Psychiatry Research 343 (2025) 116287

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Effect of agreement between clinician-rated and patient-reported PTSD symptoms on intensive outpatient treatment outcomes

Sarah C. Touponse [a], Qihang Guo [b], Tianwen Ma [b], Jessica L. Maples-Keller [c], Barbara O. Rothbaum [c], Boadie W. Dunlop [c,*]

[a] Emory University School of Medicine, 531 Asbury Circle, Annex Building Suite N340 Atlanta, GA 30322, United States
[b] Department of Biostatistics and Bioinformatics, Emory University Rollins School of Public Health, 1518 Clifton Rd NE, Atlanta, GA 30322, United States
[c] Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, 12 Executive Park Drive NE, 3rd Floor, Atlanta, GA, 30329, United States

ARTICLE INFO

Keywords:
PTSD
Depression
Clinical trial
Veteran
Military
Prolonged exposure therapy
Psychometrics

ABSTRACT

The level of agreement between self-reported posttraumatic stress disorder (PTSD) symptom severity, measured with the PTSD Checklist for DSM-5 (PCL-5), and severity measured via a clinician-rated measure, the Clinician Administered PTSD Scale for DSM-5 (CAPS-5), was tested as a predictor of the degree of improvement following a 2-week Intensive Outpatient Treatment Program (IOP). Differences in PTSD severity scores (PCL-5 minus CAPS-5) of US Veterans and service members with PTSD ($N = 483$) at Intake were used to categorize patients into 3 agreement groups: Congruent reporters, Limited Over-reporters and Extensive Over-reporters. A linear mixed model tested whether agreement group impacted the degree of improvement, measured as reduction in PCL-5 score, from IOP baseline to completion. The mean difference between the PCL-5 and CAPS-5 scores was 17.5 $\pm$ 13.1 points. Mean modeled reduction in PCL-5 scores from IOP Baseline to IOP Completion for Limited Over-reporters was -21.3 points (95 %CI -23.6, -19.1), which was significantly less than the reduction for Extensive Over-reporters (-27.6, 95 %CI -32.1, -23.1, $p<.001$), but not significantly different from Congruent reporters (-18.0, 95 %CI -22.7, -13.3, $p=.17$). Patients who most over-report their PTSD symptoms compared to trained clinicians show steepest declines in PTSD symptom severity with treatment. Personalizing treatment for PTSD may benefit from understanding the mechanisms contributing to these differences.

1. Introduction

Posttraumatic stress disorder (PTSD) can develop after exposure to a traumatic event. In the United States, estimated rates of lifetime PTSD diagnosis in veterans ranges from 3 % of those who served in World War II to 29 % of those in Operations Iraqi Freedom and Operation Enduring Freedom (Na et al., 2023). Overall, approximately 7 % of individuals with military experience will develop PTSD, most often stemming from exposure to combat, war zone deployments, training accidents, and military sexual trauma (Goldstein et al., 2016).

Treatment guidelines for PTSD recommend trauma-focused therapies as first-line treatments, including cognitive processing therapy (CPT), prolonged exposure therapy (PE), and eye movement desensitization and reprocessing (EMDR) (Hamblen et al., 2019). As treatment discontinuation can significantly hinder progress and recovery in PTSD, increased treatment adherence and completion is essential to improve outcomes (Ragsdale et al., 2020) Over the past decade, Intensive Outpatient Treatment Programs (IOP) are being used more frequently to treat PTSD and minimize drop-out (Harvey et al., 2017). IOP programs deliver more frequent and extended therapy sessions conducted over several hours per day, multiple days per week (Harvey et al., 2017; Peterson et al., 2022), improving retention rates to as high as 95 % (Harvey et al., 2017; Burton et al., 2022).

Past efforts to identify predictors of treatment outcomes and personalization of treatments for PTSD have proven relatively unsuccessful to date, particularly for military veterans. (Resick et al., 2020). Patient demographic factors such as age (van Minnen et al., 2002; Forbes et al., 2003; Rizvi et al., 2009; Resick et al., 2020), race (Rauch et al., 2021), sex (van Minnen et al., 2002; Rizvi et al., 2009), and education (Forbes et al., 2003; Rizvi et al., 2009; Barawi et al., 2020) have been investigated, however, treatment outcome associations have not been consistent. Among psychological variables, comorbid depression or depression severity have failed to emerge as a significant predictor of PTSD outcomes (van Minnen et al., 2002; Hagenaars et al., 2010; Resick

* Corresponding author.
E-mail address: bdunlop@emory.edu (B.W. Dunlop).

https://doi.org/10.1016/j.psychres.2024.116287
Received 20 July 2024; Received in revised form 3 November 2024; Accepted 20 November 2024
Available online 22 November 2024

et al., 2020). The most frequently identified predictor of outcome is pre-treatment PTSD severity, though the data is mixed, with some studies finding more severe patients improving less with treatment and others finding the opposite or no effect (Forbes et al., 2003; Barawi et al., 2020; Resick et al., 2020). Of note, the method of measurement of PTSD symptoms can also impact whether severity predicts outcomes. For example, Karatzias et al. (2007) found that lower clinician-rated symptom severity scores at baseline were associated with better treatment outcomes but that higher self-reported symptom severity scores were also associated with better treatment outcomes. Although not previously explored, it is possible that other clinical factors, such as a history of childhood sexual trauma or severity of depression, may impact the level of severity patients endorse on a PTSD self-report measure relative to a clinician measure. Further understanding the relationships between self-reported and clinician-rated measures of severity may help clarify these discrepant results.

A challenge in treating PTSD and predicting outcomes is the breadth of the twenty types of symptoms that comprise the illness. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) divides the symptoms of PTSD into four clusters: intrusive symptoms; avoidance; negative alterations in cognition and mood; and increased arousal or reactivity (American Psychiatric Association, 2013). Two widely used tools to quantify the severity of PTSD symptoms are the Clinician-Administered PTSD Scale for DSM-5 (CAPS-5) (Weathers et al., 2018) and the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5) self-report questionnaire (Blevins et al., 2015; Bovin et al., 2016). Both instruments assess PTSD severity on a scale of 0–80, with higher scores indicating more severe illness. The CAPS-5 is considered the gold standard for PTSD diagnosis and severity but requires a trained interviewer to administer the semi-structured interview, requiring 40–60 min to complete. In contrast, the PCL-5 is completed independently by the patient, requiring only about 5 min to complete.

Investigations into the respective strengths, limitations, and interchangeability of these tools have yielded valuable insights into the degree of agreement and sources of discordance between the PCL-5 and CAPS-5 measures. Every study comparing the two measures have found higher scores (absolute difference in scores ranging from 7.6 – 13.95), and higher standard deviations (up to 50 % larger), for the PCL-5 than the CAPS-5 (Weathers et al., 2018 Geier et al., 2019, Lee et al., 2022; Kramer et al., 2023; Resick et al., 2023). Studies examining correlations between PCL-5 and CAPS-5 total scores have ranged from $r = 0.66$ (Weathers et al., 2018) to $r = 0.94$ (Geier et al., 2019). Notably, only two of these studies were comprised solely of patients with PTSD (Weathers et al., 2018; Lee et al., 2022); the others had rates of PTSD ranging from 28 to 79 %. Studies that include a mixture of PTSD and non-PTSD patients yield a broader range of scores in the sample, which would generally support higher correlation coefficients. Most studies that compared self-reported and clinician-rated PTSD scoring have concluded that both measures should be used concurrently, due to the differing perspectives they capture. The higher standard deviations in PCL-5 scores suggest that there may be utility in examining differences among patients in their self-report scores.

Beyond levels of agreement in total scores, there may also be value to assessing agreement in individual item scores, as recently examined by Resick et al. (2023). High levels of agreement on total scores could derive from a mixture of over- or under-reporting individual items. Low levels of total score agreement may derive primarily from poor agreement on a small number of items, even if most items show good agreement. Identifying the items that are most discordant between clinician- and self-rated items can point to the symptoms which patients are most likely to misconstrue and inform efforts to improve the clarity of how items are worded.

This study aimed to examine the effect of the agreement level between clinician-rated and patient-reported symptoms of PTSD on treatment outcomes among military veterans with PTSD treated through a specialized 2-week IOP. Specifically, we categorized patients into three levels of agreement in PTSD symptom severity (Extensive Over-reporters, Limited Over-reporters, and Congruent reporters) based on pre-treatment CAPS-5 and PCL-5 scores. We hypothesized that outcomes for patients who showed the greatest elevation of self-reported PTSD symptoms compared to clinician assessment would show the least improvement in PTSD symptoms by the end of IOP treatment. We also explored whether clinical and demographic factors, such as a history of childhood sexual abuse and severity of depression, impacted the level of agreement between clinician-rated and patient-rated PTSD symptoms. Finally, to more closely investigate the symptoms most contributing to discordant total scores, we examined item-level differences between the CAPS-5 and PCL-5.

2. Methods

2.1. Participants and treatment model

The Emory Healthcare Veterans Program (EHVP) recruits current and former military personnel from all over the United States (Burton et al., 2022). Eligibility to participate in the IOP is determined at weekly case conferences by the clinical team. Inclusion for the current analysis required being a service member or veteran who served since 9/11/2001 and having current PTSD based on clinical structured interview. The exclusion criteria were: current, imminent risk of suicide or homicide, unmanaged psychosis or bipolar I disorder, a current substance or alcohol use disorder of a severity that could interfere with treatment, severe cognitive impairment that would interfere with care, medical illness likely to require hospitalization, or lack of a treatment team to provide ongoing care following the program (Rauch et al., 2020). The IOP treatment consists of personalized treatment plans consisting of daily individual (>95 % PE therapy) and group therapy. Daily PE therapy is based on a modified prolonged exposure manual consisting of massed-delivery format, including both imaginal and in vivo exposure (Rauch et al., 2021; Burton et al., 2022). Patients self-reported their birth sex and provided written informed consent to have their data used for research, which was approved by the Emory University Institutional Review Board (Reff: IRB00085250; approved November 6, 2015).

2.2. Assessment and measures

Assessment began with the Intake Visit, which was conducted by phone for >90 % of the participants. The Intake Visit was the only time the CAPS-5 was administered. Immediately following the Intake Visit, the patients are sent self-report questionnaires to complete, including the PCL-5. The intake assessment is conducted by masters- and PhD-level clinicians trained to administer structured clinical interviews, with inter-rater reliability maintained through monthly assessor meetings. The intake assessment includes administration of the CAPS-5 and the Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders (DIAMOND) interview to identify comorbid disorders, including major depressive disorder and substance use disorders (Tolin et al., 2018).

Upon acceptance, patients are scheduled to begin the IOP program on a Monday, the IOP Baseline Visit, at which they complete the PCL-5, which serves as the baseline measure for PTSD symptom change. They also complete the Patient Health Questionnaire-9 (PHQ-9) (Kroenke et al., 2001), and the Childhood Trauma Questionnaire Short Form (CTQ) (Bernstein et al., 2003). After completing the IOP, the patients complete the PCL-5 again, which serves as the outcome measure for the current study.

2.2.1. CAPS-5

The CAPS-5 is a semi-structured interview based on DSM-5 PTSD criteria that assesses symptom severity over the past month The current analysis utilizes the 20 criteria symptom items from the scale, which are

each scored from 0 ("absent") to 4 ("extreme or incapacitating"), for a total score of 0–80. The CAPS-5 has been shown to have good validity and strong inter-rater and test-retest reliability (Weathers et al., 2018).

2.2.2. PCL-5

The PCL-5 is a self-report questionnaire that assesses the 20 criterion symptoms for PTSD. Patients rate their symptoms on a scale from 0 ("not at all") to 4 ("extremely"). At the Intake and IOP Baseline visits, the past-month version of the PCL-5 was used; for the IOP Completion visit, the past-week version was used. Total scores range from 0 to 80, with a higher score indicating increased severity (Blevins et al., 2015). The PCL-5 is one of the most commonly used questionnaires for assessing PTSD due to its externally validity and test-retest reliability, internal consistency, and diagnostic utility, and has proved useful in assessing veterans (Blevins et al., 2015; Bovin et al., 2016) and robustly measures PTSD symptom change and is strongly correlated with improvement as assessed by the CAPS-5 (Lee et al., 2022).

2.2.3. Other measures

The PHQ-9 is a self-reported questionnaire that assesses the severity over the past week of the nine criterion symptoms for major depressive disorder. Symptoms are rated from 0 ("not at all") to 3 ("nearly every day") for the past week, and total scores range from 0 to 27 (Kroenke et al., 2001). The measure has high internal consistency and good sensitivity to change with large pre- to post-treatment effect sizes (Cohen's $d > 1.0$) in psychiatric samples (Beard et al., 2016). A score $\geq$10 has high sensitivity for detecting major depressive disorder (Manea et al., 2015). The CTQ is a self-administered 25-item scale (plus three unscored validity items) that asks about the frequency with which patients experienced a variety of childhood maltreatment experiences. Symptoms are rated on a scale of 1 ("never true") to 5 ("very often true"), for a total score ranging from 25 to 125, comprised of five subscales (Bernstein et al., 2003). The scale has demonstrated high structural validity, construct validity, and internal consistency, with adequate reliability; cross-cultural validity has not been established (Baker and Maiorino, 2010; Georgieva et al., 2021). CTQ total scores can be divided into categories of severity, ranging from "none/minimal" ($\leq$36), "low to moderate" (>36 - $\leq$51), "moderate to severe" (>51 - $\leq$68), and "severe to extreme" ($\geq$69) (Bernstein and Fink, 1998).

2.3. Definition of agreement groups

A difference score for each patient was calculated by subtracting the CAPS-5 Intake Visit score from the PCL-5 Intake Visit score, and the overall mean and standard deviation (SD) for the difference score were calculated. Patients were then categorized into one of three agreement groups. Extensive Over-reporters were those with a difference score >1 SD above the mean; Limited Over-reporters were within one SD of the mean, and Congruent reporters were >1 SD below the mean.

2.4. Statistical analyses

All statistical analyses were conducted using R version 4.3 (R Core Team 2021) or GraphPad Prism 8 (GraphPad Software, San Diego, CA, USA). Continuous variables (age, CAPS-5, PCL-5, PHQ-9, and CTQ) were summarized using means and SDs; all other variables were categorical and summarized as counts and percentages. The CAPS-5 and PCL-5 scores at intake and their difference score were plotted as histograms to visualize the distributions. The relationships between the CAPS-5 and PCL-5 at intake and between the PCL-5 scores at intake and at baseline were graphed and analyzed with Pearson's correlation. P-values and 95 % confidence intervals (CI) were reported. As the distributions of the scale scores had near normal distributions, *t*-tests were conducted on PCL-5 scores between agreement groups at baseline and at the completion of IOP. Internal consistency of the CAPS-5 at intake and the PCL-5 at all timepoints was determined using Cronbach's alpha; values >0.70 are considered to represent acceptable internal consistency (Tavakol and Dennick, 2011). To test whether the distribution of the three agreement groups differed by demographic or clinical factors, Chi-square and ANOVA tests were first performed on categorical and continuous covariates, respectively. For covariates that showed statistical significance, univariate multinomial logistic regression models were then fitted to quantify the association. Parameter estimates, 95 % confidence intervals, and p-values were reported.

A linear mixed model (LMM) was fitted to examine the changes in PCL-5 scores over the three agreement groups, starting from the Baseline PCL-5 score, adjusting for demographic and clinical factors that showed statistical significance from Chi-square and ANOVA tests. The fixed effects included agreement groups, time points, their interactions, and PHQ-9 scores, while the random effects only included a patient-specific random intercept. The final model was selected based on AIC value; race was not included due to the strong collinearity with agreement group. Parameter estimates, 95 % confidence intervals, and p-values were reported. Observed and predicted group-specific means and 1-SD error bars at different time points were reported.

For the item-level analyses, for each of the 20 DSM-5 PTSD symptoms, we subtracted the intake visit CAPS-5 item score from the respective item score on the intake visit PCL-5. These item difference scores were calculated for the combined sample, as well as within each agreement group. A Wilcoxon signed-rank test was applied to examine whether the overall PCL-5 scores were significantly different from the overall CAPS-5 scores.

For all statistical analyses, p-values <0.05 were considered as statistically significant.

3. Results

3.1. Participants

The study consisted of 483 EHVP veterans and service members who met full DSM-5 criteria for current PTSD and had CAPS-5 and PCL-5 Intake Visit scores (**Table S1,** Supplementary Material). The sample age range was ranging 20–69 years (mean: 39.80±9.30) and was 40 % (n= 193) female. The analyses of change with treatment utilized the 456 veterans (94.40 %) who completed IOP treatment.

3.2. Pre-treatment CAPS-5 and PCL-5 scores

Cronbach's alpha for the CAPS-5 at intake was 0.70, and for the PCL-5 it was >0.90 at all timepoints, indicating adequate internal consistency. At the Intake Visit, total scores for the CAPS-5, the PCL-5, and their difference score were all normally distributed (**Figure S1,** Supplementary Material). CAPS-5 and PCL-5 total scores at the Intake Visit were moderately positively correlated (r= 0.42, p<.001; Fig. 1A). The mean PCL-5 total score (53.60±14.20) was 17.50±13.1 points higher than the mean CAPS-5 total score (36.10±8.50) in the overall sample. At the Intake visit, the mean difference for Extensive Over-reporters was 36.21 (PCL-5: 68.01±9.84; CAPS-5: 31.80±6.77); for Limited Over-reporters 17.56 (54.53±10.89; 36.97±8.68), and for Congruent reporters −3.28 (33.69±9.84; 36.97±8.18).

The scores on the PCL-5 at the Intake Visit, which was when the CAPS-5 was administered, were strongly correlated with the scores on the PCL-5 at the Baseline Visit (r= 0.67, p<.001; Fig. 1B).

3.3. Relationships between agreement groups and demographic and clinical factors

Extensive Over-reporters (n= 83), Limited Over-reporters (n= 325), and Congruent reporters (n= 75) comprised the three agreement groups (Table 1). Race and PHQ-9 scores significantly differed among the three agreement groups (p=.007 and p<.001). Two univariate multinomial logistic regression models indicated that Black patients were more

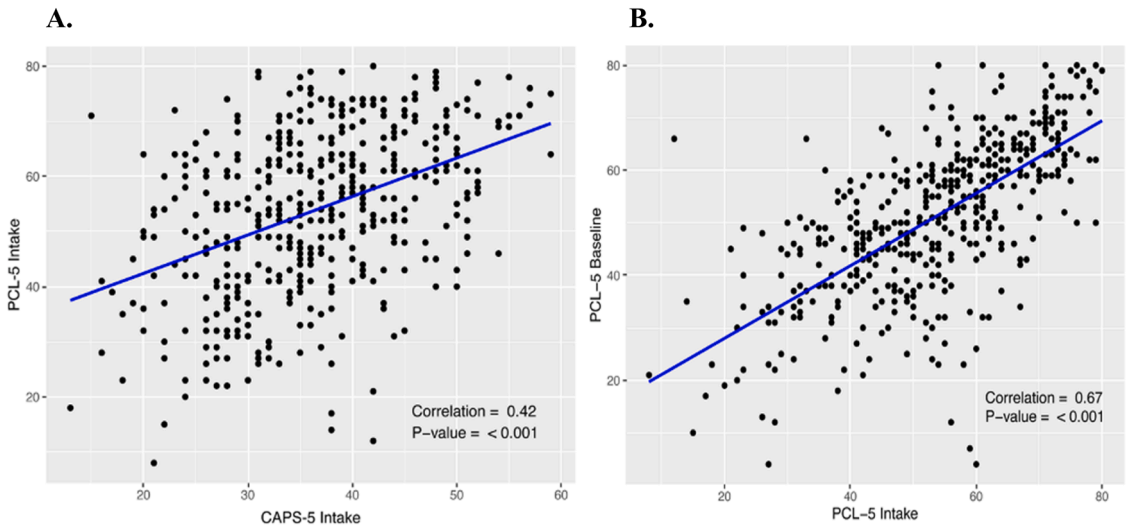

Fig. 1. Scatterplots of CAPS-5 and PCL-5 scores. A. Correlation between CAPS-5 and PCL-5 scores at the Intake Visit. B. Correlation between PCL-5 scores at Intake Visit and PCL-5 scores at IOP Baseline Visit.

likely, and multi-racial patients less likely, than White patients to be represented in the Extensive Over-reporters group compared to the Limited Over-reporters group. In addition, higher scores on the PHQ-9 reduced the odds of patients being included in the Congruent reporters group compared to Limited Over-reporters, indicating that self-report of more depressive symptoms contributed to self-report of higher PCL-5 scores (**Table S2,** Supplementary Material). Childhood trauma severity (CTQ score) was not related to agreement group.

3.4. Agreement group and treatment outcomes

The modeled mean change in PCL-5 scores for from IOP Baseline to Completion among Limited Over-reporters was −21.3 points (95 % CI −19.1, −23.6, $p<.001$) (Table 2). Observed mean PCL-5 reduction was greater for the Extensive Over-reporters (−27.2, 95 % CI −32.2, −22.3), and smaller for the Congruent reporters (−17.6, 95 % CI −22.2, −12.9, $p<.001$). These group differences were statistically significant ($p=.02$) in the linear mixed model adjusting for covariates (Fig. 2). The reduction in PCL-5 scores for Extensive Over-reporters was significantly greater than the reductions for both Limited Over-reporters ($p=.03$) and Congruent reporters ($p=.005$), but PCL-5 reductions between Limited Over-reporters and Congruent reporters did not significantly differ ($p=.16$). Additionally, for each additional point higher on the PHQ-9 at the Baseline Visit, PCL-5 scores at the IOP Completion Visit were higher, on average, by 1.22 points ($p<.001$, 95 % CI: [0.96, 1.47]).

3.5. Item-level analysis

The item-level analysis of the overall combined Intake sample (**Tables 3 and S2,** Supplementary Material) found that the top five items with the greatest positive mean differences were items 3 (flashbacks), 5 (physical reactivity), 8 (psychogenic amnesia), 15 (angry outbursts), and 18 (exaggerated startle). No item was under-reported in the combined sample. The CAPS-5 and PCL-5 scores for each of the 20-items were all significantly different ($p<.001$).

For Extensive Over-reporters (**Figure S3A,** Supplementary Material), the top five items for which the PCL-5 item score was most elevated over the respective CAPS-5 item score were items 1 (unwanted memories; 2.05 ± 1.25), 3 (2.75 ± 1.19), 5 (2.25 ± 1.16), 8 (2.14 ± 1.46), and 15 (2.14 ± 1.15). For Congruent reporters, the top five items for which the PCL-5 score was the most reduced from the respective CAPS-5 score were items 6 (internal avoidance; −0.7 ± 1.14), 9 (negative beliefs; −0.3 ± 1.26), 11 (negative emotions; −0.7 ± 1.01), 13 (detachment; −0.4 ± 0.95), and 17 (hypervigilance; −0.5 ± 1.01) (**Figure S3C,** Supplementary Material). Although Congruent reporters had PCL-5 item scores lower than CAPS-5 item scores for most items, there were six items (3, 5, 8, 15, 16 (recklessness), and 18 (exaggerated startle)) in which the PCL-5 score were actually higher than the CAPS-5 score, four of which (3, 5, 8, and 15) were among the top five most elevated items in the Extensive Over-reporters group (Table 3).

4. Discussion

4.1. Main findings

This study examined the relationship between clinician-rated and patient-reported measures of PTSD symptom severity and their association with symptomatic improvement after treatment through an IOP for veterans with PTSD. As expected, in the overall sample, mean PCL-5 total scores and individual item scores were higher than mean CAPS-5 total and respective item scores. Notably, our analysis showed substantial variability between patients in their level of agreement for the totals and many of the item scores, indicating that there may be value in categorizing patients based on their degree of concordance between the self-reported and clinician-rated measures. Indeed, although there were large reductions on average in PCL-5 scores from baseline to end of IOP treatment, patients who most over-reported their symptom severity compared to the clinician rating (defined as those with a difference in PCL-5 versus CAPS-5 scores at intake more than one standard deviation above the mean difference for the sample) showed more improvement in PCL-5 scores post-treatment compared to Congruent reporters. Given that few predictors of outcome to PTSD treatment in veterans have been previously identified, our finding that the level of agreement between self- and clinician-rated measures of PTSD severity can serve to predict the level of improvement may have utility for patient care and clinical trial design.

The overall correlation between the CAPS-5 and PCL-5 total scores in our sample was only moderate, $r = 0.42$, lower than any prior studies examining this association (Weathers et al., 2018 Geier et al., 2019; Kramer et al., 2023; Resick et al., 2023). Likely contributors to our lower correlation include: our sample was exclusively comprised of military service members with PTSD, a clinical, treatment-seeking sample rather than a clinical trials sample, and a broader diversity of the sample. These characteristics may make our sample more broadly generalizable than prior studies, and suggest that prior studies reporting stronger correlations represent optimally strong associations between the PCL-5 and

Table 1
Clinical and demographic characteristics associated with agreement group.

Patient characteristic	Extensive Over-reporters ($n = 83$)	Limited Over- reporters ($n = 325$)	Congruent reporters ($n = 75$)	p-value
Age, mean (SD)	40.86 (8.92)	39.59 (9.38)	39.53 (9.43)	0.36
Female, n (%)	34 (40.96)	132 (40.62)	27 (36.00)	0.75
Race, n (%)				**0.007**
White	30 (36.14)	190 (58.46)	50 (66.67)	
Black	32 (38.55)	97 (29.85)	15 (20.00)	
Asian/Pacific Islander	3 (3.61)	6 (1.85)	1 (1.33)	
Native American	1 (1.20)	2 (0.62)	0 (0.00)	
Multi-racial	5 (6.02)	5 (1.54)	1 (1.33)	
Missing	12 (14.46)	25 (7.69)	8 (10.67)	
Ethnicity, n (%)				0.81
Hispanic/Latino	7 (8.43)	30 (9.23)	5 (6.67)	
Non-Hispanic/Latino	66 (79.52)	274 (84.31)	63 (84.00)	
Missing/Not reported	10 (12.05)	21 (6.46)	7 (9.33)	
Education Level, n (%)				0.16
Less than a Bachelor's degree	20 (24.10)	114 (35.08)	29 (38.67)	
Bachelor's degree or higher	56 (67.47)	198 (60.92)	43 (57.33)	
Missing/Not reported	7 (8.43)	13 (4.00)	3 (4.00)	
Employment status, n (%)				0.83
Employed	27 (32.53)	124 (38.15)	31 (41.33)	
Unemployed	21 (25.30)	75 (23.08)	16 (21.33)	
Retired	5 (6.02)	31 (9.54)	6 (8.00)	
Disabled/Unable to work	15 (18.07)	49 (15.08)	10 (13.33)	
Student	4 (4.82)	25 (7.69)	8 (10.67)	
Missing/Not reported	11 (13.25)	21 (6.46)	4 (5.33)	
Rank, n (%)				0.38
Enlisted (E1–9)	62 (74.70)	277 (85.23)	62 (82.67)	
Officer (WO2–5, O2–6)	13 (15.66)	36 (11.08)	10 (13.33)	
Missing/Not reported	8 (9.64)	12 (3.69)	3 (4.00)	
Branch of service, n (%)				0.96
Army	52 (62.65)	198 (60.92)	48 (64.00)	
Air Force	7 (8.43)	40 (12.31)	11 (14.67)	
Marines	7 (8.43)	32 (9.85)	7 (9.33)	
Navy	7 (8.43)	29 (8.92)	5 (6.67)	
Coast Guard	2 (2.41)	5 (1.54)	0 (0.00)	
Multiple	4 (4.82)	13 (4.00)	3 (4.00)	
Missing/Not reported	4 (4.82)	8 (2.46)	1 (1.33)	
Clinical Measures				
CAPS-5 score at intake, mean (SD)	31.80 (6.77)	36.97 (8.68)	36.97 (8.18)	**<0.001**
PCL-5 score at intake, mean (SD)	68.01 (6.91)	54.53 (10.89)	33.69 (9.84)	**<0.001**
PCL-5 score at IOP baseline, mean (SD)	60.32 (12.16)	51.53 (13.33)	39.42 (14.28)	**<0.001**
PCL-5 score at IOP completion, mean (SD)	35.43 (21.04)	29.77 (19.01)	21.14 (16.10)	**<0.001**
MDD present, n (%)	28 (33.73)	95 (29.23)	20 (26.67)	0.60
PHQ9 score at intake, mean (SD)	17.59 (4.67)	16.60 (5.13)	14.15 (5.71)	**<0.001**
CTQ score, mean (SD)	52.01 (25.77)	50.44 (21.27)	48.83 (19.33)	0.38
Substance use disorder present, n (%)	16 (19.28)	33 (10.15)	9 (12.00)	0.074

Bold face indicates statistical significance defined as p-value 0.05. Percentages are calculated with agreement group.

CAPS-5.

Although previously used in studies of major depressive disorder (Dorz et al., 2004; Dunlop et al., 2011; Hershenberg et al., 2020), no prior work has examined differences between self-reported and clinician-rated measures of PTSD symptom severity as a predictor of treatment outcomes. Past studies in depression have found that patients who over-reported depression symptom severity compared to clinician raters had slower time to remission and reduced overall response rates to acute treatment compared to congruent or under-reporting patients (Dunlop et al., 2011). In contrast, and contrary to our hypothesis, patients who most over-reported PTSD symptoms at treatment baseline demonstrated more improvement in self-reported symptom severity at

Table 2
Linear mixed model results of time variable, race, PHQ-9, and agreement groups.

Covariate	Estimate	95 % Lower	95 % upper	P-value	Overall P-Value
Time: Completion	−21.32	−23.58	−19.06	**<0.001**	**<0.001**
Time: Baseline	Reference				
Extensive Over-reporters	9.22	4.54	13.90	**<0.001**	**<0.001**
Congruent reporters	−10.03	−14.47	−5.59	**<0.001**	
Limited Over-reporters	Reference				
PHQ-9 Score at Intake	1.22	0.96	1.47	**<0.001**	**<0.001**
Time: Completion * Extensive Over-reporters	−5.91	−11.35	−0.47	**0.034**	**0.020**
Time: Completion * Congruentreporters	3.74	−1.43	8.92	0.157	
Time: Completion * Limited Over-reporters	Reference				
Time: Baseline * Extensive Over-Reporters					
Time: Baseline * Congruent reporters					
Time: Baseline * Limited Over-reporters					

Bold font indicates statistical significance defined as p-value < 0.05.

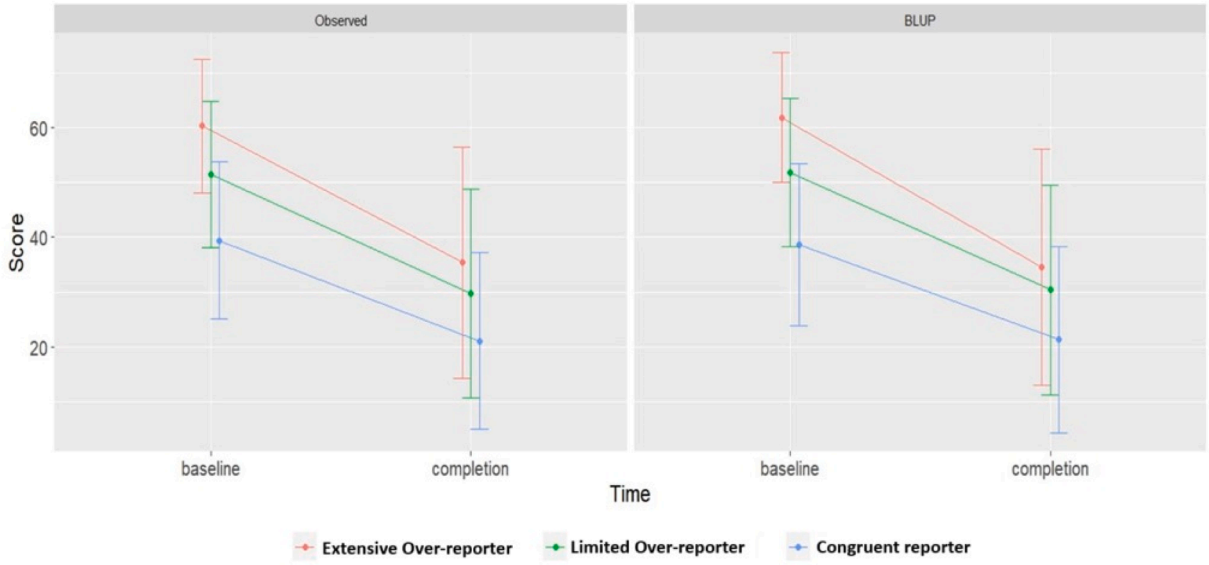

Fig. 2. Change in observed and modeled PCL-5 scores from Baseline to Completion of IOP, by Agreement Group. Observed and best linear unbiased prediction (BLUP) estimate for group-specific PCL-5 means and 1-SD error bars at baseline and IOP completion. BLUP is the based on the final linear mixed models. The two plots show similar patterns.

post-treatment than other patients.

A potential explanation for this result is that, through the process of treatment, Extensive Over-reporting patients may learn how to better distinguish the concepts that the items are inquiring about. The 'halo effect' refers to a cognitive bias that distorts one's perception of a person or thing due to one's overall impression of it (Thorndike, 1920); in the case of self-report questionnaires, some patients may respond to questions in a global manner rather than to each item on its own terms, resulting in upward scoring of unrelated items (Edwards et al., 1984). Alternatively, high levels of over-reporting may reflect a manner of distress expression, whereby some patients may select the extreme ends of score options as a means of conveying their need for help. Supporting this view, a recent report by Resick and colleagues (2023) of active-duty veterans treated for PTSD found that after treatment, the proportion of items endorsed at the highest option (4) on the PCL-5 reduced substantially, even though the proportion of rater-assessed CAPS-5 items endorsed at the maximal score was essentially unchanged. It may be that patients who are categorized as Extensive Over-reporters at pre-treatment may benefit from targeted psychoeducation about the variety of symptoms that make up PTSD and the features that define and distinguish them, which may reduce both halo and distress expression effects.

Another potential application of these results may be for reducing placebo response and clarifying outcomes from clinical trials, specifically by informing trial eligibility criteria, the choice of outcome measures, and the interpretation of treatment outcomes for patients based on their pre-treatment level of symptom severity agreement. It is possible that a psychoeducation module for trial participants about the symptoms of PTSD between screening and baseline visits could improve patient understanding about the types of psychological and behavioral phenomena each item incorporates. Such psychoeducation may help reduce discrepancies between clinician- and patient-reported scales and reduce the contributions to placebo response that stem from measurement error.

Differences between the CAPS-5 and PCL-5 total and item scores have been shown in recent studies (Kramer et al., 2023; Resick et al., 2023); however, by creating categories of patients based on the total difference score, we were able to identify the item-level differences that most impacted total score concordance across the agreement groups. In the combined sample, the mean item scores on the PCL-5 were substantially higher (>1 point mean difference) than those on the CAPS-5 for the items assessing flashbacks (item 3), physical reactivity to trauma reminders (item 5), psychogenic amnesia (item 8), and angry outbursts (item 15). Kramer et al. (2023) found that the CAPS-5 and PCL-5 items with the lowest correlations between intensity and severity included flashbacks (item 3), avoidance (items 6 and 7), hypervigilance (item 17), and exaggerated startle (item 18). Similarly, Resick et al. (2023) found that items 3, 5 (physical reactivity to trauma reminders), 6, and 17 had the lowest CAPS-5/PCL-5 correlation at baseline. Taken together, these results indicate that patients may not conceptualize certain PTSD symptoms in the same way clinician interviewers do. CAPS assessment requires clarification that symptoms are related to the traumatic event, whereas the instructions for the PCL-5 do not emphasize this temporal sequence. These results also support the possibility that score discrepancy may stem from Extensive Over-reporters having a reduced understanding of what the PCL-5 symptom items represent compared to clinicians (Kraemer et al., 2024), and their increased improvement may be due to increased understanding of these symptom constructs. Interestingly, the symptoms that Congruent reporters minimized did not overlap with the items most elevated in the Extensive Over-reporters. Specifically, the items Congruent reporters scored the lowest compared to the trained raters were avoidance of internal reminders, hypervigilance, negative emotions, negative beliefs, and detachment.

Resick et al. (2023) compared item-level CAPS-5 and PCL-5 scores and argued that the baseline elevations in PCL-5 scores they observed were most likely not due to patients' increased distress reporting among their sample of active-duty military personnel. In the combined sample, our results largely replicated those of Resick et al., finding elevated PCL-5 scores in the combined sample. However, by categorizing patients into the three agreement groups, we identified a subset of Extensive Over-reporters whose elevations of PCL-5 scores were substantially greater than those of other patients. This finding suggests that high distress expression manifested through selection of extreme responses on the questionnaire, may be a driver for some patients' poor congruence with clinician severity assessments.

4.2. *Strengths and limitations*

This study had several important strengths. First, the sample was the

Table 3
Means and standard deviations of item-level difference scores by agreement group.

Item Number (Description)	Combined Sample (N = 483)	Extensive Over-reporters (n = 83)	Limited Over-reporters (n = 325)	Congruent reporters (n = 75)
1 (Unwanted memories)	1.10 (1.37)	**2.05 (1.25)**	1.05 (1.27)	−0.02 (1.10)
2 (Unpleasant dreams)	0.77 (1.24)	1.63 (1.14)	0.68 (1.15)	−0.05 (1.13)
3 (Flashbacks)	**1.76 (1.28)**	**2.75 (1.19)**	1.69 (1.16)	0.72 (0.98)
4 (Psychological reactivity)	0.87 (1.15)	1.65 (1.11)	0.83 (1.01)	−0.09 (1.06)
5 (Physical reactivity)	**1.14 (1.28)**	**2.25 (1.16)**	1.03 (1.11)	0.05 (1.07)
6 (Avoidance, internal)	0.58 (1.25)	1.41 (1.03)	0.60 (1.11)	**−0.72 (1.14)**
7 (Avoidance, external)	0.89 (1.36)	1.92 (1.14)	0.81 (1.21)	−0.21 (1.36)
8 (Psychogenic amnesia)	**1.18 (1.41)**	**2.14 (1.46)**	1.09 (1.29)	0.28 (1.12)
9 (Negative beliefs)	0.83 (1.31)	1.68 (1.24)	0.81 (1.15)	**−0.30 (1.26)**
10 (Excessive blaming)	0.77 (1.49)	1.76 (1.50)	0.68 (1.37)	−0.19 (1.26)
11 (Negative emotions)	0.66 (1.19)	1.60 (1.04)	0.67 (0.99)	**−0.72 (1.01)**
12 (Loss of interest)	1.08 (1.34)	1.83 (1.30)	1.11 (1.22)	−0.14 (1.10)
13 (Detachment)	0.70 (1.22)	1.59 (0.98)	0.66 (1.14)	**−0.40 (0.95)**
14 (Lack of positive emotions)	0.71 (1.21)	1.62 (0.99)	0.62 (1.09)	−0.16 (1.25)
15 (Angry outbursts)	**1.23 (1.23)**	**2.14 (1.15)**	1.19 (1.06)	0.09 (1.13)
16 (Recklessness)	0.91 (1.15)	1.60 (1.41)	0.86 (1.02)	0.19 (0.79)
17 (Hypervigilance)	0.56 (1.13)	1.41 (0.93)	0.51 (1.02)	**−0.47 (1.01)**
18 (Exaggerated startle)	**1.38 (1.28)**	1.92 (1.27)	1.40 (1.20)	0.51 (1.24)
19 (Impaired concentration)	1.00 (1.16)	1.76 (1.06)	1.01 (1.03)	−0.14 (1.01)
20 (Insomnia)	0.79 (1.11)	1.29 (1.01)	0.83 (1.05)	−0.19 (0.98)

Bold face reflects the top 5 items contributing to reduced agreement in each group.

largest single study sample to compare CAPS-5 and PCL-5 agreement (N = 483), all of whom had PTSD, and was very diverse in terms of gender, race, and military rank, supporting the generalizability of our findings. Second, the sample was very well-characterized, enabling the evaluation of a broad array of covariates. Third, the patients in this sample received high quality, evidence-based therapy in a massed-delivery format in the IOP, allowing strong testing of the predictive value of agreement group and which minimized treatment drop-out below 6 %. A potential limitation of the study is that the CAPS-5 was assessed only at intake, and not repeated at IOP completion due to the substantial time required to conduct the scale, particularly within a clinical treatment setting. Repeating the CAPS-5 at post-treatment would have given a more complete measure of symptom change that incorporated both clinician and patient perspectives. Clinical trials that use the CAPS-5 as an outcome measure could explore whether agreement level changes post-treatment or as a function of degree of improvement. Another measurement-related limitation was that at baseline patients completed the past-month version of the PCL-5, whereas at IOP completion the past-week version was used. Whether using different time frames has an impact on the degree of change with treatment has not been established. Measuring the outcome after only 2 weeks of treatment could be considered as a limitation, though other analyses of this sample have found that approximately 85 % of IOP-treated veterans maintained their improvement in PTSD symptoms out to one year of follow-up. Finally, the term "over-reporter" implies that the clinician rating is more accurate, which is not objectively provable.

5. Conclusions

This study contributes to the growing literature on advancing PTSD treatment for veterans by utilizing a novel approach of categorizing patients into groups based on the level of agreement between the clinician and patient PTSD severity scores. In addition to replication of our findings in other datasets, future studies should examine other factors that could drive differences in symptom reporting. Previously completed trials, particularly those in which an active treatment failed to prove superior to placebo, could be reanalyzed (if both clinician-administered and self-reported symptom measures were collected) to examine whether outcomes differed by agreement level, or if agreement level moderated the observed effects. Given the substantial variability in levels of symptom agreement, this analysis supports the recommendation that both clinician-rated and patient-reported measures of outcomes be utilized in clinical research to fully understand the benefits of treatment.

CRediT authorship contribution statement

Sarah C. Touponse: Data curation, Formal analysis, Writing – original draft. **Qihang Guo:** Formal analysis, Writing – review & editing. **Tianwen Ma:** Formal analysis, Writing – review & editing. **Jessica L. Maples-Keller:** Data curation, Methodology, Writing – review & editing. **Barbara O. Rothbaum:** Methodology, Writing – review & editing. **Boadie W. Dunlop:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

JM-K has received research funding and consulting payments from COMPASS Pathways, and receives support from the Wounded Warrior Project (WWP), and Multidisciplinary Association of Psychedelic Studies. BOR has or recently had funding from Wounded Warrior Project, NIMH, National Science Foundation, and the Bob Woodruff Foundation, receives royalties from Oxford University Press, Guilford, APPI, Psych Campus, and Emory University, and received advisory board payments from Jazz Pharmaceuticals, Bioserenity, Cerebral Therapeutics, Otsuka, Psychwire, and Senseye. BOR owns equity in Virtually Better, Inc. that creates virtual environments. BWD has received research support from Boehringer Ingelheim, Compass Pathways, Intra-Cellular Therapies, NIMH, Otsuka, Usona Institute, and has served as a consultant for Biohaven, Cerebral Therapeutics, Myriad Neuroscience, and Otsuka. SJC, QG, and TM report no disclosures.

Acknowledgments

The authors acknowledge, with gratitude, critical support from Wounded Warrior Project, who has supported The Emory Healthcare Veterans Program (EHVP) and served as a partner in the Warrior Care Network-dedicated to filling gaps in mental health care for the invisible wounds of war in service members, Veterans and military families.

Funding

The Wounded Warrior Project funded the care and data collection of the veterans treated through this clinical program. The Wounded

Warrior Project had no role in the study design, analysis and interpretation of data, writing of the report, or decision to submit the article for publication.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2024.116287.

Data availability statement

Research data are not shared.

References

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. American Psychiatric Association, Washington D.C fifth ed.

Baker, A.J.L., Maiorino, E., 2010. Assessments of emotional abuse and neglect with the CTQ: issues and estimates. Child. Youth Serv. Rev. 32, 740–748. https://doi.org/10.1016/j.childyouth.2010.01.011.

Barawi, K.S., Lewis, C., Simon, N., Bisson, J.I., 2020. A systematic review of factors associated with outcome of psychological treatments for post-traumatic stress disorder. Eur. J. Psychotraumatol. 11, 1774240. https://doi.org/10.1080/20008198.2020.1774240.

Beard, C., Hsu, K.J., Rifkin, L.S., Busch, A.B., Björgvinsson, T., 2016. Validation of the PHQ-9 in a psychiatric sample. J. Affect. Disord. 193, 267–273. https://doi.org/10.1016/j.jad.2015.12.075.

Bernstein, D.P., Stein, J.A., Newcomb, M.D., Walker, E., Pogge, D., Ahluvalia, T., Stokes, J., Handelsman, L., Medrano, M., Desmond, D., Zule, W., 2003. Development and validation of a brief screening version of the childhood trauma questionnaire. Child. Abuse Negl. 27, 169–190. https://doi.org/10.1016/s0145-2134(02)00541-0.

Bernstein, D.P., Fink, L., 1998. Childhood Trauma Questionnaire: A Retrospective Self Report. Manual. The Psychological Corporation. Harcourt Brace and Company, San Antonio, TX.

Blevins, C.A., Weathers, F.W., Davis, M.T., Witte, T.K., Domino, J.L., 2015. The posttraumatic stress disorder checklist for *DSM-5* (PCL-5): development and initial psychometric evaluation. J. Trauma Stress. 28, 489–498. https://doi.org/10.1002/jts.22059.

Bovin, M.J., Marx, B.P., Weathers, F.W., Gallagher, M.W., Rodriguez, P., Schnurr, P.P., Keane, T.M., 2016. Psychometric properties of the PTSD checklist for diagnostic and statistical manual of mental disorders–fifth edition (PCL-5) in veterans. Psychol. Assess. 28, 1379–1391. https://doi.org/10.1037/pas0000254.

Burton, M.S., Rothbaum, B.O., Rauch, S.A.M., 2022. The role of depression in the maintenance of gains after a prolonged exposure intensive outpatient program for posttraumatic stress disorder. Depress. Anxiety. 39, 315–322. https://doi.org/10.1002/da.23240.

Dorz, S., Borgherini, G., Conforti, D., Scarso, C., Magni, G., 2004. Comparison of self-rated and clinician-rated measures of depressive symptoms: a naturalistic study. Psychol. Psychother. 77, 353–361. https://doi.org/10.1348/1476083041839349. Pt 3.

Dunlop, B.W., Li, T., Kornstein, S.G., Friedman, E.S., Rothschild, A.J., Pedersen, R., Ninan, P., Keller, M., Trivedi, M.H., 2011. Concordance between clinician and patient ratings as predictors of response, remission, and recurrence in major depressive disorder. J. Psychiatr. Res. 45, 96–103. https://doi.org/10.1016/j.jpsychires.2010.04.032.

Edwards, B.C., Lambert, M.J., Moran, P.W., McCully, T., Smith, K.C., Ellingson, A.G., 1984. A meta-analytic comparison of the beck depression inventory and the Hamilton Rating Scale for depression as measures of treatment outcome. Br. J. Clin. Psychol. 23, 93–99. https://doi.org/10.1111/j.2044-8260.1984.tb00632.x.

Forbes, D., Creamer, M., Hawthorne, G., Allen, N., McHugh, T., 2003. Comorbidity as a predictor of symptom change after treatment in combat-related posttraumatic stress disorder. J. Nerv. Ment. Dis. 191, 93–99. https://doi.org/10.1097/01.NMD.0000051903.60517.98.

Georgieva, S., Tomas, J.M., Navarro-Pérez, J.J., 2021. Systematic review and critical appraisal of Childhood Trauma Questionnaire - Short Form (CTQ-SF). Child. Abuse Negl. 120, 105223. https://doi.org/10.1016/j.chiabu.2021.105223.

Goldstein, R.B., Smith, S.M., Chou, S.P., Saha, T.D., Jung, J., Zhang, H., Pickering, R.P., Ruan, W.J., Huang, B., Grant, B.F., 2016. The epidemiology of DSM-5 posttraumatic stress disorder in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions-III. Soc. Psychiatry Psychiatr. Epidemiol. 51, 1137–1148. https://doi.org/10.1007/s00127-016-1208-5.

Hagenaars, M.A., van Minnen, A., Hoogduin, K.A.L., 2010. The impact of dissociation and depression on the efficacy of prolonged exposure treatment for PTSD. Behav. Res. Ther. 48, 19–27. https://doi.org/10.1016/j.brat.2009.09.001.

Hamblen, J.L., Norman, S.B., Sonis, J.H., Phelps, A.J., Bisson, J.I., Nunes, V.D., Megnin-Viggars, O., Forbes, D., Riggs, D.S., Schnurr, P.P., 2019. A guide to guidelines for the treatment of posttraumatic stress disorder in adults: an update. Psychotherapy 56, 359–373. https://doi.org/10.1037/pst0000231.

Harvey, M.M., Rauch, S.A.M., Zalta, A.K., Sornborger, J., Pollack, M.H., Rothbaum, B.O., Laifer, L.M., Simon, N.M., 2017. Intensive treatment models to address posttraumatic stress among post-9/11 warriors: the warrior care network. Focus (Madison) 15, 378–383. https://doi.org/10.1176/appi.focus.20170022.

Hershenberg, R., McDonald, W.M., Crowell, A., Riva-Posse, P., Craighead, W.E., Mayberg, H.S., Dunlop, B.W., 2020. Concordance between clinician-rated and patient reported outcome measures of depressive symptoms in treatment resistant depression. J. Affect. Disord. 266, 22–29. https://doi.org/10.1016/j.jad.2020.01.108.

Karatzias, A., Power, K., McGoldrick, T., Brown, K., Buchanan, R., Sharp, D., Swanson, V., 2007. Predicting treatment outcome on three measures for post-traumatic stress disorder. Eur. Arch. Psychiatry Clin. Neurosci. 257, 40–46. https://doi.org/10.1007/s00406-006-0682-2.

Kramer, L.B., Whiteman, S.E., Petri, J.M., Spitzer, E.G., Weathers, F.W., 2023. Self-rated versus clinician-rated assessment of posttraumatic stress disorder: an evaluation of discrepancies between the PTSD Checklist for DSM-5 and the clinician-administered PTSD scale for DSM-5. Assessment 30, 1590–1605. https://doi.org/10.1177/10731911221113571.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9. J. Gen. Intern. Med. 16, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Lee, D.J., Weathers, F.W., Thompson-Hollands, J., Sloan, D.M., Marx, B.P., 2022. Concordance in PTSD symptom change between DSM-5 versions of the clinician-administered PTSD scale (CAPS-5) and PTSD checklist (PCL-5). Psychol. Assess. 34, 604–609. https://doi.org/10.1037/pas0001130.

Manea, L., Gilbody, S., McMillan, D., 2015. A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. Gen. Hosp. Psychiatry. 37, 67–75. https://doi.org/10.1016/j.genhosppsych.2014.09.009.

van Minnen, A., Arntz, A., Keijsers, G.P.J., 2002. Prolonged exposure in patients with chronic PTSD: predictors of treatment outcome and dropout. Behav Res Ther 40, 439–457. https://doi.org/10.1016/S0005-7967(01)00024-9.

Na, P.J., Schnurr, P.P., Pietrzak, R.H., 2023. Mental health of U.S. combat veterans by war era: results from the National health and Resilience in veterans study. J. Psych. Res. 158, 36–40. https://doi.org/10.1016/j.jpsychires.2022.12.019.

Peterson, A.L., Blount, T.H., Foa, E.B., Brown, L.A., McLean, C.P., Mintz, J., Schobitz, R. P., DeBeer, B.R., Mignogna, J., Fina, B.A., Evans, W.R., Synett, S., Hall-Clark, B.N., Rentz, T.O., Schrader, C., Yarvis, J.S., Dondanville, K.A., Hansen, H., Jacoby, V.M., Lara-Ruiz, J., Straud, C.L., Hale, W.J., Shah, D., Koch, L.M., Gerwell, K.M., Young-McCaughan, S., Litz, B.T., Meyer, E.C., Blankenship, A.E., Williamson, D.E., Roache, J.D., Javors, M.A., Sharrieff, A.M., Niles, B.L., Keane, T.M., Consortium to Alleviate PTSD, 2023. Massed vs intensive outpatient prolonged exposure for combat-related posttraumatic stress disorder: a randomized clinical trial. JAMA Netw. Open 6. https://doi.org/10.1001/jamanetworkopen.2022.49422 e2249422.

R Core Team, 2021. A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.r-project.org/.

Ragsdale, K.A., Watkins, L.E., Sherrill, A.M., Zwiebach, L., Rothbaum, B.O., 2020. Advances in PTSD treatment delivery: evidence base and future directions for intensive outpatient programs. Curr. Treat Options Psych. 7, 291–300. https://doi.org/10.1007/s40501-020-00219-7.

Rauch, S.A.M., Yasinski, C.W., Post, L.M., Jovanovic, T., Norrholm, S., Sherrill, A.M., Michopoulos, V., Maples-Keller, J.L., Black, K., Zwiebach, L., Dunlop, B.W., Loucks, L., Lannert, B., Stojek, M., Watkins, L., Burton, M., Sprang, K., McSweeney, L., Ragsdale, K., Rothbaum, B.O., 2021. An intensive outpatient program with prolonged exposure for veterans with posttraumatic stress disorder: retention, predictors, and patterns of change. Psychol. Serv. 18, 606–618. https://doi.org/10.1037/ser0000422.

Resick, P.A., LoSavio, S.T., Wachen, J.S., Dillon, K.H., Nason, E.E., Dondanville, K.A., Young-McCaughan, S., Peterson, A.L., Yarvis, J.S., Mintz, J., STRONG STAR Consortium, 2020. Predictors of treatment outcome in group or individual cognitive processing therapy for posttraumatic stress disorder among active duty military. Cogn. Ther. Res. 44, 611–620. https://doi.org/10.1007/s10608-020-10085-5.

Resick, P.A., Straud, C.L., Wachen, J.S., LoSavio, S.T., Peterson, A.L., McGeary, D.D., Young-McCaughan, S., Taylor, D.J., Mintz, J., 2023. STRONG STAR consortium, consortium to alleviate PTSD, 2023. A comparison of the CAPS-5 and PCL-5 to assess PTSD in military and veteran treatment-seeking samples. Eur. J. Psychotraumatol. 14, 2222608. https://doi.org/10.1080/20008066.2023.2222608.

Rizvi, S.L., Vogt, D.S., Resick, P.A., 2009. Cognitive and affective predictors of treatment outcome in cognitive processing therapy and prolonged exposure for posttraumatic stress disorder. Behav. Res. Ther. 47, 737–743. https://doi.org/10.1016/j.brat.2009.06.003.

Tavakol, M., Dennick, R., 2011. Making sense of Cronbach's alpha. Int. J. Med. Educ. 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd.

Thorndike, E.L., 1920. A constant error in psychological ratings. J. Appl. Psychol. 4, 25–29.

Tolin, D.F., Gilliam, C., Wootton, B.M., Bowe, W., Bragdon, L.B., Davis, E., Hannan, S.E., Steinman, S.A., Worden, B., Hallion, L.S., 2018. Psychometric properties of a structured diagnostic interview for DSM-5 anxiety, mood, and obsessive-compulsive and related disorders. Assessment 25, 3–13. https://doi.org/10.1177/1073191116638410.

Weathers, F.W., Bovin, M.J., Lee, D.J., Sloan, D.M., Schnurr, P.P., Kaloupek, D.G., Keane, T.M., Marx, B.P., 2018. The clinician-administered PTSD scale for DSM–5 (CAPS-5): development and initial psychometric evaluation in military veterans. Psychol. Assess. 30, 383–395. https://doi.org/10.1037/pas0000486.