American Psychological Association

Journal of Consulting and Clinical Psychology

Manuscript version of

Patterns and Predictors of Change in Trauma-Focused Treatments for War-Related Posttraumatic Stress Disorder

Brett T. Litz, Danielle S. Berke, Nora K. Kline, Kevin Grimm, Luke Rusowicz-Orazem, Patricia A. Resick, Edna B. Foa, Jennifer S. Wachen, Carmen P. McLean, Katherine A. Dondanville, Adam M. Borah, John D. Roache, Stacey Young-McCaughan, Jeffrey S. Yarvis, Jim Mintz, Alan L. Peterson

Funded by:

- U.S. Department of Defense

© 2019, American Psychological Association. This manuscript is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final version of record is available via its DOI: https://dx.doi.org/10.1037/ccp0000426

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

CHORUS Advancing Public Access to Research

Patterns and Predictors of Change in Trauma-Focused Treatments for War-Related Posttraumatic Stress Disorder

(in press, *Journal of Consulting and Clinical Psychology*)

Brett T. Litz and Danielle S. Berke

VA Boston Healthcare System and Boston University School of Medicine

Nora Kline

VA Boston Healthcare System

Kevin Grimm

Arizona State University

Patricia A. Resick

Duke University Medical Center

Edna B. Foa

University of Pennsylvania

Jennifer S. Wachen

VA Boston Healthcare System and Boston University School of Medicine

Carmen P. McLean

VA Palo Alto Health Care Sytem and Stanford University School of Medicine

Katherine A. Dondanville

University of Texas Health Science Center at San Antonio

Adam M. Borah

Carl R. Darnall Army Medical Center, Fort Hood, Texas

John D. Roache and Stacey Young-McCaughan

University of Texas Health Science Center at San Antonio

Jeffrey S. Yarvis

Carl R. Darnall Army Medical Center

Jim Mintz

University of Texas Health Science Center at San Antonio

Alan L. Peterson

University of Texas Health Science Center at San Antonio and South Texas Veterans Health Care System

for the STRONG STAR Consortium

Author Note

Brett T. Litz and Danielle S. Berke, Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System, and Department of Psychiatry, Boston University School of Medicine; Nora Kline, Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System; Kevin Grimm, Department of Psychology, Arizona State University; Patricia A. Resick, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center; Edna B. Foa, Center for the Treatment and Study of Anxiety and Department of Clinical Psychology in Psychiatry, Perelman School of Medicine, University of Pennsylvania; Jennifer S. Wachen, Behavioral Science Division, National Center for PTSD, VA Boston Healthcare System, and Department of Psychiatry, Boston University School of Medicine; Carmen P. McLean, Dissemination and Training Division, National Center for PTSD, VA Palo Alto Health Care System, Menlo Park, California, and Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine; Katherine A. Dondanville,

Department of Psychiatry, University of Texas Health Science Center at San Antonio; Adam M. Borah, Department of Behavioral Health, Carl R. Darnall Army Medical Center, Fort Hood, Texas; John D. Roache and Stacey Young-McCaughan, Department of Psychiatry, University of Texas Health Science Center at San Antonio; Jeffrey S. Yarvis, Department of Behavioral Health, Carl R. Darnall Army Medical Center, Fort Sam Houston, Texas; Jim Mintz, Department of Psychiatry, University of Texas Health Science Center at San Antonio; Alan L. Peterson, Department of Psychiatry, University of Texas Health Science Center at San Antonio and Research and Development Service, South Texas Veterans Health Care System, San Antonio, Texas; for the STRONG STAR Consortium.

Adam Borah is now with the Department of Behavioral Medicine and Health Services, Central Texas Veterans Health Care System, Temple, Texas. Jeffrey Yarvis is now with the 21st Combat Support Hospital, Fort Hood, Texas.

Funding for this work was made possible by the U.S. Department of Defense through the U.S. Army Medical Research and Materiel Command, Congressionally Directed Medical Research Programs, Psychological Health and Traumatic Brain Injury Research Program awards W81XWH-08-02-0114 (Brett Litz), W81XWH-08-02-109 (Alan Peterson), W81XWH-08-02-0111 (Edna Foa), W81XWH-08-02-0114 (Brett Litz), and W81XWH-08-02-0116 (Patricia Resick). The views expressed herein are solely those of the authors and do not reflect an endorsement by or the official policy or position of the U.S. Army, the Department of Defense, the Department of Veterans Affairs, or the U.S. Government.

Correspondence concerning this article should be addressed via email to Brett Litz litzb@bu.edu.

Abstract

Objective: We evaluated patterns and predictors of change from three efficacy trials of trauma-focused cognitive-behavioral treatments (TF-CBT) among service members (N = 702; mean age = 32.88; 89.4% male; 79.8% non-Hispanic/Latino). Rates of clinically significant change were also compared with other trials. **Method**: The trials were conducted in the same setting with identical measures. The primary outcome was symptom severity scores on the PTSD Symptom Scale – Interview Version (PSS-I; Foa et al., 1993). **Results**: Symptom change was best explained by baseline scores and individual slopes. TF-CBT was not associated with better slope change relative to Present-Centered Therapy, a comparison arm in 2 trials. Lower baseline scores (β = .33, p < .01) and higher ratings of treatment credibility (β = -.22, p < .01) and expectancy for change (β = -.16, p < .01) were associated with greater symptom change. Older service members also responded less well to treatment (β = .09, p < .05). Based on the Jacobson and Truax (1991) metric for clinically significant change, 31% of trial participants either recovered or improved. **Conclusions**: Clinicians should individually tailor treatment for service members with high baseline symptoms, older patients, and those with low levels of credibility and expectancy for change.

Keywords: posttraumatic stress disorder, active duty military, war-related PTSD

Public Health Significance: Three randomized trials demonstrated that trauma-focused cognitive-behavioral therapies for war-related posttraumatic stress disorder in active duty military personnel had rates of clinically significant change that were substantially smaller than in studies of civilians and not different from Present-Centered Therapy. This study highlights the importance of tailoring treatment for service members with high baseline symptoms, older patients, and those with lower levels of credibility and expectancy for change.

Patterns and Predictors of Change in Trauma-Focused Treatments for War-Related Posttraumatic Stress Disorder

Randomized clinical trials (RCTs) to evaluate the treatment of war-related posttraumatic stress disorder (PTSD) in active duty military service members is a nascent area of clinical research. Nevertheless, since 2004, efficacy trials of PTSD treatments have been used as the basis for best practice recommendations in the U.S. Department of Veterans Affairs and U.S. Department of Defense (VA/DoD; 2017). For good reason, the 2017 guidelines recommend trauma-focused cognitive-behavioral treatments (TF-CBT) for PTSD; broadly speaking these therapies are first-line because of very strong efficacy evidence. In contexts where the guidelines are disseminated and used (Rosen et al., 2016; Sayer et al., 2017), they hold the promise of ensuring evidence-based, uniform, time-limited care, which may enhance patient and care-provider confidence and satisfaction and result in improved outcomes (Karlin et al., 2010). However, the existing practice guidelines have noteworthy, but addressable, limitations.

Specifically, the guidelines can offer only *nomothetic* prescriptions for care. TF-CBT, as tested in trials, is recommended broadly, regardless of potentially key patient characteristics (e.g., demographics, military service variables, types of military trauma), or clinical presentations (e.g., pressing comorbid problems). This problem is a byproduct of the state of evidence in the field, which is based on high-quality efficacy trials, which were not designed to determine whether a treatment should be used in all patient populations or clinical contexts. The existing efficacy trials only test outcomes based on overall average effect sizes. Mean effect size hides potential heterogeneity of effect sizes (some patients have impressive outcomes, some do not change at all, and some do worse; Kraemer, Frank, & Kupfer, 2006; Kravitz, Duan, & Braslow, 2004), and substantial effect sizes may not be meaningful clinically or functionally

significant. Nomothetic recommendations may not be better than usual care when patients do not match the average patient in a given trial. Moderator analyses could identify subgroups that do better or worse, but efficacy trials are not typically designed or powered a priori to answer these questions, and, although post-hoc moderator analyses can be revealing, they may not replicate because trials are chiefly samples of convenience (not population-based; e.g., Rothwell, 2005).

Because efficacy trials are not designed to generalize to all patient samples and contexts, basing practice recommendations on aggregated trial results limits the utility of the VA/DoD guidelines. For example, high-quality parallel efficacy trials of war veterans (Polusny et al., 2015; Schnurr et al., 2007; Schnurr et al., 2003) and service members (Engel et al., 2014) have substantially smaller effect sizes than trials of female sexual assault victims (e.g., Foa et al., 1999; Resick, Nishith, Weaver, Astin, & Feuer, 2002), yet these trials are culled and assumed to be equipotent with respect to war-related PTSD treatment practices. There is also heterogeneity of effect sizes across trials, but the guidelines do not use an effect size threshold as a scoring criterion. The broader problem is that there is no uniform benchmark for defining and comparing clinically meaningful change in trials (or practice), which could help clinicians, patients, and policy makers determine what constitutes sufficient efficacy.

In this study, we analyzed participant-level metadata from three large, high-quality RCTs of TF-CBT with treatment-seeking active duty U.S. military service members with PTSD. Our goal was to explore patterns of change and predictors of change that might refine evidence-based treatment recommendations for service members with PTSD. We had four specific aims: (1) to describe the participants to allow clinicians to determine if these service members match a given clinical context; (2) to determine the best-fitting patterns of change in PTSD symptom severity over the course of treatment and follow-up intervals, with the goal of unpacking mean slopes in

various arms to determine subgroups that respond differently to various treatments; (3) to test predictors of subgroup outcomes that could aid in clinical decision making; and (4) to derive indices of clinically significant change to benchmark the results by comparing identical, clinically significant change indices from other high-quality trials.

Two studies tested trajectories of self-reported outcomes among war veterans with PTSD (Elliott, Biddle, Hawthorne, Forbes, & Creamer, 2005; Schumm, Walter, & Chard, 2013). Appealing to these precedents, we hypothesized that the following trajectories of change would best fit the data (in order of prevalence): (a) *small gains* (steady but at best moderate change); (b) *no gains* (no statistical change from baseline); and (c) *enduring gains* (substantive linear improvement). If latent classes did not best fit the data, we planned to test additional change parameters that would best fit the data. We hypothesized that the severity of baseline mental health symptoms is a proxy for *case complexity* and predicted that cases with higher scores would be less responsive to treatment. We assumed that these service members may have more pressing problems than PTSD or their problems may be more entrenched or multifaceted (see Ruscio & Holohan, 2006). We also hypothesized that TF-CBT would result in enduring positive gains relative to a comparison therapy. Consistent with research showing that veterans with PTSD have better outcomes when they buy into a therapy (Hundt, Harik, Barrera, Cully, & Stanley, 2016), we hypothesized that pretreatment ratings of credibility and expectancy would also predict positive gains. Finally, an exploratory prediction was that older service members would make more positive gains because they would be more motivated to change due to the greater likelihood that enduring problems would affect their ability to advance their careers and their ability to lead and help others to a greater degree than younger service members.

Method

Participants and Procedures

This observational study was a separate planned and funded project as part of the *South Texas Research Organizational Network Guiding Studies on Trauma And Resilience* (STRONG STAR) consortium. Participants were military service members recruited, assessed, and treated at the Carl R. Darnall Army Medical Center at Fort Hood, Texas. We merged identical data elements from three parallel trials of service members with PTSD. The first trial compared Cognitive Processing Therapy (CPT), a TF-CBT, to Present-Centered Therapy (PCT), a non-TF supportive and problem-solving treatment. Each therapy was provided in a 90-minute group format delivered twice weekly (N = 108; Resick et al., 2015. A second trial compared CPT, conducted in a 90-minute group format, delivered twice weekly, to CPT, conducted in 60-minute individual therapy (N = 268; Resick et al., 2017). The group therapies had fixed schedules (missed sessions were not rescheduled, four missed sessions led to termination). The third trial (Foa et al., 2018; N = 326) compared 10 individual, 90-minute sessions of another TF-CBT, Prolonged Exposure (PE), over 8 weeks (PE-spaced); 10 individual PCT sessions delivered over 8 weeks; and 10 individual, 90-minute sessions of massed PE delivered over 2 weeks (PE-massed).

The Institutional Review Boards (IRB) at Brooke Army Medical Center and the University of Texas Health Science Center at San Antonio (UTHSCSA), as well as the U.S. Army Medical Research and Materiel Command Human Research Protection Office, approved the three clinical trials. This archival study was approved by the IRB at VA Boston Healthcare System. All participants provided informed consent for the trials. To monitor safety during the progress of each trial and to ensure that participants' benefits exceeded risk, a Data and Safety

Monitoring Board, which was independent from the investigators and sponsor of the research, monitored the trials. Across trials, recruitment was based on referrals by providers and self-referrals by service members. Prescreening criteria included active duty status, deployment to Iraq and/or Afghanistan, aged 18 to 65 years, psychiatric medication stability, and availability.

Each trial had the same primary endpoint, namely PTSD symptom severity. A pool of blinded independent evaluators (IEs) were trained to criterion to administer the PTSD Symptom Scale, Interview Version (PSS-I; Foa, Riggs, Dancu, & Rothbaum, 1993) at four time points: baseline, posttreatment, and 6- and 12-months posttreatment (see patient flow diagrams for each trial in Supplementary Figure 1). Participants also filled out a battery of questionnaires (see the published trials), a subset of which were used in this paper.

Measures

Demographics and military service characteristics. Demographic information was collected with a standardized self-report form.

The PTSD Symptom Scale, Interview Version. The PSS-I (Foa et al., 1993) is a 17-item, structured clinical interview that evaluates the frequency and intensity of *DSM-IV-TR*-based PTSD symptoms in the past two weeks on a 4-point scale ranging from 0 (*not at all*) to 3 (*very much*). These scores are summed to create a total severity score. The PSS-I is comparable to the Clinician Administered PTSD Scale (Foa & Tolan, 2000), the gold standard PTSD assessment with veterans (the PSS-I was chosen because it takes considerably less time). The internal consistency of the PSS-I across the three trials was .83.

Each PSS-I was taped, and at least 5% of each IE's interviews were selected for co-rating by two experts. Fidelity was evaluated using 124 co-rated interviews. Interrater reliability

comparing expert and IE diagnosis was substantial (K = .87), and the absolute agreement of total severity scores was excellent (two-way random effects intraclass correlation coefficient = .98).

The PTSD Checklist - Stressor-Specific treatment version for *DSM-IV* (PCL-S). The PCL-S (Weathers, Litz, Herman, Huska, & Keane, 1993) is a 17-item, self-report measure that assessed PTSD symptoms at the same intervals as the PSS-I and over the course of treatment, indexed to the Criterion-A event used for the PSS-I. When administered at baseline and at each posttreatment interval, ratings were based on the past month; when administered during treatment, ratings were *in the past week* (Trials 1-2) or *since the last session* (Trial 3). Within treatment, the PCL-S was administered at every session in Trial 3 (excluding session 1 and session 10) and at even-numbered sessions in Trials 1 and 2. The internal consistency of the PCL-S across trials was 0.85. The PCL-S was used as an additional test of the hypothesis that greater baseline symptom burden would lead to worse outcomes, and, in an exploratory fashion, to examine patterns of change *within* treatment.

Beck Depression Inventory - II (BDI-II). The BDI-II is a 21-item measure of the affective and somatic symptoms of depression (Beck, Steer, & Brown, 1996). Items are scored on a 0 (*no disturbance*) to 3 (*maximal disturbance*) scale. Items are summed to create a total score. The internal consistency across trials was 0.91. The BDI-II was used in this study as an indicator of comorbid mental health burden at baseline.

Credibility and Expectancy Questionnaire (CEQ). The CEQ (Devilly & Borkovec, 2000) is a 6-item scale comprised of two 3-item subscales that assesses participants' expectancy that the treatment could help them and how credible the treatment is in their minds, respectively. The CEQ has been shown to have strong internal consistency and test-retest reliability (Devilly

& Borkovec, 2000). The internal consistency of the expectancy and credibility subscales were 0.77 and 0.85, respectively, across the three trials.

Data Analysis Plan

Sample characteristics. Demographic/military service characteristics were compared to the U.S. Army in 2014 (U.S. Department of Defense, n.d.). Prevalence differences were assessed using a normal approximation of the multinomial distribution.

Modeling patterns of change in PTSD symptom severity. Growth mixture modeling (GMM; e.g., Muthén & Shedden, 1999) was used to generate best-fitting trajectories of change in PSS-I total severity scores. We chose GMM because it identifies shared (group) patterns of change (see Ram & Grimm, 2009). If group trajectories failed to best fit the data, we planned to explore other ways of modeling change until a best-fitting model was identified.

Analyses of clinically significant change. We sought an index of the magnitude of treatment effects that: (1) could be used prognostically (main effects) or predictively (moderators); (2) could be used with individuals (and, therefore, in clinical practice) and be aggregated in standardized ways; (3) would allow for the incorporation of intent-to-treat data; and (4) has been used in various trials in the PTSD field. Two methods recommended by Jacobson and Truax (1991) met these criteria. The first task is to determine a reasonable cutoff between the patient/dysfunctional and nonpatient/functional populations for posttreatment scores. Because there is no consensus or sufficient empirical basis for these values (any recommended cut score from a given trial is measure- and sample-specific and therefore not generalizable), we used Jacobson and Traux's (1991) empirical recommendation, in which a posttreatment score that is 2 *SD*s beyond the range of the baseline mean of a reference group qualifies as clinically significant change. In our case, the reference group is all randomized

participants. For the subgroup who had posttreatment PSS-I data, the PSS-I score that indexed sufficient positive clinical change at posttreatment was 12.54. The second method involves determining clinically significant change from pre- to post-treatment, which involves the calculation of a *reliable change index* (RCI) for each participant to ensure that symptom changes are not due to an artifact of measurement error. The RCI is computed according to the following formula: (x2 – x1) / Sdiff, where x1 is a participant's pretreatment PSS-I total score and x2 represents the participant's posttreatment PSS-I total score. Sdiff represents the standard error of difference between these two test scores and was calculated from the internal consistency of the PSS-I at baseline, as suggested by Martinovich, Saunders, and Howard (1996). An RCI value that is a Z-score $\geq$1.96 ($\geq$2 SD from the mean) reflects change that is statistically superior to measurement error (at .05 alpha). Following Jacobson and Truax (1991), we classified participants as *recovered* if they passed the 2*SD* criterion and the RCI criterion (a PSS-I change score of 10.39 corresponded to an RCI of 1.96), *improved* if they passed the RCI criterion but not the 2*SD* criterion, *unchanged* if they did not pass the RCI criterion, and *deteriorated* if worsened scores passed the RCI criterion. We reported the rates of these categories in two ways: (1) using participants who had posttreatment data (n = 475) and, (2) adding participants with missing follow-up data to the unchanged category, consistent with an intent-to-treat analysis (N = 702). We also conducted chi-square analyses to compare the percentage of participants who either recovered or improved versus those who were unchanged or deteriorated. Finally, we asked colleagues who conducted high-quality efficacy trials of TF-CBT to generate these categories using identical formulas, providing comparison benchmarks.

Results

Demographic and Military Service Characteristics

Relative to the U.S. Army in 2014 (U. S. Department of Defense, n.d.), there were more male service members, more noncommissioned officers (NCOs; E-4 to E-6) and senior NCOs (E-7 to E-9), and a smaller proportion of junior enlisted (E-1 to E-3), as well as fewer officers, randomized into the three STRONG STAR trials (see Table 1). The service members randomized into the three trials were also older (a higher percentage above 26 years, see Table 1; mean age = 32.88, *SD* = 7.39) and less likely to be Caucasian relative to the U.S. Army. The study group also entailed a substantially higher proportion of married and divorced/separated service members relative to the U.S. Army (see Table 1).

Changes in PSS-I Scores

Before analyzing the data, we examined each participant's raw data to discern potential patterns of change. Supplementary Figure 2 depicts the raw PSS-I scores against the number of days since baseline for each measurement. The latter was necessitated by the considerable dispersion of time of follow-up measurement in the three trials. The expected posttreatment assessment interval is represented by a thin vertical line at 3 months, and the expected 6- and 12-month follow-up intervals are also represented by thin vertical lines. The density changes over time demonstrated missing data, and there was no obvious pattern of change.

Models considered. Changes in PSS-I scores over time were modeled with several different growth models to examine the overall change pattern for the sample. Considered growth models included the *intercept only* growth model, the *linear* growth model, the *quadratic* growth model, and the *two-part linear spline* growth model. In all models, the number of days since the baseline assessment was utilized as the time metric. The fit statistics of the models considered are shown in Table 2. Of the models considered, a two-part linear spline with an estimated knot point was the best fitting model. Parameter estimates of this model suggest the

following: The mean of the pre-knot (a knot is the point at which the distribution can be partitioned to reflect slope change) random slope (β_1) was -2.06 points per month, indicating that PSS-I scores declined, on average, from the baseline assessment to the knot point, which was located at 94.42 days after baseline. The mean of the post-knot random slope (β_2) was 0.03 and not significantly different from 0, suggesting that there was no systematic mean change in PSS-I scores 94.42 days postbaseline. The mean of the intercept (β_0), which was located at the knot point, was 18.80 and represents the mean PSS-I score at 94.42 days. The variance of the intercept and the pre-knot slope were significantly different from zero, suggesting that individuals differed in their PSS-I scores at 94.42 days and in the rate of change prior to 94.42 days. The variance of the post-knot slope was not significantly different from zero; participants did not differ in their change patterns after 94.42 days.

Examination of differential change patterns in PSS-I scores. Growth mixture models, which allow for the random coefficients of the model to follow a finite mixture distribution, were fit to determine if there were shared distinct latent trajectories. In growth mixture models, the number of classes is unknown, as are the ways in which the classes differ (assuming the existence of latent classes). Following Ram and Grimm (2009), the following models were considered: (a) multiclass models with $\boldsymbol{\beta}_k$, $\gamma_k = \gamma$, $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$, and $\Psi_k = \Psi$; (b) multiclass models with $\boldsymbol{\beta}_k$, γ_k, $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$, and $\Psi_k = \Psi$; (c) multiclass models with $\boldsymbol{\beta}_k$, γ_k, $\boldsymbol{\Sigma}_k$, and $\Psi_k = \Psi$; and (d) multiclass models with $\boldsymbol{\beta}_k$, γ_k, $\boldsymbol{\Sigma}_k$, and Ψ_k. In the search for latent classes, all models were built on the two-part linear spline model with an unknown knot point.

The fit statistics of the fitted latent class models are shown in Supplementary Table 1. Of the 2-class growth mixture models considered, the only model that converged to a proper solution was the 2-class model with β_k allowed to vary over classes. This model had a higher

Bayesian Information Criterion (BIC = 11, 941) than the 1-class (growth) model (BIC = 11,939), indicating that the 2-class model did not fit better than the 1-class model. The entropy for this model was .46, indicating low confidence in class assignments. Due to the lack of improvement, 3-class models were not considered. Due to the lack of convergence with the 2-class model with $\boldsymbol{\beta}_k$ and γ_k estimated for each class, the subsequent 2-class models assumed $\gamma_k = \gamma$.

As a follow-up to the growth mixture models, a series of latent class growth models were fit. These models utilize the finite mixture distribution, but $\boldsymbol{\Sigma}_k = 0$ for all classes. Thus, all between-person differences in change are assumed to come from the class differences, as there are no between-person differences in the intercepts or slopes within each class. Due to severe convergence issues, models with only $\boldsymbol{\beta}_k$ allowed to vary over classes are reported with the knot point fixed to the value from the 1-class model. In this series of models, the fit of the models, with respect to the BIC, improved with increasing the number of latent classes up until the 4-class model (i.e., the 5-class model had a higher BIC). All of the models had poor entropy, indicating low confidence in the classifications. Given this, there was no support to any latent class growth model. Furthermore, given that the bilinear spline growth model had a lower BIC than any of these models, we considered the bilinear spline growth model as the best representation of the data.

The predicted individual growth trajectories in PSS-I are shown in Figure 1, which depicts the individual PSS-I growth curves predicted by the bilinear-spline model. Because there was no change after the posttreatment assessment, the prediction models described below used individual PSS-I slope scores from pre- to post-treatment as the dependent measure.

Predictors of Individual PSS-I Slopes

Participants were clustered within one of three clinical trials and all participants were nested within therapists (i.e., multiple service members were treated by the same therapist across all arms across trials). Participants assigned to a trial arm in which therapy was delivered in a group format were further nested into group cohorts. This created three levels of nesting in the group therapy arms (Level 1 = participants; Level 2 = therapists; Level 3 = group cohorts) and two levels of nesting in the individual therapy arms. A multi-level model building approach using Mplus (Muthén & Muthén, 2017) was employed to evaluate the data structure with respect to PSS-I change. First, a series of null models (or intercept-only models) was constructed to estimate the intraclass correlation (ICC) of PSS-I slopes. Separate null models were specified for each cluster. In order to account for the fact that participants were partially nested within groups (i.e., half the participants in Trial two and all participants in Trial three received individual treatment), multiple-arm partial nesting models were specified by constraining group variance to zero for participants in individual therapies. Variability between group cohorts was only estimated among those participants assigned to group interventions. Results revealed nonsignificant variance between cluster means (see Supplementary Table 2). Consequently, ordinary least squares regression was used to test hypotheses.

As predicted, participants who began treatment with greater baseline PTSD (β_{PCL} = .39, p < .001; $\beta_{PSS\text{-}I}$ = .33, p < .001) and depression symptom severity scores (β = .25, p < .001) had less improvement from pre- to post-treatment. Also, as predicted, favorable expectancies about therapy (β = -.16, p < .001) and credibility ratings (β = -.22, p < .001) were associated with improved PSS-I scores. Participants assigned to TF-CBT arms, relative to those assigned to any PCT arm, did not differ in slope in PTSD symptoms from pre- to post-treatment (β = .04, p =

.30). Also, counter to predictions, older participants experienced less improvement in PSS-I scores ($\beta = .09$, $p = .014$).

Exploratory Analysis of Within-Treatment Change

Because each trial administered a PCL repeatedly over the course of treatment, we examined within-treatment patterns of change that may aid clinical decision-making regarding when to consider a change in treatment due to non-responsiveness. We explored whether a session number could be identified after which slope change leveled off, predicting changes in PCL scores (excluding the PE-M arm) over the course of treatment with Growth Mixture Models (see Supplementary Table 3). While a 2-class model had adequate entropy, the fit statistics difference with a 1-class model was negligible, and the classes only separated 48/592 participants. The 1-class model also provided a better test of our exploratory question (Supplementary Figure 3 shows the box-plots of mean scores). This model had a pre-knot decrease of ~7.30 PCL points per month and a post-knot leveling off (to ~4.6 PCL points/month). The knot-point was approximately at the 6^{th} or 7^{th} session. Given that 95% of the cases that did not show reliable change by session 7 did not have reliable change between session 7 and posttreatment this suggests session 7 as a shift point for patients who do not make reliable change. This finding needs to be replicated particularly because there was a high standard deviation value of the slopes (pre-knot SD=~14 points, post-knot SD=~6).

Rates of Clinically Significant Change

Table 3 provides the number and percentage of participants across the three trials classified as recovered, improved, unchanged, and deteriorated. We present two sets of values, one based on participants who had posttreatment data (completers) and one with all randomized participants, adding participants with missing posttreatment data to the unchanged category. This

employs a last observation carried forward, intent-to-treat method. Results indicated that rates of clinically significant change (recovered and improved) were significantly lower than rates of nonpositive-change (unchanged and deteriorated) for completers (Standardized Difference Score = 8.87, $p \leq .001$; $d = .81$) and for intent-to-treat (Standardized Difference Score = 18.78, $p \leq .001$; $d = 1.42$) samples. To benchmark these results, Table 3 also contains identically calculated change categories (also indexed from baseline to posttreatment) from high-quality trials of civilians and war veterans and an observational study of veterans in VA care (for the trials, the results are collapsed across treatment arms, when applicable). Structured and manualized psychotherapies for PTSD provided to service members in garrison lead to substantially less clinically significant change than civilian trials. Excluding participants with missing posttreatment data (completers), the sum of recovery and improved rates for the three STRONG STAR trials was 31%. The recovery and improvement rate was 39% for the Schnurr et al. (2007) VA cooperative studies program trial of female veterans comparing PE and PCT. The rate was 33% for the Shiner et al. (2018) VA outpatient medical record study of medications for PTSD. The rate was 39% for the two combined Foa et al. civilian trials, one comparing naltrexone vs. placebo, all treated with PE (Foa et al., 2013), and another comparing varenicline plus smoking cessation plus PE versus varenicline plus smoking cessation only (we used the PE arm only; Foa et al., 2017). For the Resick, Nishith, Weaver, Astin, and Feuer (2002) trial, comparing CPT, PE, and a waitlist group in civilian female sexual assault victims, the recovery and improved rate was 83%.

Discussion

Prior to STRONG STAR, no trial had tested the efficacy of TF-CBT to treat active-duty service members with PTSD in garrison (existing trials tested telehealth approaches, chiefly in

primary care; e.g., Engel et al., 2014). This means that there was a considerable knowledge gap in the trials used to establish the VA/DoD guidelines for specialty care treatment for PTSD *in the military context*. The expectation was that providing evidence-based treatment for PTSD relatively early after deployment would lead to greater efficacy in trials of veterans because the PTSD would be putatively less chronic and entrenched. Service members seeking treatment in garrison were also expected to be more motivated to change to help them remain fit for duty and continue their military career advancement. Nevertheless, we anticipated the need to leverage the combined trials to determine the types of cases who respond well and those who do not. The goal was to address limitations of the VA/DoD PTSD treatment guidelines by identifying predictors of distinct patterns of change that could help clinicians individualize care to minimize dropout and treatment failure.

We found a striking degree of individual variation in PTSD change. Not surprisingly, given the raw data, no group patterns of latent change in PTSD symptoms were found. Instead, PTSD symptom change was best explained by individual pretreatment PTSD symptom burden and individual slope change. The latter suggests clinicians should expect that service members seeking treatment for PTSD will have unique treatment response potentials. The finding that service members with lower baseline PTSD severity scores had greater slope change (as was the case with depression symptom severity) was consistent with our hypothesis. This suggests that baseline symptom burden may be a proxy for case complexity. Given that more severe and diverse symptoms usually have multiple complex causes and maintaining factors, complex cases may require an individualized approach (Delgadillo, Huey, Bennett, & McMillan, 2017). For example, outcomes may be improved if other, more pressing problems (see Rosen, Adler, & Tiet, 2013) are addressed successfully (e.g., Cloitre, Koenen, Cohen, & Han, 2002). It may be

useful to generate a problem list, prioritize targets for change, and use TF-CBT when PTSD is the primary problem or when various problems are implicated by exposure to trauma (e.g., new-onset or worsening problems). It is of note in a multiple-target approach that, although various types of assessment information are gathered in routine practice, there are no systematic, evidence-based methods of synthesizing assessment data to inform treatment decisions or signal need for a strategy change. All that is known are the determinants of the initiation of TF-CBT, which is based on clinician and patient preference and clinic culture (Sayer et al., 2017).

Unexpectedly, TF-CBT was not associated with greater change in PTSD symptom severity relative to PCT. This null result is consistent with prior trials of veterans with PTSD (Frost, Laska, & Wampold, 2014). In fact, meta-analyses and comprehensive reviews have underscored the general state of equipoise among credible bona fide treatments (e.g., Steenkamp, Litz, Hoge, & Marmar, 2015). In these studies, PCT was no less credible and no less associated with positive expectancies for change than TF-CBT. It can be argued that PCT should be used as a first-line treatment for war-related PTSD, particularly for service members in garrison who have relatively low baseline symptom burden. However, it should be noted that PCT is a manualized therapy that requires training to a high level of fidelity. As compared to most TF-CBTs, PCT is arguably simpler to learn, easier to use, and less invasive and demanding of therapists and patients. Nonetheless, few military providers have been trained in PCT.

A substantial majority of service members failed to make clinically significant gains within and across trials. Active duty service members treated in garrison did not uniquely and substantively benefit from TF-CBT relative to trials in veteran populations. The rate of clinically significant change roughly matches other parallel group trials of veterans (e.g., Schnurr et al., 2007) and replicates the starkly contrasting results of trials of war-related and civilian PTSD

(Steenkamp & Litz, 2013). There is sufficient evidence to counter the expectation that large effect sizes in trials of civilian sexual assault victims (see Table 3) will generalize to the active duty context. Indeed, it may be clinically useful to have treatment guidelines that are uniquely applicable to combat- or war-related trauma.

Consistent with predictions, pretreatment credibility and expectancy ratings modestly predicted greater change in PTSD symptom severity. It may be that the content and processes of the therapies as explained at the outset of the trials failed to resonate or instill positive expectancies with an unspecified percentage of participants. Perhaps some service members were not comfortable with the personal and time requirements of psychotherapy, and the process for some may have been inconsistent with the action-oriented nature of the military culture. Service members may also require a more sustained and detailed collaborative discussion and question-and-answer period with their clinician to accurately and positively absorb messages about the treatment process (Hoge et al., 2014). Patients in trials are not consumers with the ability to choose among various options (although dropout might be considered exercising choice), so we cannot know whether the therapy assigned might have run counter to personal preference.

Counter to our expectations, older service members responded less well to treatment than younger service members. However, this result needs to be taken in the context of the uniquely skewed demographic characteristics of the combined sample of the three STRONG STAR trials, which had higher pay grades (more NCOs and senior NCOs), older age, and greater marriage rates relative to the U.S. Army. Consequently, service members in the STRONG STAR trials that responded less well to treatment were not only older but also likely managing greater leadership responsibilities and well-established relationships. It may be that the early career and civilian status of therapists (most of whom were postdoctoral trainees) created degrees of

distrust, disconnection, and emotional distance, which made frank disclosure difficult for older service members. It may be that older service members in positions of leadership were not comfortable losing degrees of personal control and, consequently, may have failed to meet the boundary conditions for treatment. If this hypothesis was confirmed, the implication is that some therapists practicing in military settings, in service of earning and building trust, may need to have a candid discussion about the presence of a legitimate age/culture clash and work on compassionately challenging some patients to try on new ways of relating.

We also found that PTSD symptom severity scores did not change from posttreatment to any follow-up assessment point. This means that substantive change that occurred over the course of treatment was maintained. Conversely, the majority of service members who did not substantively change over the course of therapies appeared not to have any delayed benefit. In an exploratory analysis, we estimated that if service members did not show PTSD symptom severity change by session 7 it was likely that they would not subsequently change. The clinical implication is that clinicians should routinely assess PTSD symptoms over the course of treatment and that sustained nonresponse should trigger a midcourse reconceptualization and shift in treatment strategy. Indeed, the VA/DoD PTSD treatment guideline recommends routine monitoring of treatment progress. However, no guidance is provided about what constitutes insufficient change and what should be done if patients are nonresponsive. With respect to the former issue, scores that fail to exceed measurement error, using the RCI criterion, should be considered as sufficient evidence for nonresponse. Because the RCI can be easily calculated using cumulative local clinic baseline scores of any measure and is an *individual change index*, it is ideal for this purpose (see Shiner, Watts, Pomerantz, Young-Xu, & Schnurr, 2011). With respect to the latter, clinicians should revise their conceptualization about the most effective

change agent or generate alternative targets and a revised treatment plan. It may be useful for clinicians also to get frank feedback from service members about their experience of the therapy and the requirements and demands of a given treatment to preclude obstacles to subsequent successful engagement (if applicable).

This study had noteworthy validity threats. First, the study was observational. Although we planned to combine the three STRONG STAR clinical trials to conduct meta-analyses to answer practice-relevant questions (by ensuring common data elements and uniform procedures), the results are correlational and subject to potential third variables. Particularly with respect to the finding that TF-CBT was not superior to PCT, notwithstanding addressing cluster effects, the finding is confounded by trial cohort and delivery modality, although Trial 3 and the Schnurr et al. (2007) trial failed to find a difference between PE and PCT at follow-up. Although there was a moderate effect size difference between group CPT and group PCT in Trial 1 indexed by PCL-S scores (as planned), there were no differences between these two arms indexed by the PSS-I. Finally, the clinical significance benchmark we chose, although commonly used, has been criticized because it may be imprecise and does not account for regression to the mean or variable practice effects (Speer & Greenbaum, 1995; Tingey, Lambert, Burlingame, & Hansen, 1996). However, modifications to the approach have not shown substantive incremental validity (Atkins, Bedics, McGlinchey, & Beauchaine, 2005). Finally, the trials had inclusionary criteria that may not match clinical care (Westen, Novotny, & Thompson-Brenner, 2004), and the trials were not population-based random samples. Although all participants underwent treatment in the same setting (Fort Hood), each study group was a sample of convenience and in the aggregate demographically skewed.

Notwithstanding these methodological issues, the individualized nature of change in this analysis, coupled with a majority not showing clinically significant change, suggest two pathways to reduce the risk of treatment failure. One involves increasing the likelihood that appropriate cases get TF-CBT. If a clinical context matches the demographic characteristics of our combined sample, we recommend that clinicians individualize treatment plans for older service members, service members with high symptom burden, and those with relatively low ratings of the credibility or confidence in TF-CBT. Because we only tested these variables as main effects in our analyses, interactions and mediational associations need subsequent testing.

The other way forward entails generating and testing ideographic approaches to the treatment of military- and war-related PTSD. The aim would be to generate actuarial algorithms and empirically supported decision aids that clinicians can use if specific cases match given profiles, which is a hybrid nomothetic (profile-based) and ideographic (patient-based) approach (Kessler et al., 2017). For an individualized approach to be viable, it will need to lead to more clinically significant change than usual care and manualized TF-CBT. Given that most VA cases do not receive TF-CBT for varying reasons (Watts et al., 2014), notwithstanding patient choice and biased clinician decision-making (Hundt et al., 2016), clinicians can provide an untapped knowledge source to generate hypotheses about profiles and replicable individual treatment planning.

Another related research avenue entails leveraging the dissemination, training, and certification processes in place in the DoD and VA and evaluating the ways in which clinicians are modifying TF-CBT to better fit patient presentations, trauma types, resources, and capabilities (e.g., Stirman et al., 2017). Researchers can test the effectiveness of the modified approaches relative to existing TF-CBT manuals. Finally, in the clinical context, independent

evaluators could use qualitative methods to assess treatment nonresponders' experience, which may also reveal testable hypotheses about unaddressed targets, mismatches of treatment approach, and patient capabilities.

References

Atkins, D. C., Bedics, J. D., McGlinchey, J. B., & Beauchaine, T. P. (2005). Assessing clinical significance: Does it matter which method we use? *Journal of Consulting and Clinical Psychology, 73,* 982-989. https://doi.org/10.1037/0022-006x.73.5.982

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

Cloitre, M., Koenen, K. C., Cohen, L. R., & Han, H. (2002). Skills training in affective and interpersonal regulation followed by exposure: A phase-based treatment for PTSD related to childhood abuse. *Journal of Consulting and Clinical Psychology, 70,* 1067-1074. https://doi.org/10.1037//0022-006x.70.5.1067

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*, 835-853. https://doi.org/10.1037/ccp0000231

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry, 31,* 73-86. https://doi.org/10.1016/s0005-7916(00)00012-4

Elliott, P., Biddle, D., Hawthorne, G., Forbes, D., & Creamer, M. (2005). Patterns of treatment response in chronic posttraumatic stress disorder: An application of latent growth mixture modeling. *Journal of Traumatic Stress, 18,* 303-311. https://doi.org/10.1002/jts.20041

Engel, C. C., Bray, R. M., Jaycox, L. H., Freed, M. C., Zatzick, D., Lane, M. E., ... Katon, W. J. (2014). Implementing collaborative primary care for depression and posttraumatic stress disorder: Design and sample for a randomized trial in the U.S. military health system. *Contemporary Clinical Trials, 39,* 310-319.

https://doi.org/10.1016/j.cct.2014.10.002

Foa, E. B., Asnaani, A., Rosenfield, D., Zandberg, L. J., Gariti, P., & Imms, P. (2017). Concurrent varenicline and prolonged exposure for patients with nicotine dependence and PTSD: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 85,* 862-872. https://doi.org/10.1037/ccp0000213

Foa, E. B., Dancu, C. V., Hembree, E. A., Jaycox, L. H., Meadows, E. A., & Street, G. P. (1999). A comparison of exposure therapy, stress inoculation training, and their combination for reducing posttraumatic stress disorder in female assault victims. *Journal of Consulting and Clinical Psychology, 67,* 194-200. https://doi.org/10.1037//0022-006x.67.2.194

Foa, E. B., McLean, C. P., Zang, Y., Rosenfield, D., Yadin, E., Yarvis, J. S., ... Peterson, A. L.; for the STRONG STAR Consortium. (2018). Effect of prolonged exposure therapy delivered over 2 weeks vs 8 weeks vs present-centered therapy on PTSD symptom severity in military personnel: A randomized clinical trial. *Journal of the American Medical Association, 319,* 354-364. https://doi.org/10.1001/jama.2017.21242

Foa, E. B., Riggs, D. S., Dancu, C. V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post- traumatic stress disorder. *Journal of Traumatic Stress, 6,* 459-473. https://doi.org/10.1007/bf00974317

Foa, E. B., & Tolin, D. F. (2000). Comparison of the PTSD symptom scale–interview version and the clinician- administered PTSD scale. *Journal of traumatic stress, 13,* 181-191. https://doi.org/10.1023/A:1007781909213

Foa, E. B., Yusko, D. A., McLean, C. P., Suvak, M. K., Bux, D. A., Oslin, D., ... Volpicelli, J. (2013). Concurrent naltrexone and prolonged exposure therapy for patients with comorbid alcohol dependence and PTSD: A randomized clinical trial. *Journal of the*

American Medical Association, *310*, 488-495. https://doi.org/10.1001/jama.2013.8268

Frost, N. D., Laska, K. M., & Wampold, B. E. (2014). The evidence for present- centered therapy as a treatment for posttraumatic stress disorder. *Journal of Traumatic Stress*, *27*, 1-8. https://doi.org/10.1002/jts.21881

Hoge, C. W., Grossman, S. H., Auchterlonie, J. L., Riviere, L. A., Milliken, C. S., & Wilk, J. E. (2014). PTSD treatment for soldiers after combat deployment: Low utilization of mental health care and reasons for dropout. *Psychiatric Services*, *65*, 997-1004. https://doi.org/10.1176/appi.ps.201300307

Hundt, N. E., Harik, J. M., Barrera, T. L., Cully, J. A., & Stanley, M. A. (2016). Treatment decision-making for posttraumatic stress disorder: The impact of patient and therapist characteristics. *Psychological Trauma: Theory, Research, Practice, and Policy*, *8*, 728-735. https://doi.org/10.1037/tra0000102

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

Karlin, B. E., Ruzek, J. I., Chard, K. M., Eftekhari, A., Monson, C. M., Hembree, E. A., ... Foa, E. B. (2010). Dissemination of evidence- based psychological treatments for posttraumatic stress disorder in the Veterans Health Administration. *Journal of Traumatic Stress*, *23*, 663-673. https://doi.org/10.1002/jts.20588

Kessler, R. C., Van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Ebert, D. D., ... Zaslavsky, A. M. (2017). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences*, *26*, 22-36. https://doi.org/10.1017/S2045796016000020

Kraemer, H. C., Frank, E., & Kupfer, D. J. (2006). Moderators of treatment outcomes: Clinical, research, and policy importance. *Journal of the American Medical Association, 296*, 1286-1289. https://doi.org/10.1001/jama.296.10.1286

Kravitz, R. L., Duan, N., & Braslow, J. (2004). Evidence- based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly, 82*, 661-687. https://doi.org/10.1111/j.0887-378x.2004.00327.x

Martinovich, Z., Saunders, S., & Howard, K. (1996). Some comments on "Assessing clinical significance." *Psychotherapy Research, 6*, 124-132. https://doi.org/10.1080/10503309612331331648

Muthén, L. K., & Muthén, B. O. (2017). Mplus (Version 8) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*, 463-469. https://doi.org/10.1111/j.0006-341x.1999.00463.x

Polusny, M. A., Erbes, C. R., Thuras, P., Moran, A., Lamberty, G. J., Collins, R. C., ... Lim, K. O. (2015). Mindfulness-based stress reduction for posttraumatic stress disorder among veterans: a randomized clinical trial. *Journal of the American Medical Association, 314*, 456-465. https://doi.org/10.1001/jama.2015.8361

Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33*, 565-576. https://doi.org/10.1177/0165025409343765

Resick, P. A., Nishith, P., Weaver, T. L., Astin, M. C., & Feuer, C. A. (2002). A comparison of

cognitive-processing therapy with prolonged exposure and a waiting condition for the treatment of chronic posttraumatic stress disorder in female rape victims. *Journal of Consulting and Clinical Psychology, 70*, 867-879. https://doi.org/10.1037//0022-006x.70.4.867

Resick, P. A., Wachen, J. S., Dondanville, K. A., Pruiksma, K. E., Yarvis, J. S., Peterson, A. L., & Mintz, J.; and the STRONG STAR Consortium. (2017). Effect of group vs individual cognitive processing therapy in active-duty military seeking treatment for posttraumatic stress disorder: A randomized clinical trial. *JAMA Psychiatry, 74*, 28-36. https://doi.org/10.1001/jamapsychiatry.2016.2729

Resick, P. A., Wachen, J. S., Mintz, J., Young-McCaughan, S., Roache, J. D., Borah, A. M., ... Peterson, A. L.; on behalf of the STRONG STAR Consortium. (2015). A randomized clinical trial of group cognitive processing therapy compared with group present-centered therapy for PTSD among active duty military personnel. *Journal of Consulting and Clinical Psychology, 83*, 1058-1068. https://doi.org/10.1037/ccp0000016

Rosen, C., Adler, E., & Tiet, Q. (2013). Presenting concerns of veterans entering treatment for posttraumatic stress disorder. *Journal of Traumatic Stress, 26*, 640-643. https://doi.org/10.1002/jts.21841

Rosen, C. S., Matthieu, M. M., Stirman, S. W., Cook, J. M., Landes, S., Bernardy, N. C., ... Watts, B. V. (2016). A review of studies on the system-wide implementation of evidence-based psychotherapies for posttraumatic stress disorder in the Veterans Health Administration. *Administration and Policy in Mental Health and Mental Health Services Research, 43*, 957-977. https://doi.org/10.1007/s10488-016-0755-0

Rothwell, P. M. (2005). External validity of randomised controlled trials: "To whom do the

results of this trial apply?" *The Lancet*, *365*(9453), 82-93. https://doi.org/10.1016/s0140-6736(04)17670-8

Ruscio, A. M., & Holohan, D. R. (2006). Applying empirically supported treatments to complex cases: Ethical, empirical, and practical considerations. *Clinical Psychology: Science and Practice*, *13*, 146-162. https://doi.org/10.1111/j.1468-2850.2006.00017.x

Sayer, N. A., Rosen, C. S., Bernardy, N. C., Cook, J. M., Orazem, R. J., Chard, K. M., ... Schnurr, P. P. (2017). Context matters: Team and organizational factors associated with reach of evidence-based psychotherapies for PTSD in the Veterans Health Administration. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*, 904-918. https://doi.org/10.1007/s10488-017-0809-y

Schnurr, P. P., Friedman, M. J., Engel, C. C., Foa, E. B., Shea, M. T., Chow, B. K., ... Bernardy, N. (2007). Cognitive behavioral therapy for posttraumatic stress disorder in women: A randomized controlled trial. *Journal of the American Medical Association*, *297*, 820-830. https://doi.org/10.1001/jama.297.8.820

Schnurr, P. P., Friedman, M. J., Foy, D. W., Shea, M. T., Hsieh, F. Y., Lavori, P. W., ... Bernardy, N. C. (2003). Randomized trial of trauma-focused group therapy for posttraumatic stress disorder: Results from a Department of Veterans Affairs cooperative study. *Archives of General Psychiatry*, *60*, 481-489. https://doi.org/10.1001/archpsyc.60.5.481

Schumm, J. A., Walter, K. H., & Chard, K. M. (2013). Latent class differences explain variability in PTSD symptom changes during cognitive processing therapy for veterans. *Psychological Trauma: Theory, Research, Practice, and Policy*, *5*, 536-544. https://doi.org/10.1037/a0030359

Shiner, B., Watts, B. V., Pomerantz, A., Young- Xu, Y., & Schnurr, P. P. (2011). Sensitivity of the SF- 36 to PTSD symptom change in veterans. *Journal of Traumatic Stress, 24*, 111-115. https://doi.org/10.1002/jts.20613

Shiner, B., Westgate, C. L., Gui, J., Maguen, S., Young-Xu, Y., Schnurr, P. P., & Watts, B. V. (2018). A retrospective comparative effectiveness study of medications for posttraumatic stress disorder in routine practice. *Journal of Clinical Psychiatry, 79*(5), e1-e10. https://doi.org/10.4088/jcp.18m12145

Speer, D. C., & Greenbaum, P. E. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology, 63,* 1044-1048. https://doi.org/10.1037//0022-006x.63.6.1044

Steenkamp, M. M., & Litz, B. T. (2013). Psychotherapy for military-related posttraumatic stress disorder: Review of the evidence. *Clinical Psychology Review, 33,* 45-53. https://doi.org/10.1016/j.cpr.2012.10.002

Steenkamp, M. M., Litz, B. T., Hoge, C. W., & Marmar, C. R. (2015). Psychotherapy for military-related PTSD: A review of randomized clinical trials. *Journal of the American Medical Association, 314*, 489-500. https://doi.org/10.1001/jama.2015.8370

Stirman, S. W., Finley, E. P., Shields, N., Cook, J., Haine-Schlagel, R., Burgess, J. F., ... Monson, C. (2017). Improving and sustaining delivery of CPT for PTSD in mental health systems: A cluster randomized trial. *Implementation Science*, *12*, 32. https://doi.org/10.1186/s13012-017-0544-5

Tingey, R., Lambert, M., Burlingame, G., & Hansen, N. (1996). Assessing clinical significance: Proposed extensions to method. *Psychotherapy Research*, *6*, 109-123.

https://doi.org/10.1080/10503309612331331638

U.S. Department of Defense. (n.d.) *2014 demographics: Profile of the military community*. Washington, DC: Office of the Deputy Assistant Secretary of Defense. Retrieved from http://download.militaryonesource.mil/12038/MOS/Reports/2014-Demographics-Report.pdf

U.S. Department of Veterans Affairs & U.S. Department of Defense. (2017). *VA/DoD clinical practice guideline for the management of posttraumatic stressdDisorder and acute stress disorder*. Washington, DC: Author. Retrieved from www.healthquality.va.gov/guidelines/MH/ptsd/

Watts, B. V., Shiner, B., Zubkoff, L., Carpenter-Song, E., Ronconi, J. M., & Coldwell, C. M. (2014). Implementation of evidence-based psychotherapies for posttraumatic stress disorder in VA specialty clinics. *Psychiatric Services*, *65*, 648-653. https://doi.org/10.1176/appi.ps.201300176

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). *The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility.* Paper presented at the annual convention of the International Society for Traumatic Stress Studies, San Antonio, TX.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, *130*, 631-663. https://doi.org/10.1037/0033-2909.130.4.631

Table 1 *Demographic and Military Service Characteristics*

		All Trials	US Army 2014	Percentage Difference [CI]	Standard Difference Score
Gender					
	Male	89.4%	86.1%	3 [3, 3]	46.45**
Age					
	≤ 25	16.9%	39.7%	-23 [-23, -23]	165.57**
	26 to 30	28.6%	21.9%	7 [6, 7]	36.37**
	31 to 40	37.1%	27.1%	10 [10, 10]	56.40**
	≥ 41	17.4%	11.3%	6 [6, 7]	29.51**
Military grade					
	E-1 to E-3	0.9%	19.5%	-19 [-19, -18]	143.20**
	E-4 to E-6	80.1%	50.2%	31 [31, 31]	244.19**
	E-7 to E-9	15.1%	10.9%	4 [4, 5]	20.78**
	Officer	1.7%	16.3%	-15 [-15, -14]	106.94**
Education					
	GED/high school	79.6%	76.5%	3 [3, 3]	32.94**
	Associates/ bachelor's degree	17.7%	15.0%	3 [2, 3]	14.20**
	Advanced degree	1.7%	7.9%	-6 [-7, -6]	41.43**

Race					
	American Indian/ Alaskan Native	2.0%	0.7%	1[1, 2]	4.87**
	Asian	1.1%	4.1%	-3 [-3, -3]	-19.34**
	Native Hawaiian or Other Pacific Islander	1.1%	1.0%	.01 [.00, .01]	0.50
	Black or African American	25.7%	21.6%	4 [4, 4]	22.54**
	White	56.0%	67.7%	-12 [-12, -12]	-100.30**
Ethnicity[a]					
	Hispanic/Latino	19.9%	–	–	–
	Not Hispanic/Latino	79.8%	–	–	–
Relationship status					
	Never married	9.9%	35.1%	-25 [-26, -25]	188.36**
	Married	70.4%	59.3%	11 [11, 11]	90.57**
	Divorced/ separated/ widowed/ other	19.7%	5.6%	0.14 [0.14,0.15]	51.80**

Note. CI = confidence interval; E-1 to E-3 = junior enlisted; E-4 to E-6 = junior noncommissioned officers; E-7 to E-9 = senior noncommissioned officers. [a]Ethnicity data for the U.S. Army 2014 were not collected in the 2014 Demographics Profile of the Military Community.

**$p < .001$.

Table 2

Model	-2 log-likelihood	Estimated Parameters	Bayesian Information Criterion (BIC)	Akaike Information Criterion (AIC)
Intercept only	12,428	3	12,447	12,434
Linear[a]	12,143	6	12,182	12,155
Quadratic	11,947	10	12,013	11,967
Bilinear spline	11,867	11	11,939	11,889

Fit Statistics for Growth Models Fit to PSS-I Scores
Note. PSS-I = PTSD Symptom Scale – Interview Version.

[a]Model did not converge to a proper solution.

Table 3

Clinical Change Categories in PTSD Symptoms by Sample

	Clinical Change Categories (all pre- to post-treatment)			
	Recovered	Improved	Unchanged	Deteriorated
Combined STRONG STAR trials				
Completers[a] % (*n*)	20.6% (98)	10.5% (50)	66.5% (316)	2.3% (11)
Intent to treat[b] % (*n*)	13.96% (98)	7.12% (50)	77.35% (543)	1.57% (11)
Schnurr et al., 2007[c]				
Completers % (*n*)	28% (65)	11% (27)	59% (138)	2% (5)
Intent to treat % (*n*)	22.9% (65)	9.51% (27)	65.8% (187)	1.8% (5)
Foa et al., 2013, combined with Foa et al., 2017[d]				
Completers % (*n*)	38.6% (44)	19.3% (22)	41.2% (47)	0.9% (1)
Intent to treat % (*n*)	28.9% (44)	14.5% (22)	55.9% (85)	0.7% (1)
Resick, Nishith, Weaver, Astin & Feuer, 2002[e]				
Completers % (*n*)	72.7% (80)	10% (11)	17.3% (19)	0% (0)
Intent to treat % (*n*)	47% (80)	.06% (11)	47% (80)	0% (0)
Shiner et al., (2018)[f]				
Completers % (*n*)	11% (327)	22% (641)	57% (1,676)	10% (287)

Note. PTSD = posttraumatic stress disorder.

[a]Completers excludes participants without posttreatment data; [b]Intent to treat includes all randomized participants; [c,e]used the Clinician Administered PTSD Scale for *DSM-IV* (CAPS-IV); [d]used the PTSD Symptom Scale – Interview version (PSS-I) (we included only the participants who received PE in the two trials); [f]used the PTSD Checklist for *DSM-IV* (PCL-IV).

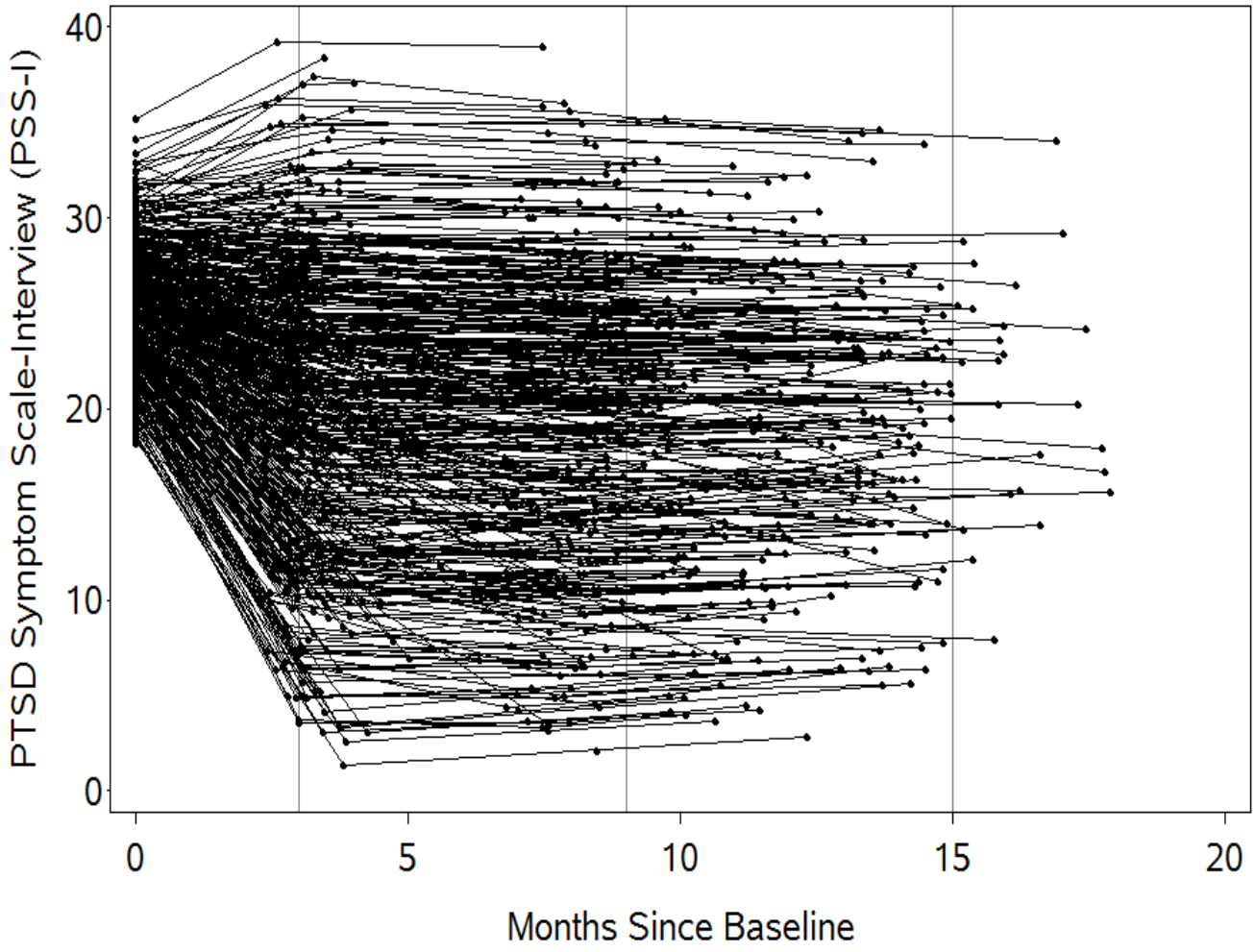

Figure 1. Predicted individual growth trajectories in PSS-I scores estimated by the best-fitting bilinear spline growth model. PSS-I = PTSD Symptom Scale – Interview version; PTSD = posttraumatic stress disorder.