psychotherapy and prevention research, addressing questions such as how the association between depression and anxiety changes over time during depression treatment (Wright et al., 2014), how baseline nicotine dependence and momentary negative affect lead to varying urge to smoke over time after smoking-cessation interventions (Lanza, Vasilenko, Liu, Li, & Piper, 2014), and how daily stress and negative affect and their associations change over the course of mindfulness-based stress reduction (Snippe, Dziak, Lanza, Nylilieck, & Wichers, 2017). TVEM can be used to address substantive questions regarding the temporal course of change in treatment effects (e.g., when a treatment effect starts to become active or diminish), time-varying moderation of treatment effects (e.g., how clients with certain characteristics respond to treatment over time), and covariation between psychological phenomena of interest (e.g., when two symptom dimensions start to be decoupled during treatment).

Of the previously tested moderators of GAD treatment, duration of GAD diagnosis is a promising candidate for applying TVEM. In Newman and Fisher (2013), clients with longer GAD duration showed greater symptom reduction in response to single-component interventions (cognitive therapy, self-control desensitization) compared with CBT. This was the case even though the original study found no differences in the overall efficacy of these interventions (Borkovec, Newman, Pincus, & Lyle, 2002). The moderation effect was fully mediated by increasing flexibility in daily and interdaily oscillations of anxiety, with chronic GAD patients showing greater flexibility in single-component interventions than CBT. Such results suggest that focused interventions requiring more intensive practice of fewer skills may have been more effective for changing entrenched patterns of generalized anxiety (in those with longer duration) than being exposed to a diverse set of coping skills, as was the case for CBT (Newman & Fisher, 2013).

Another moderator to consider is depression. Despite high comorbidity between GAD and depression (e.g., Kessler, Berglund, et al., 2005), moderating effects of comorbid depression in GAD treatment have yet to be examined. Comorbid major depressive disorder predicted a lower probability of naturalistic remission from GAD over the course of 8 years (Bruce et al., 2005) and less
symptom reduction from brief psychodynamic and supportive non-directive therapies for GAD (Crum Christof et al., 2004). Interestingly, comorbid depressive symptoms predicted better or comparable outcomes from either individual (Newman, Przeworski, Fisher, & Borkovec, 2010) or group (Wetherell et al., 2005) CBT, suggesting that CBT is helpful for those with comorbid depression. Nonetheless, because Newman et al. (2010) averaged across CBT and single-component treatments (SCD and CT), and neither study examined effects of comorbid depression in AR, it remains untested whether depression symptoms would predict differential outcomes in GAD between CBT and AR.

There is a reason to believe, however, that CBT would be more effective than AR in GAD patients with comorbid depression. CBT includes cognitive therapy, which has been empirically supported to treat both anxiety and depression (Chambless & Ollendick, 2001). In addition, given the overlap between worry and rumination (e.g., Szkodny & Newman, 2017), cognitive interventions directly targeting worry in GAD may generalize to rumination, a symptom commonly found in depression (e.g., Nolen-Hoeksema, 2000). In fact, a meta-analysis showed lower overall efficacy for relaxation treatment than for CBT in treating depression (Jorm, Morgan, & Hetrick, 2008). Moderation by comorbid depressive symptoms may be examined as a time-varying effect to determine whether and when depression predicts better outcomes from CBT than AR. Specifically, it remains to be tested whether a moderation effect begins to be observed after cognitive therapy is introduced in CBT.

The present study, a secondary analysis of Borkovec and Costello (1993), aimed to extend the sparse literature on moderation of GAD treatment outcomes using a novel analytic approach that allows for testing time-varying moderation effects. We are not aware of another study that has applied TVEM to examine whether different treatments for GAD led to varying outcomes over time contingent on patients’ baseline characteristics. In the original study (Borkovec & Costello, 1993), AR and CBT showed better outcomes than nondirective (ND) therapy at posttreatment on GAD symptom measures, including the Hamilton Anxiety Rating Scale (HARS), the Penn State Worry Questionnaire, the State–Trait Anxiety Inventory, and the Clinician’s Severity Rating (Cohen’s d = −.48 to −.92 for AR vs. ND; −.91 to −1.06 for CBT vs. ND), with no differences between the two. CBT and AR also showed better maintenance of gains compared with ND at 1-year follow-up (Cohen’s d = −.13 to −.33 for AR vs. ND; −.19 to −.60 for CBT vs. ND).

We applied TVEM to thrice daily anxiety ratings of GAD patients across 10 weeks of treatment. We hypothesized that longer GAD duration would predict better outcomes from AR than ND or CBT, and these effects would be more prominent in the later phase of treatment, reflecting the benefit of repeated practice of fewer skills for GAD patients reporting longer duration of illness. We also predicted that higher depressive symptoms would be associated with greater anxiety reduction from CBT than ND or AR, and these effects would emerge after cognitive therapy was introduced, which could generalize to depressive symptoms. Because TVEM focused on self-reported ratings of anxiety from pretreatment to fading sessions, we also ran supplementary multilevel analyses to test whether a similar pattern of moderation was replicated on clinician-rated GAD symptom measures from pretreatment to a 1-year follow-up. Although moderation between AR and CBT was our primary interest, we still retained ND, the control condition, in analyses to contextualize the moderation between AR and CBT.

Method

Participants and Procedure

Five hundred eight clients were recruited from agencies and news advertisements, and 442 were ruled out because they did not meet criteria for GAD. Diagnosis was made using a structured diagnostic interview (the Anxiety Disorders Interview Schedule-Revised [ADIS-R]; Di Nardo & Barlow, 1988) based on the Diagnostic and Statistical Manual of Mental Disorders criteria (3rd ed., rev.; DSM–III-R; American Psychiatric Association, 1987). A pair of trained independent assessors conducted interviews within a period of a week. Clients were included if both assessors agreed that GAD was the principal diagnosis and rated its severity as equal to or greater than 4 (moderate on a 0- to 5-point scale). Exclusion criteria were meeting panic disorder criteria, currently receiving other psychotherapy, having prior exposure to therapies provided in the study, having medical conditions that could contribute to anxiety symptoms, taking antidepressant medication, and having severe depression, substance abuse, psychosis, or organic brain syndrome.

Among the remaining 66 clients, 43 were females and 23 were males. Sixty participants identified as White, three as Black, two as Hispanic, and one as Native American. Mean age was 37 years (SD = 13.01). Participants were randomly assigned to ND, which was a control condition (ND, n = 20), AR (n = 23), or CBT (n = 23). Based on pretreatment assessment, two participants (one in AR and one in CBT) were diagnosed with comorbid major depressive disorder, with clinician’s severity ratings of 4. Eight additional participants (two in ND, three in AR, and three in CBT) were diagnosed with dysthymia at pretreatment, with severity ratings between 3 and 4. Four clients in each condition were on anxiety medication, but their dosage was held constant during the study. Over the course of treatment, eight clients dropped out early and three were removed from the treatment because of exacerbating depressive symptoms. The number of dropouts did not significantly differ between treatment conditions, χ²(2, N = 66) = 2.99, p = .23. The average number of sessions completed by the dropouts was 5.09 (range = 1 to 11). In addition, gender and ethnicity were not significantly different across therapy conditions, χ²(2, N = 66) = .17, p = .92, and χ²(6, N = 66) = 6.81, p = .34, respectively. There was also no significant gender or ethnicity difference based on dropout status, χ²(1, N = 66) = .47, p = .49, and χ²(3, N = 66) = 2.20, p = .53, respectively. Neither treatment condition nor dropout status significantly predicted age, F(2, 62) = .41, p = .66, r² = .01, and F(1, 63) = .60, p = .44, r² = .01, respectively.

Clients participated in 12 therapy sessions, twice per week for 6 weeks. Then, there was a 2-week posttreatment assessment, followed by two fading sessions occurring in the subsequent 2 weeks. The first four therapy sessions took 90 min each, and remaining sessions were 60 min. ND used an exploratory approach with an emphasis on supportive listening. AR involved psychoeducation, self-monitoring of anxiety, and learning and repeated practice of relaxation techniques including diaphragmatic breathing and progressive muscle relaxation. CBT included AR techniques (introduced in Sessions 1 through 3) as well as self-control desensitization (SCD; introduced in Session 4) and cognitive therapy (introduced in Session 5).
Measures

Client daily diary. Across the 10 weeks of treatment, clients made thrice daily (morning, afternoon, and evening) ratings of anxiety level for the preceding period on a 0 to 100 scale (0 = complete relaxation, 100 = maximum discomfort). Ratings occurred across 70 days (compliance was 72%), leading to a maximum of 210 data points (average = 151 ratings/participant). Two-week retest reliability was .80 and convergent and discriminant validity were demonstrated in Newman and Fisher (2013).

Duration of generalized anxiety disorder. Duration of GAD symptoms was assessed through clients’ self-report as a part of the ADIS-R diagnostic interviews at pretreatment. In the GAD section, clients reported how long the tension, anxiety, and worry had been a problem for them. In case a client responded with “all my life,” assessors inquired further by asking the client to anchor the onset of their symptoms to specific life events or periods (e.g., “since first grade”). Of 66 clients, five clients (7.6% of the sample) reported having symptoms throughout their lives even with further prompting, and their age was coded as their duration of GAD symptoms. Interrater reliability of self-reported GAD duration between two diagnostic interviews with different assessors at pretreatment was good (intraclass correlation coefficient [ICC] = .84, 95% CI [.73, .91]).

Hamilton Anxiety Rating Scale (Hamilton, 1959). The HARS is a clinician-rated measure with 14 items rated on a 5-point Likert scale ranging from 0 (absent) to 4 (severe), reflecting affective, cognitive, and somatic symptoms of anxiety. Scores range from 0 to 56, with ≤17 suggesting mild anxiety, 18 to 24 suggesting mild to moderate anxiety, and 25 to 30 for moderate to severe severity. The measure has good interrater reliability (e.g., Bruss, Gruenberg, Goldstein, & Barber, 1994; ICC = .89 in the present study). It was administered at pretreatment and at 6- and 12-month follow-ups by trained interviewers who were blind to treatment condition.

Clinician’s Severity Rating (CSR; Di Nardo & Barlow, 1988). The CSR has a 0 (none) to 8 (very severely disturbing/disabling) Likert scaled score assigned by trained interviewers blind to treatment condition to reflect degree of impairment associated with GAD as part of the ADIS-R administered at pre- and posttreatment, 6-month, and 1-year follow-up. Interrater reliability of CSRs in the current study was good (ICCs = .77–1.0).

Hamilton Depression Rating Scale (HDRS; Riskind, Beck, Brown, & Steer, 1987). The original HDRS is a 17-item interviewer-rated measure of depression (Hamilton, 1960). We rescored the HDRS based on the revised 12-item version, which removed content overlap with anxiety symptoms and thus, increased its specificity to depression. Three items were rated on a 0 (absent) to 2 (clearly present) scale, and nine items were on a 0 (absent) to 4 (severe) scale. Possible scores ranged from 0 to 42. The measure has good internal consistency (α = .73; α = .71 in the present sample) and discriminant validity between depression and GAD (Riskind et al., 1987). The HDRS was administered by the first of the two diagnostic interviewers at pretreatment and who was blind to treatment condition.

Analytic Approach

Dropouts and those removed from treatment were included in all analyses. We used TVEM (Li et al., 2015) to examine time-varying moderation of treatment effects by self-reported duration of GAD diagnosis and comorbid depressive symptoms (HDRS scores) between ND, AR, and CBT for GAD. TVEM uses a spline-based approach in modeling regression intercepts and coefficients as smooth functions of time. A complex function of time is divided into many intervals based on the number of knots (dividing points). The specific shape of change for each interval is estimated using polynomial terms. Because TVEM uses a nonparametric approach to estimate intercept and slope functions, it does not impose a specific shape of change (e.g., linear, curvilinear). Therefore, it is especially suited to capturing complex and nonlinear patterns of change compared with usual analytic methods for intensive longitudinal data (e.g., multilevel modeling).

We ran two TVEM models, first using ND as a reference group to compare AR and CBT with ND, and then using AR as a reference group to compare between AR and CBT. Treatment conditions were coded using dummy variables. The following is the model using ND as the reference group:

Anxiety_{ij} = \beta_0(t) + \beta_1(t) \times AR_i + \beta_2(t) \times CBT_i + \beta_3(t) \times duration_i + \beta_4(t) \times AR_i \times duration_i + \beta_5(t) \times CBT_i \times duration_i + \beta_6(t) \times HDRS_i + \beta_7(t) \times AR_i \times HDRS_i + \beta_8(t) \times CBT_i \times HDRS_i + e_{ij}

Anxiety_{ij} was thrice daily anxiety rating for subject i at assessment j measured at time t. To facilitate interpretation, HDRS scores were standardized (M = 0, SD = 1), and GAD duration was grand-mean centered. Given the broad range of GAD duration in the current sample (0.58 to 61.67 years), duration was also divided by 10 so that 10 years represented one unit of this variable. Beta estimates (e.g., \beta_6(t), \beta_7(t)) were modeled as flexible, smooth coefficient functions of time. The coefficient \beta_6(t) was an intercept function describing mean thrice daily anxiety over time for participants in the ND condition (i.e., AR = 0; CBT = 0) with average levels of GAD duration (i.e., duration = 0) and depressive symptoms (i.e., HDRS = 0). The coefficients \beta_1(t) and \beta_2(t) indicated time-varying effects of AR and CBT relative to ND on thrice daily anxiety at the mean level of both moderators (GAD duration, depression symptoms). The coefficients \beta_3(t) and \beta_4(t) each indicated time-varying effects of duration and depression on thrice daily anxiety for individuals in the ND condition while holding effects of the other moderator constant. Coefficient functions of most interest were \beta_5(t) + \beta_6(t) + \beta_7(t) + \beta_8(t), which represented time-varying moderation of AR and CBT’s efficacy relative to ND on thrice daily anxiety by GAD duration and depressive symptoms. A random error term (e_{ij}) was also included.

Analyses used SAS macro suite, %TVEM (TVEM SAS Macro, 2017, Version 3.1.1). TVEM offers two spline-based estimation methods: a penalized truncated power spline (P-spline) and an unpenalized B-spline (Li et al., 2017). Whereas P-spline automatically selects number of knots (splitting points for splines) based on information criteria to determine the optimal model, B-spline requires a user to select the number of knots by manually comparing models with different numbers of knots in an iterative fashion for each time-varying effect. Given that our model included multiple time-varying estimates, using B-spline would have required very extensive model selection in an iterative fashion for each of the time-varying estimates. Therefore, we used a P-spline method as in prior studies (e.g., Lanza, Vasilenko, & Russell, 2016). Following the recommendation of Li et al. (2017), we set 10 knots as the maximum for each
time-varying estimate to allow sufficient complexity. To account for the nesting of assessments within participants, robust standard errors were used.

Another important feature of TVEM is that it treats time as a continuous variable and accommodates for unequal spacing of measurements (Li et al., 2017; Tan et al., 2012), which was the case for our thrice daily anxiety ratings (morning, afternoon, and evening). TVEM also deals with missing data by using all available data for each participant at each time point. We coded time of assessment in real time (e.g., 10:00 a.m., 2:00 p.m.), and estimates were produced based on the assumption that the shape of change in regression coefficients was a smooth curve. This is a reasonable assumption because treatment and moderation effects are unlikely to change radically within a few hours across a day. Given that the goal of the study was to examine the pattern of weekly, session-to-session changes, not within-process days, thrice daily ratings provided good coverage of the time axis despite differences in spacing between assessments. We therefore used the raw data from the client diaries (not daily averages) for TVEM analyses.

The %TVEM macro provides intercept and slope estimates and 95% CIs across continuous time. Because TVEM results are continuous and cannot be summarized into discrete statistics, we represented them graphically by plotting the estimated intercept and slope functions and corresponding 95% pointwise CIs over time. Three lines were drawn for each coefficient of interest, one for the estimate, and the other two for the upper and lower limits of the estimate’s CI. To facilitate interpretation of findings, we plotted the time-varying trajectory of daily anxiety and the time-varying effects of GAD duration and depression on daily anxiety by treatment condition. If a CI for daily anxiety did not contain zero at a given time point, it meant that daily anxiety at the time was significantly different from zero (at the mean levels of GAD duration and depression). Similarly, if a CI for the effect of GAD duration or depression did not include zero at a time point, the effect was significantly different from zero. If CIs for the effect of GAD duration or depression did not overlap between treatment conditions at a time point, this indicated significant moderation of treatment effects between conditions. In addition, if CIs for a predictor at two different time points did not overlap, the effect of the predictor was significantly different between the two times. We refer to specific pointwise estimates in the results section to facilitate precise interpretation, reporting both unstandardized (b) and standardized coefficients (β). It is also of note that TVEM CIs tend to be wider at the beginning and end of a time series because there are fewer available data points. Therefore, coefficient estimates for those time points provide less certainty.

In addition, we conducted supplementary multilevel modeling analyses. Given that the TVEM model focused on moderation of treatment effects from pre- to posttreatment based on clients’ self-reported anxiety, we tested whether a similar pattern of results held for pretreatment to follow-up change in symptoms assessed through clinician-rated measures. Multilevel analyses were run in Mplus 8.1, employing full information maximum likelihood (FIML) estimation to handle missing data (Muthén & Muthén, 2017). For the outcome, we used a composite of the HARS and the CSR, created by standardizing and averaging the two scores. Using a composite helps to increase reliability and validity of assessment (Cronbach, 1950; Jensen & McFarland, 1993; Nunnally & Bernstein, 1994) as well as to lower experiment-wise Type I error rate relative to running separate analyses for single measures (Horowitz, Inouye, & Siegelman, 1979).

Missing data was only present in the outcome (15%) and not predictors. Because there were two follow-up assessments, at 6 months and 1 year, time was modeled as a continuous predictor. Fixed effects included intercept, time, treatment condition (dummy variables), moderators, and interactions between variables. As in TVEM analyses, both moderators (GAD duration, HDRS scores) were entered simultaneously. Intercept and time were treated as random effects. To make comparisons between each treatment condition (ND, AR, and CBT), we ran analyses twice, first using ND as a reference group, and then using AR as a reference group. Cohen’s d was calculated following Wolf (1986) and Dunst, Hamby, and Trivette (2004).

Results

Descriptive Statistics and Baseline Differences

The sample had mild to moderate anxiety on average based on pretreatment HARS (M = 20.12, SD = 4.69). Average pretreatment CSR was 4.80 on a 0 to 8 scale (SD = .63), consistent with the inclusion criteria of ≥4. HARS and CSR scores were highly positively correlated across the four assessment points, r(168) = .88, p < .001, justifying the use of a composite. The pretreatment composite anxiety score was not significantly different between treatment conditions, F(2, 63) = .39, p = .68, ƞp² = .01, or between treatment completers and dropouts, F(1, 64) = .52, p = .47, ƞp² = .01.

Average self-reported duration of GAD was 16.85 years (SD = 16.07) and ranged from .58 to 61.67 years. Mean HDRS score was 11.60 (SD = 3.67, range = 4.5–22.5), comparable with the average HDRS scores observed in other GAD samples (Riskind et al., 1987). GAD duration was correlated neither with the CSR, r(66) = −.08, p = .52, nor with HDRS scores, r(66) = .02, p = .87. Client age was correlated with GAD duration, r(66) = .53, p < .001. Also, GAD duration and depressive symptoms were not significantly different between treatment conditions, F(4, 114) = 1.38, p = .25, ƞp² = .05, or between treatment completers and dropouts, F(2, 56) = 2.43, p = .10, ƞp² = .08.

Time-Varying Moderation of Treatment Effects by GAD Duration and Depression

Figure 1 shows the time-varying trajectory of daily anxiety by treatment condition. Over time, AR showed significant reduction in daily anxiety (Day 1: β = .31, b = 35.06, 95% CI [31.82, 38.29]; Day 70: β = −.31, b = 21.32, 95% CI [16.02, 26.62]), whereas daily anxiety in those who received ND (Day 1: β = .21, b = 31.03, 95% CI [28.99, 33.08]; Day 70: β = .26, b = 31.98, 95% CI [29.22, 34.74]) and CBT (Day 1: β = .22, b = 31.64, 95% CI [28.42, 34.86]; Day 70: β = −.09, b = 25.90, 95% CI [21.37, 30.43]) did not change significantly. However, given that average anxiety level in CBT at Day 67 was significantly lower than the beginning anxiety (β = −.14, b = 24.72, 95% CI [21.32, 28.12]), the nonsignificant change between Days 1 and 70 in those who received CBT might be caused by flaring of the CIs at the edge of time series. When comparing between conditions, AR led to significantly lower daily anxiety than ND starting in Week 2 and throughout the treatment period. AR also showed lower daily anxiety than CBT between Weeks 7 and 9. CBT exhibited lower daily anxiety than ND between Week 2 and Week 5 and during fading sessions (Weeks 9–10).
When the effect of GAD duration on daily anxiety was examined, it showed a different pattern across treatment conditions (see Figure 2). In those receiving ND, longer GAD duration predicted higher anxiety starting in Week 1 and until the end of fading sessions. This effect continued to rise over time, with the effect significantly stronger during fading sessions (Week 10) relative to Week 1, Day 3 (Week 1): $\beta = .10, b = 1.11$, 95% CI [.06, 2.16]; Day 70 (Week 10): $\beta = .38, b = 4.09$, 95% CI [2.69, 5.49]. In those receiving CBT, longer GAD duration also predicted higher anxiety between Weeks 2 and 4 and during Weeks 6 through 10. GAD duration did not have a significant effect on daily anxiety in those receiving AR throughout the treatment period. At the same time, longer GAD duration predicted less anxiety reduction in ND relative to AR during Weeks 4 through 10. This difference between ND and AR became significantly stronger in the later phase, peaking at the end of fading sessions, Day 23 (Week 4): $\beta = -.12, b = -1.24$, 95% CI [−2.03, −.45]; Day 70 (fading sessions): $\beta = -.43, b = -4.64$, 95% CI [−7.16, −2.12]. Longer GAD duration also predicted less anxiety reduction in those receiving CBT relative to AR during Weeks 2, 3, and 7 through 10. The
difference between CBT and AR also peaked during fading sessions, although not significantly differently from the earlier phase of treatment, Day 8 (Week 3): $\beta = .17, b = 1.91, 95\%$ CI $[.02, 3.80]$; Day 70 (fading sessions): $\beta = .54, b = 6.18, 95\%$ CI $[1.65, 10.70]$. CIs for CBT and ND overlapped throughout the treatment period, indicating no significant moderation by GAD duration between the two conditions.1

The effect of comorbid depression on daily anxiety also varied across treatment conditions over time (see Figure 3). In ND, more severe depression predicted higher daily anxiety throughout treatment, except for Weeks 2 through 4. In AR, higher depression also predicted greater daily anxiety throughout treatment, except the first 3 days. However, CBT showed a different pattern in which more severe depression predicted lower daily anxiety during post-treatment and fading sessions (Weeks 8–9). In fact, higher depression predicted less anxiety reduction in those receiving AR relative to CBT starting in Week 5 and during most of treatment period (except the last 3 days, likely because of imprecision of estimates). This moderation effect became stronger over time, peaking during fading sessions (Day 58 [fading sessions]: $\beta = -.47, b = -8.66, 95\%$ CI $[-10.79, -6.54]$) and was significantly stronger during fading sessions than the earlier phase (Day 36 [Week 5]: $\beta = -.23, b = -4.29, 95\%$ CI $[-5.90, -2.68]$). Higher depression also predicted less anxiety reduction in ND relative to CBT during Week 1 and Weeks 5 through 10, except the last 3 days. Between Weeks 5 and 10, the moderation effect increased over time, peaking during Week 9, Day 31 (Week 5): $\beta = -.11, b = -1.82, 95\%$ CI $[-3.39, -2.5]$; Day 57 (fading sessions): $\beta = -.41, b = -9.07, 95\%$ CI $[-10.98, -6.15]$. When ND and AR were compared, depression led to significantly higher daily anxiety in those receiving ND relative to AR during the first 4 days, but this effect became nonsignificant for the rest of the treatment period.2

Multilevel Analyses: Pretreatment to Follow-Up

Moderation of Treatment Effects

In a multilevel model testing moderation of treatment effects on clinician-rated anxiety measures from pretreatment to 1-year follow-up, there was significant moderation by GAD duration. Longer GAD duration predicted more symptom reduction at 1-year follow-up in response to AR relative to CBT ($\beta = -.04, standard error [SE] = .02, p = .009, d = -.64$) but not relative to ND ($\beta = -.02, SE = .02, p = .30, d = -.26$). Moderation by

---

1 Because age was correlated with GAD duration, we ran supplementary analyses to examine whether the effects of GAD duration were attributable to the effects of age. When we included both GAD duration and age as time-varying moderators in one TVEM model, the pattern of results on GAD duration stayed the same. In addition, results on age were distinct from what we observed for GAD duration. Whereas longer GAD duration predicted better outcomes in AR than CBT and ND, older age predicted better outcomes for both AR and CBT relative to ND.

2 Although we present results from a model including both moderators, we also ran models separately for each moderator given our modest sample size. Results were consistent with what we present here. We ran a supplementary analysis including time of day for each assessment as a time-invariant covariate. There was a significant fixed effect for time of day ($\beta = -5.46, t [9137] = -4.54, p < .001$), but including the covariate did not change the results of TVEM analyses. Multilevel analyses were also run using the daily diary data as an outcome in Mplus. Missing data was addressed using the FIML estimation. To improve the plausibility of the missing at random assumption of maximum likelihood estimation (Collins, Schafer, & Kam, 2001; Silvia, Kwapiel, Eddington, & Brown, 2013), we controlled for time of day in multilevel analyses. Moderation effects by GAD duration and depression were not significant ($p > .1$), except that GAD duration predicted less anxiety reduction in ND than AR at a marginal level ($\beta = -.06, SE = .03, p = .08, d = -.43$). Results are consistent with the previous findings (e.g., Shiuko et al., 2014; Wright et al., 2014), which showed that TVEM tends to be more sensitive to detecting group differences than multilevel modeling. We therefore chose to interpret TVEM diary findings and not the multilevel model diary findings.
GAD duration was not significant between CBT and ND (β = −.02, SE = .01, p = .9, d = −.42; see Figure 4). Meanwhile, comorbid depression also significantly moderated treatment effects between CBT and ND such that higher depression predicted greater symptom reduction in response to CBT relative to ND (β = −.06, SE = .03, p = .03, d = −.52). Moderation was not significant between CBT and AR (β = −.007, SE = .03, p = .98, d = −.02) or between AR and ND (β = −.06, SE = .03, p = .09, d = −.42; see Figure 4).

Discussion

The primary aim of this study was to investigate whether duration of GAD and comorbid depressive symptoms moderated treatment outcomes between AR, CBT, and ND for GAD and how moderation effects changed over time. Hypotheses on time-varying moderation were largely supported. Patients who reported longer duration of GAD symptoms responded better to AR than CBT and ND, and this moderation effect was more salient in the later phase of treatment (posttreatment, fading sessions). In addition, patients with higher depressive symptoms responded better to CBT than AR and ND, and this moderation became more pronounced over time, peaking during fading sessions. A similar pattern of results also emerged when pretreatment to follow-up symptom change was examined using clinician-rated symptom measures.

When looking at the individual treatments, in ND, both longer GAD duration and higher comorbid depressive symptoms predicted less reduction in anxiety. Such findings are inconsistent with results from supportive expressive therapy wherein GAD duration did not predict outcome. However, comorbid depression did negatively predict outcome from this treatment (Cris-Christoph et al., 2004). What our findings add, however, is a better picture of symptom change during the treatment period for a nondirective supportive treatment (albeit not supportive-expressive therapy). Such results highlighted that those with higher levels of baseline depression did not do as well as those with lower depression from ND in the initial and later phases of treatment, and those with longer duration of GAD were increasingly worse off over the course of ND treatment compared with those with shorter duration of the illness.

We also found that depressive symptoms predicted a better response to CBT than AR and ND, and increasingly so over time with the moderation effects peaking during fading sessions. This finding can be interpreted in the context of prior evidence that CBT and AR may operate through different mechanisms of change. In a prior study comparing CBT and AR for GAD (Donegan & Dugas, 2012), reduction in worrisome cognitions had a greater influence on subsequent reduction in somatic anxiety in CBT than AR. Thus, change in negative repetitive thinking may be a more important mechanism in CBT than AR. This is consistent with the theorized mechanism for why CBT would be more effective than AR for GAD clients with higher depression. Anxiety and depression both involve repetitive negative thoughts (e.g., anticipation of danger and harm, thoughts of failure and hopelessness; Beck, Brown, Steer, Eidelson, & Riskind, 1987). Cognitive therapy, which is a core element of CBT, may be better than AR at targeting such perseverative negative thoughts. The pronounced moderation between AR and CBT in the later phase of treatment may have resulted from the effects of cognitive therapy (which is introduced in Week 3). However, there are some caveats because other interventions (AR, self-control desensitization) preceded cognitive therapy in CBT. Therefore, although cognitive therapy was the most distinctive feature of CBT versus AR (the two other CBT interventions were relaxation-based), we cannot rule out the possibility that the observed moderation effect was driven by a combination of the interventions. Furthermore, in the case of the ND versus CBT comparison, the increasing trend in the moderation effect from Week 5 to Week 10 might have been related to the

3 In multilevel follow-up analyses including both age and GAD duration as moderators, age did not significantly moderate treatment effects, but the moderation between CBT and AR was still significant (with ND falling nonsignificantly between AR and CBT). Overall, the moderating effects of GAD duration cannot be attributed to the effects of age.
nonsignificant effect of depression on daily anxiety in ND between Week 2 and Week 4. If not for the unexpected and potentially spurious nonsignificance in ND between Week 2 and Week 4, moderation by depression between CBT and ND might have been more time-invariant. It is also noteworthy that the moderation effect was temporarily significant during Week 1 before becoming nonsignificant in Weeks 2 to 4. This initial moderation might have been driven by therapeutic effects of the treatment rationale delivery in the first session of CBT (e.g., Ahmed & Westra, 2009). Overall, the present findings should be considered preliminary, and replication is needed to clarify the underlying mechanisms of the depression moderation effects.

In the GAD literature, research on mechanisms of change for AR has been more equivocal than for CBT. Although reduction in muscle tension was considered the primary mechanism of AR, findings (e.g., Conrad, Isaac, & Roth, 2008) have not consistently supported this. Additional mechanisms have been proposed, including increased mindfulness (e.g., Hayes-Skelton, Usmani, Lee, Roemer, & Orsillo, 2012) and habituation to relaxation-induced anxiety (Newman, Lafreniere, & Jacobson, 2018). It is difficult to pinpoint exact mechanisms of change, but our findings suggest that AR was at least more effective for GAD patients with longer duration of the disorder than CBT. Evidence for the AR versus ND comparison was more mixed, with longer GAD duration predicting greater anxiety reduction from AR relative to ND during the posttreatment period (TVEM) but not at 1-year follow-up (multilevel analyses). In fact, ND outperformed CBT for clients with longer GAD duration at 1-year follow-up at a marginally significant level (although the moderation was no longer present once the effects of age were controlled for). Nonetheless, among the three conditions, AR overall appeared most advantageous for GAD clients reporting longer duration of illness. Additional research on mechanisms of change in AR would help to elucidate how best to tailor treatments for chronic GAD patients.

At the same time, it is possible that mechanisms of change related to GAD duration might not be specific to the type of skills taught in AR. Newman and Fisher (2013) found that compared with CBT, self-reported duration of GAD also predicted better outcomes from other single-component interventions (cognitive therapy, self-control desensitization). These authors suggested that having GAD for a longer duration may mean that bad habits are more entrenched. Whereas CBT focuses on more skills with less practice of any one skill, AR intensively focuses on relaxation skills. Thus, for individuals who report longer duration of GAD, the benefits of AR or any single-component intervention may come from a more intensive focus on repeated practice of fewer skills necessary to break entrenched habits. On the other hand, more skills with less practice of any one skill, as in CBT, may not be as helpful for longstanding GAD patients. In line with this interpretation, moderation by GAD duration between AR and CBT was more salient in the later phase of treatment than the early phase, indicating increasing relative benefits of AR compared to CBT for GAD clients with longer duration.

A recent systematic review (Schneider, Arch, & Wolitzky-Taylor, 2015) revealed that most existing findings on moderation of anxiety disorder treatments are based on single studies, and there is a lack of replication efforts in our field. For instance, in GAD, there have been no prior replications of moderation findings. Given that a chronic course of symptoms is common in GAD (e.g., Wittchen, Zhao, Kessler, & Eaton, 1994), it is particularly encouraging that we replicated the moderating effect of longer GAD duration predicting a better response to single components than combined CBT. The current findings inform ways to tailor treatments via intensive focus on teaching fewer skills. Findings suggesting the possible added value of cognitive therapy for those with higher comorbid depressive symptoms also have important implications, because depression is highly comorbid with GAD as well as other anxiety disorders (e.g., Kessler, Chiu, Demler, Merikangas, & Walters, 2005). Thus, these findings may generalize to other anxiety disorders. In sum, continued practice of a few selected skills in single-component interventions for those with longer duration of GAD and change in maladaptive thought patterns through cognitive therapy for those with higher comorbid depressive symptoms may be crucial mechanisms for the moderation effects.

It is important to note limitations of the current study. First, although the number of repeated measurements was high, our sample and cell sizes were modest. This might have limited statistical power to detect a significant effect, especially differences between treatment conditions. Replication of the moderation effects for both pre- to posttreatment and pretreatment to follow-up increases our confidence in the findings, but given the small sample size and its tendency to lower reliability of findings, our results should be considered preliminary. Future replication with larger samples is strongly recommended. In a similar vein, we had missing data from attrition (17%) and less-than-perfect compliance with the daily diary (72%). On the one hand, TVEM and multilevel analyses both handle unbalanced data within and across individuals by using all available data for every individual at each time point. The subgroup of 55 treatment completers (83% of the sample) also showed 80% daily diary compliance, a typical recommended cutoff for EMA data collection (Stone & Shiffman, 2002). This indicates that the majority of the sample provided sufficient sampling across time. We used FIML for multilevel analyses but could not do so for TVEM analyses. Because of its novelty, there has been limited research or validation of missing data methods in TVEM. To our knowledge, no statistical software currently supports multiple imputation or FIML in TVEM either. Thus, we note this as a limitation and call for further research on missing data techniques in TVEM.

In addition, moderation by depression between AR and CBT appeared to become nonsignificant at the end of the 10-week period. However, because TVEM estimates at the boundaries (beginning and end) of time have greater uncertainty, that is, wider CIs, the nonsignificance may reflect the imprecision of estimates rather than the absence of an effect. The fact that the CIs overlapped for only the last 3 days supports this view. It is also important to note that TVEM analyses covered a time period beyond pre- to posttreatment. It included the active treatment phase (6 weeks) as well as the 2-week posttreatment assessment and additional 2 weeks of fading sessions. Our results in fact showed significant moderation between AR and CBT during posttreatment, the usual time point of interest for evaluating moderation effects. Therefore, we believe that the current findings provide at least preliminary implications for treatment selection between AR and CBT for GAD for those with higher depression.

Another limitation is that the original study (Borkovec & Costello, 1993) excluded individuals with severe depression (as assessed through a diagnostic interview) and those taking antidepressants. Based on HDRS scores, the current sample still represented a range of depression severity (none to severe), but it is
possible that the sample did not include a sufficient number of GAD patients with the most severe depressive symptoms. This may have led to a restricted range of depression scores, reducing statistical power and lowering generalizability of the findings. Replication using a larger sample of GAD patients with a greater range of depressive symptoms is warranted.

Because the original randomized controlled trial (Borkovec & Costello, 1993) was conducted before the publication of DSM–IV (American Psychiatric Association, 1994), the study used DSM–III–R (American Psychiatric Association, 1987) criteria. Diagnostic criteria for GAD have stayed the same since DSM–IV, but there were changes from DSM–III–R to DSM–IV, such as exclusion of autonomic hyperactivity symptoms. Although Abel and Borkovec (1995) found 100% overlap between clients diagnosed initially with DSM–III–R criteria and subsequently meeting criteria for DSM–IV, the current findings await replication with the updated criteria. We also want to note that GAD duration was assessed through self-report and could have been subject to recall bias. Such bias could have increased measurement error, lowering the magnitude of effects of GAD duration (i.e., wider CI). At the same time, a systematic review found that individuals can recall past experiences with sufficient accuracy, especially with external probes as in semistructured diagnostic interviews (Hardt & Rutter, 2004). Using self-reported GAD duration also has utility because therapists routinely rely on client self-report to assess duration of symptoms. These reports were also stable across a 2-week period in the current study. Nonetheless, it is important to be mindful of potential recall bias in interpreting our results.

Our TVEM analyses focused on treatment and posttreatment data, but future studies can include intensive repeated data from follow-up assessments. To reduce participant burden, researchers can employ less frequent assessments (e.g., daily instead of thrice daily) while increasing the total time period of assessment. Another extension would be to use finite mixture TVEM (Dziak, Li, Tan, Shiffman, & Shiyo, 2015), which combines finite mixture modeling with non- or semiparametric regression modeling to identify latent classes of individuals based on distinct trajectories of change or time-varying effects (e.g., symptom reduction vs. nonresponse). In addition, although TVEM assumes a smooth shape of change over time, a subset of clients with depression or anxiety disorders may show sudden gains (i.e., large reduction in symptoms between sessions), which sometimes lead to lasting gains (e.g., Aderka, Nickerson, Bøe, & Hofmann, 2012). Other analytic techniques such as growth curve models or dynamical systems models might be more appropriate for modeling such discontinuous patterns of change. In addition, because TVEM focuses on an average pattern of change in the sample, it can be complemented by person-specific analyses, which can capture individual trajectories of change and between-person variability in the trajectories.

Another area for further research is examining the interaction between GAD duration and depression in moderating treatment outcomes. Because of the modest sample size, the current study did not have sufficient statistical power to test three-way interactions (GAD Duration × Depression × Treatment Condition). However, such research would help to clarify clinical implications for GAD patients with both long duration and depression. As demonstrated in recent work (e.g., Huibers et al., 2015), using a regression-based index to integrate predictive information from multiple moderators would be a promising way to extend the current findings.

Despite the clear need for further research, the present study was the first to replicate superior efficacy of a single-component intervention for longstanding GAD patients compared with a treatment package, and to find added benefits of CBT for GAD patients with higher depression relative to AR. We should also mention that single-component comparison studies of CBT for GAD are very rare. Other than the current data (Borkovec & Costello, 1993), we are aware of only two other studies (Borkovec et al., 2002; Dugas et al., 2010). Studies that include a ND comparison are even more rare (the current data set being the only one of which we are aware). Therefore, the current data offers a unique opportunity to examine whether individual differences predicted differential responses from AR versus CBT versus ND, and how this plays out across therapy. Another unique feature of the study was testing moderation through a novel analytic approach, TVEM, to examine not only which treatment worked for whom but also how moderation effects changed over time from pre- to posttreatment. This allowed for exploring processes and potential mechanisms behind moderation effects. For instance, more pronounced moderation by GAD duration in the later phase of treatment suggested that repeated practice of fewer skills contributed to better outcomes for chronic GAD patients in AR than CBT. In addition, stronger moderation by depressive symptoms in the later phase following the introduction of cognitive therapy within CBT suggested that cognitive therapy might be especially beneficial for GAD patients with higher depression. These findings were also replicated for the pretreatment to follow-up period using multilevel modeling, adding to confidence in the results. The present study demonstrates how incorporating intensive longitudinal data in treatment studies can help to explicate the underlying mechanisms for treatment and moderation effects, which then can facilitate improving and personalizing treatments for individual clients.

References
TIME-VARYING MODERATION IN GAD TREATMENT


Received December 4, 2017

Revision received November 19, 2018

Accepted November 26, 2018