Received: 1 June 2023 | Revised: 2 September 2023 | Accepted: 4 September 2023

DOI: 10.1002/jts.22975

ISTSS International Society for Traumatic Stress Studies WILEY

RESEARCH ARTICLE

Minimal important difference metrics and test–retest reliability of the PTSD Checklist for *DSM*-5 with a primary care sample

Brittany E. Blanchard[1] | **Morgan Johnson**[1] | **Sarah B. Campbell**[1,2] | **David E. Reed II**[3] | **Shiyu Chen**[1] | **Patrick J. Heagerty**[4] | **Brian P. Marx**[5,7] | **Debra Kaysen**[6] | **John C. Fortney**[1,2]

[1]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington, USA

[2]VA Health Services Research and Development Center for Innovation for Veteran-Centered and Value-Driven Care, VA Puget Sound Health Care System, Seattle, Washington, USA

[3]Department of Health Systems and Population Health, University of Washington School of Public Health, Seattle, Washington, USA

[4]Department of Biostatistics, University of Washington School of Public Health, Seattle, Washington, USA

[5]VA Boston Healthcare System, National Center for PTSD, Boston, Massachusetts, USA

[6]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Palo Alto, California, USA

[7]Department of Psychiatry, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, USA

Correspondence
Brittany E. Blanchard PhD, Assistant Professor and Director, Division of Population Health Analytics Team, Department of Psychiatry and Behavioral Sciences, School of Medicine, University of Washington, 1959 NE Pacific Street, Box 356560, Seattle, WA 98195-6560, 206-221-5498.
Email: bblancha@uw.edu

Funding information
Patient-Centered Outcomes Research Institute, Grant/Award Number: PCS-1406-19295; National Center for Advancing Translational Sciences, Grant/Award Number: KL2TR002317

Abstract

The PTSD Checklist for *DSM-5* (PCL-5) is a measure of posttraumatic stress disorder (PTSD) symptom severity that is widely used for clinical and research purposes. Although previous work has examined metrics of minimal important difference (MID) of the PCL-5 in veteran samples, no work has identified PCL-5 MID metrics among adults in primary care in the United States. In this secondary analysis, data were evaluated from primary care patients ($N = 971$) who screened positive for PTSD and participated in a large clinical trial in federally qualified health centers in three U.S. states. Participants primarily self-identified as women (70.2%) and White (70.3%). We calculated test–retest reliability using clinic registry data and multiple distribution- and anchor-based metrics of MID using baseline and follow-up survey data. Test–retest reliability (Pearson's r, Spearman's ρ, intraclass correlation coefficient) ranged from adequate to excellent (.79–.94), with the shortest time lag demonstrating the highest reliability estimate. The MID for the PCL-5 was estimated using multiple approaches. Distribution-based approaches indicated an MID range of 8.5–12.5, and anchor-based approaches indicated an MID range of 9.8–11.7. Taken together, the MID metrics indicate that PCL-5 change scores of 9–12 likely reflect real change in PTSD symptoms and indicate at least an MID for patients, whereas PCL-5 change scores of 5 or less likely are not reliable. These findings can help inform clinicians

using the PCL-5 in similar populations to track patient responses to treatment and help researchers interpret PCL-5 score changes in clinical trials.

In 2020, an estimated 13,000,000 adults in the United States had a diagnosis of posttraumatic stress disorder (PTSD; National Center for PTSD, 2023). The past-year prevalence of PTSD is approximately 12.5% in primary care settings (Spottswood et al., 2017) and is even higher in publicly funded safety-net clinics such as federally qualified health centers (FQHCs), with estimates ranging from 16% to 24% (Han et al., 2016; Lathan et al., 2021). Given the prevalence of PTSD, identifying effective treatments for the disorder is a public health priority. To determine if a PTSD treatment is effective, it is necessary to know whether the observed change in PTSD symptoms is both reliable (i.e., not due to measurement error) and clinically meaningful. There are varying approaches to measuring and defining change considered to be minimally detectable by patients, of clinical importance, and/or statistically reliable (Hays & Peipert, 2019; see King, 2011, and Sedaghat, 2019). The two most commonly used methods are distribution-based methods and anchor-based methods.

Distribution-based methods encompass multiple formulas that use sample statistics to estimate the minimal change score that is not due to chance or measurement error (de Vet et al., 2006; Hays & Peipert, 2021; Jacobson & Truax, 1991; Wright et al., 2012). Distribution-based formulas are often variations of within-person *t* tests that incorporate sample standard deviations (i.e., baseline or change score standard deviations) and standard error of measurement (SEM). Although some researchers note that a limitation of these metrics is their reliance on group-level variances that may not apply to a given individual (Hays & Peipert, 2021; see Wells et al., 2001), this is a common practice for determining whether an individual has improved during and after treatment.

A well-known distribution-based approach is the reliable change index (RCI) formula (Jacobson and Truax, 1991), a change score that reflects change that is not due to measurement error 95% of the time. In the RCI formula, individual-level change is divided by the adjusted SEM. This estimate is compared with 1.96, which represents the critical *z* value for a 95% confidence interval. Later, researchers multiplied the SEM by 1.96 to derive the change score indicative of reliable change, sometimes denoted as RC > 1.96 (Marx et al., 2022). Other similar distribution-based approaches include variations to SEM that account for multiple assessment points multiplied by 1.96. This formula is referred to by many names, including "minimal detectable change" (Beaton et al., 2001; de Vet et al., 2006), "smallest real difference" (Beckerman et al., 2001), "smallest detectable change," and "coefficient of repeatability" (see Hays & Peipert, 2021, for a review). Other distribution-based approaches include multiplying an effect size estimate by the standard deviation and multiplying the baseline standard deviation by 0.5 (Norman et al., 2003; see Sedaghat, 2019, for a review). Thus, there are many formulas but no consensus about which should be applied to group- and/or individual-level change (King et al., 2019; c.f., Wright et al., 2012). In addition to limited generalizability beyond the sample, the biggest criticism of these approaches is their failure to account for the patient-perceived importance of change.

Anchor-based methods address this limitation (see Sedaghat, 2019). The construct these methods assess is also referred to by many names, including "minimal clinically important difference" (MCID; Guyatt et al., 1987; Jaeschke et al.,1989), "minimally important change," "minimally detectable difference," and "subjectively significant difference" (de Vet et al., 2006; Guyatt et al., 2002); in the present study, we refer to this construct as "minimal important difference" (MID). Such methods require data on patient-reported or clinician-rated indicators of change. A common approach is to calculate mean change scores for a given outcome measure among patients who report minimal change on another "anchor" measure. Anchor-based approaches are superior to distribution-based methods in that they incorporate patient perception of change. However, patient perception of change can be impacted by recall and response biases, and there have been few psychometric evaluations of anchors used for the purpose of establishing MIDs.

When anchor-based data are not available or feasible to acquire, distribution-based methods are sometimes used as a proxy. Some researchers caution against conflating these metrics (see Hays & Peipert, 2021; Terwee et al., 2011; Turner et al., 2010), whereas others recommend using multiple anchor- and distribution-based MID metrics and triangulating the findings to identify a plausible MID range (e.g., Sedaghat, 2019). For the present study, we adhered to the latter recommendations.

Research examining any metric of MID for patient-reported PTSD outcome measures remains extremely limited, thus hindering the clinical and research applications of these measures. The PTSD Checklist for *DSM-5* (PCL-5; Weathers et al., 2013) is a measure that quantifies PTSD symptom severity based on criteria outlined in the

Diagnostic and Statistical Manual of Mental Disorders (5th ed.; *DSM-5*; American Psychiatric Association [APA], 2013). The PCL-5 is commonly used in clinical trials to evaluate the effectiveness of PTSD treatments as well as in measurement-based care (MBC) for PTSD, which involves multiple administrations of the PCL-5 over time to monitor treatment response and inform treatment decisions. Previous examinations of the original PCL (Weathers et al., 1993), based on diagnostic criteria in the fourth edition of the *DSM* (*DSM*-IV; APA, 1994), using distribution- and anchor-based methods estimated an MID range of approximately 5–10 points (Stefanovics et al., 2018). To our knowledge, only one study has reported any change score indicative of a MID for the PCL-5 (Marx et al., 2022).

In a sample of White U.S. male veterans, Marx and colleagues (2022) calculated an RCI, denoted as RC > 1.96, for the PCL-5 and reported a PCL-5 MID change score of 15–18 points as indicative of reliable change. Because the calculation of an RCI requires an estimate of a measure's reliability to calculate SEM, Marx et al. (2022) used test–retest estimates from a previous study that also used a veteran sample (Bovin et al., 2016). We are aware of three estimates of test–retest reliability for the English version of the PCL-5, derived from samples of U.S. military veterans ($n = 51$; Bovin et al., 2016), U.S. undergraduate students ($n = 55$; Blevins et al., 2015), and Welsh adults ($n = 51$; Roberts et al., 2021), with test–retest estimates ranging from .82 to .86. However, there are currently no test–retest estimates for the English version of the PCL-5 derived with United States–based clinical adults outside of the Veterans Affairs (VA) health care system (see Forkus et al., 2023).

Broadly, evidence of the psychometric properties of a given measure's scores can only be generalized to similar populations. As stated by Borsboom and Molenaar (2015), "Strictly speaking, such reliability estimates signify properties of test scores rather than of tests themselves (e.g., administering the test to different populations will ordinarily lead to different values for reliability)" (p. 419). Reliability, especially the test–retest estimate, is particularly relevant to MID.

More specific to trauma, the types of traumatic events experienced, prevalence and correlates of PTSD, and prevalence of treatment-seeking vary between veteran and nonveteran communities (see Levahot et al., 2018). Further, there is evidence that veterans may have lower treatment expectancies compared to nonveteran civilians (Gobin et al., 2018). These lower expectancies among veterans may bias responses on patient-reported outcomes (PROs) and anchor measures, which likely impact MID estimates.

Based on several recommendations in the literature (e.g., King, 2011; Revicki et al., 2007; Sedaghat, 2019; Wells et al., 2001), we used multiple distribution- and anchor-based methods to estimate MID in a sample of adult primary care patients. To derive reliable change score estimates using RCI (i.e., RC > 1.96), we first calculated the PCL-5 test–retest reliability in the sample. Like MID, there is a lack of consensus regarding best practices, so we examined test–retest reliability across multiple time lags using three of the most common approaches: Pearson's *r*, Spearman's rho, and the intraclass correlation coefficient (ICC). Our objective was to estimate an MID range and test–retest reliability of the PCL-5. No apriori hypotheses were specified.

METHOD

Participants

Data for this secondary analysis were derived from a randomized pragmatic comparative effectiveness trial conducted across 12 FQHCs (see Fortney et al., 2021). The trial compared two primary care–based approaches for delivering evidence-based treatment to patients with PTSD and/or bipolar disorder. Patients were randomized to receive care from a primary care team or a specialty mental health team. PTSD treatments included psychotropic medication and/or psychotherapy (i.e., behavioral activation or cognitive processing therapy). The results indicated large effect sizes for each arm, as assessed using Veterans 12-Item Short Form Healthy Survey Mental Health Component summary scores (VR-12 MCS; Jones et al., 2021), which served as the primary outcome (i.e., Cohen's *d*s = 0.79 and 0.87), and no statistically significant difference between arms (see Fortney et al., 2021, for details). Of the 1,004 patients who participated in the trial, we restricted the present sample to data from the 978 participants who screened positive for PTSD on the abbreviated six-item PCL (i.e., a score of 14 or higher; Han et al., 2016; Lang & Stein, 2005), with additional inclusion criteria for specific distribution- and anchor-based analyses (*n*s = 69–971).

Procedure

We used baseline and 6-month survey data for demographic information and MID analyses. For test–retest analyses, we used PCL-5 scores from an online clinical registry (see Unützer et al., 2002, 2012) used by providers for MBC. The frequency of PCL-5 administration for MBC reflected both standardized encounter schedules and variation across patients based on need. In one study arm, the PCL-5 was administered via a smartphone app and sent to the clinical registry, which accounted for many

short time lags between PCL-5 administrations (Fortney et al., 2020). Therefore, the shortest time lag between PCL-5 administrations also varied from patient to patient. Written informed consent was obtained for all study participants in the trial. All procedures and materials for the clinical trial were approved by the Institutional Review Boards of the University of Arkansas for Medical Sciences, University of Michigan, and University of Washington. The trial was designed and conducted in close collaboration with consumer and policy advisory boards. The pragmatic trial was preregistered at ClinicalTrials.gov (Identifier:NCT02738944).

Measures

Demographic characteristics

Sociodemographic information was obtained via items from a self-report survey. Age, gender, ethnicity, and race, as well as socioeconomic status, assessed using 2016 federal poverty level data (United States Census Bureau, 2016) and rurality, classified using Rural–Urban Commuting Area (RUCA) Categorization D (Economic Research Service, 2020), were measured at baseline.

PTSD symptoms

The PCL-5 (Weathers et al., 2013) is a 20-item measure used to assess *DSM-5* symptoms of PTSD. Responses are scored on a scale of 0 (*not at all*) to 4 (*extremely*), and scores are summed, with higher total scores indicating greater symptom severity (range: 0–80). PCL-5 change scores (i.e., PCL-$5_{\Delta 0-6}$) were calculated by subtracting 6-month PCL-5 scores from PCL-5 baseline scores; positive change scores indicate decreases in symptom severity (i.e., symptom improvement). PCL-5 scores have demonstrated evidence of convergent and discriminant validity, test–retest reliability (rs = .82–.86; Blevins et al., 2015; Bovin et al., 2016), and internal consistency reliability (Cronbach's $\alpha = .94$; Blevins et al., 2015). In the current sample, the PCL-5 total score demonstrated good internal reliability, with Cronbach's $\alpha = .93$, $\omega = .93$.

Health-related quality of life

The VR-12 (Jones et al., 2001) is a 12-item measure used to assess health-related mental and physical quality of life during the past 4 weeks. For the first anchor-based analysis, we used Item 9, "How much of the time during the past 4 weeks have you felt calm and peaceful?," with response options ranging from 1 (*none of the time*) to 6 (*all of the time*) and lower scores reflecting better functioning. For the second anchor-based analysis, we used the VR-12 Mental Health Component Summary score (MCS), which is normed to the U.S. population ($M = 50$, $SD = 10$; baseline range: -0.35–63.40; 6-month range: -0.81–65.58). MCS change scores ($MCS_{\Delta 6-0}$) were calculated by subtracting baseline MCS scores from 6-month MCS scores. A positive $MCS_{\Delta 6-0}$ indicates improvement in mental health-related quality of life. VR-12 MCS scores have demonstrated evidence of validity as well as good internal consistency reliability (e.g., Cronbach's $\alpha = .90$; Kazis et al., 2001). In the present sample, VR-12 scores demonstrated good internal consistency reliability, Cronbach's $\alpha = .83$, $\omega = .80$.

Perceptions of recovery

The Recovery Assessment Scale (RAS; Corrigan et al., 2005) measures recovery in relation to mental health concerns across five domains. For the third anchor-based analysis, only Item 14 (i.e., "My symptoms interfere less and less with my life") was used, with response options ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) and higher scores reflecting higher levels of recovery orientation.

Data analysis

Overview

We first estimated the test–retest reliability of the PCL-5 using clinical registry data. Next, we estimated MID for the PCL-5 using distribution- and anchor-based metrics with survey data. To explore the magnitude of PTSD symptom severity change (PCL-$5_{\Delta 0-6}$) that was empirically associated with no change or a change less than MID, we also examined mean change scores among patients who reported no change on the VR-12 anchor-based outcomes in the survey data.

Test–retest reliability

To estimate PCL-5 test–retest reliability, we used registry PCL-5 data from study participants who completed at least two PCL-5 assessments within 30 days in the context of MBC. We created four mutually exclusive analytic subsamples with different lag times. To create time lags, we examined the distribution and aimed to create the shortest time lag possible with an adequate sample size ($n \sim$ 50–100; see de Vet et al., 2011; Kennedy, 2022). After the 0–3-day lag ($n = 88$), we created subsets similar to those in previous work based on mean lag. The 4–7-day lag

($n = 152$) and 22–30-day lag ($n = 59$) subsets were created to compare estimates to Blevins et al. (2015; $M_{lag} = 6$ days) and Bovin et al. (2016; $M_{lag} = 28$ days). The 8–21 day-lag subset ($n = 248$) was created with the remaining data. Participants could only contribute one set of PCL-5 test–retest data points. If a participant had three or more PCL-5 administrations within the determined lag, the pair of PCL-5 scores with the shortest lag time was included. Participant demographic data (e.g., age, gender, race) from the baseline research survey were used for descriptive statistics.

Test–retest reliability was assessed using Pearson's correlation coefficients, Spearman's rank correlation coefficients, and ICCs across the four subsamples. ICC estimates and their 95% confidence intervals (CIs) were calculated using a single-rating, consistency, two-way mixed-effects model (Qin et al., 2019). There were no missing data in this analysis. Although the ICC is often recommended, we included correlation coefficients to compare our findings with existing literature. Analyses were conducted using the ICC9 macro (Hankinson et al., 1995; Hertzmark & Spiegelman, 2010) and PROC CORR in SAS (Version 9.4). We also examined the mean change between the first and second PCL-5 clinical administrations across the four time lag subsamples.

Distribution-based methods

The calculation of distribution-based methods requires at least a baseline PCL-5 score ($N = 971$). We used variations of the effect size and SEM-based formulas, including RCI. Test–retest reliability coefficient from our sample with the shortest time lag was used to calculate the SEM. Table3 includes the equations and sample sizes for each distribution-based approach. Given the lack of consensus regarding the use of baseline versus change score–derived standard deviations, we provide equations using both standard deviations in Table 3.

Anchor-based methods

We examined correlations between outcome change scores and proposed anchor-based indicators. Using Revicki et al.'s (2007) guidelines for establishing MID for PROs, anchors were considered adequate if the correlations were at least .30 or larger. We examined the Spearman's rho correlations between PCL-$5_{\Delta 0–6}$ and proposed anchor-based measures (i.e., VR-12 Item 9, VR-12 MCS, and RAS Item 14).

Continuous anchor

MCS change score ($MCS_{\Delta 6–0}$) was used as a patient-reported indicator of improvement in mental health-related quality of life and was coded using baseline and 6-month survey data. To be included in this analysis, patients had to endorse a 5–10-point change in MCS score. A change of 5 points was selected based on previous research indicating this threshold for determining the MID for the MCS, which corresponds to 0.5 of a standard deviation (Norman et al., 2003). We used a range of 5–10 points to ensure an adequate sample size.

Categorical anchors

To examine PCL-$5_{\Delta 0–6}$ for patient-reported indicators of clinical change with categorical response options (i.e., VR-12 Item 9 and RAS Item 14) without known MID ranges, we first examined empirical logits and subsample means for each item at baseline and 6-months. We could not assume equivalent logits across different categorical response options, so we chose changes in response options based on previous operationalizations of MID, which included the phrasing "a little" to increase face validity. Specifically, we compared PCL-5 change scores between patients who reported one response category higher on the Likert scale at follow-up than they did at baseline, as Hays and Peipert (2021) have expressed concerns about the inclusion of all patients who report any change. For VR-12 Item 9 ("felt calm and peaceful"), participants in the analytic sample ($n = 73$) who endorsed both *none of the time* at baseline and *a little of the time* at 6 months were included. For RAS Item 14 ("symptoms interfere less"), the analytic sample ($n = 69$) was limited to participants who endorsed either (a) *disagree* at baseline and *neither agree nor disagree* at 6 months or (b) *neither agree nor disagree* at baseline and *agree* at 6 months. All anchor-based results and sample sizes are reported in Table 4.

Exploratory and sensitivity analyses

To explore PCL-5 change scores not indicative of at least MID, we also examined the mean PCL-$5_{\Delta 0–6}$ among subsamples of participants who reported a change less than the established MID of 5 on the MCS (i.e., change scores ranging from -2.5 to 2.5). For sensitivity analyses, we excluded data from participants with PCL-$5_{\Delta 0–6}$ outliers, defined as scores beyond the 90th percentile ($n = 10–21$), and repeated the anchor-based analyses described previously.

RESULTS

Sample descriptive statistics

Table 1 presents descriptive statistics across all analytic samples. The full sample used to evaluate

TABLE 1 Sample sociodemographic and clinical descriptive statistics

Variable	*M*	*SD*	*n*	%
Age (years)	39.4	12.85		
Gender				
Women			679	70.2
Men			276	28.5
Another gender			12	1.2
Latina/o/e ethnicity			121	12.5
Racial identity				
African American or Black			114	11.9
Arabic or Middle Eastern			2	0.2
Asian American or Pacific Islander			3	0.3
Multiracial			62	6.5
Native American or Alaskan Native			35	3.6
Other			69	7.2
White			676	70.3
Below 2016 Federal Poverty Level			604	66.0
Rural[a]			479	49.4
Veteran			50	5.2
Clinic location				
Arkansas			235	24.2
Michigan			339	34.9
Washington			397	40.9
PCL-5 descriptive statistics				
Baseline survey	48.0	17.7		
6-month survey	36.6	20.1		
Change score ($PCL\text{-}5_{\Delta 0-6}$)	10.8	18.4		

Note: $N = 971$. PCL-5 = Posttraumatic Stress Disorder Checklist for *DSM-5*.
[a]Defined using Rural–Urban Commuting Area (RUCA) Code Categorization D.

distribution-based metrics ($N = 971$) includes all patients who screened positive for PTSD and contributed a baseline PCL-5 score. The mean baseline survey PCL-5 score was 48.0 ($SD = 17.7$), which exceeds every recommended clinical cutoff score derived for the PCL-5 English version among different populations, indicating a probable PTSD diagnosis (range: 22–42 and above; see Forkus et al., 2023, for a review). The mean participant age was 39.4 years ($SD = 12.9$). The sample consisted primarily of people who self-identified as women (70.2%) who were living below the 2016 federal poverty threshold (66.0%). Participants self-identified as African American (11.9%), Arabic or Middle Eastern (0.2%), Asian American or Pacific Islander (0.3%), multiracial (6.5%), Native American or Alaskan Native (3.6%), White (70.3%), or another race (7.2%). Approximately 13% of participants self-identified as Latina/o/e or Hispanic. Demographic characteristics for select test–retest analyses are provided in Supplementary Table S1.

Test–retest reliability

Test–retest coefficients estimated using Pearson's *r*, Spearman's rho, and the ICC were very similar or identical within—but not between—time lags. Evidence of test–retest reliability was excellent in the 0–3-day lag subsample (.94), good in the 4–7- and 8–21-day lag subsamples (.84–.86 and .82, respectively), and adequate-to-good in the 22–30-day lag subsample (.79–.80; Table 2).

Distribution-based results

The distribution-based estimate using the 0.5*standard deviation approach (Norman et al., 2003) was 8.9. When using the PCL-5 change score standard deviation to evaluate MID using the same approach, the estimate was similar (i.e., 9.2). When using the SEM-derived approaches, the estimates ranged from 8.5 to 12.5. All calculations based on

TABLE 2 Test–retest reliability estimates for the PTSD Checklist for *DSM-5*

Variable			Reliability estimate[a]						Days between PCL-5 administration		PCL-5 score change			
Lag time (days)	*n*	%	ICC	95% CI	*r*	95% CI	*ρ*	95% CI	*M*	*SD*	*M*	*SD*	$PCL\text{-}5_B$ SEM[b]	$PCL\text{-}5_{\Delta 0\text{–}6}$ SEM[b]
0–3	88	16.1	.94	[.91, .96]	.94	[.91, .96]	.94	[.91, .96]	1.2	1.0	−1.2	6.9	4.33	4.51
4–7	152	27.8	.86	[.81, .90]	.86	[.81, .90]	.84	[.79, .88]	6.3	0.9	−2.4	10.0	6.62	6.89
8–21	248	45.3	.82	[.77, .86]	.82	[.77, .85]	.82	[.77, .85]	14.4	3.6	−4.2	12.6	7.50	7.81
22–30	59	10.8	.79	[.68, .87]	.80	[.68, .88]	.80	[.67, .87]	26.4	2.4	−3.5	12.1	7.91	8.23
Overall	547	100.0							11.3	7.7	−3.2	11.1		

Note: PTSD = posttraumatic stress disorder; *DSM-5* = *Diagnostic and Statistical Manual of Mental Disorders* (5th ed); ICC = intraclass correlation coefficient; CI = confidence interval; SEM = standard error of measurement.

[a]Pearson's *r* and Spearman's *ρ* were calculated along with the ICC.

[b]SEM was based on reliability estimates of .94, .86, .82, and .80, respectively, for each lag time category.

change score standard deviations yielded slightly higher estimates than those using baseline standard deviations. The smallest estimate was found using 1.96*SEM and the largest was using RCI > 1.96.

Anchor-based results

All three proposed anchor-based measures exhibited a Spearman's rho correlation with $PCL\text{-}5_{\Delta 0\text{–}6}$ greater than 0.3 (i.e., .40 for the MCA, −.31 for VR-12 Item 9, and .37 for RAS Item 14), $p < .001$. The continuous anchor-based $MCS_{\Delta 6\text{–}0}$ (i.e., 5–10-point change in MCS score) approach indicated a mean $PCL\text{-}5_{\Delta 0\text{–}6}$ of 9.8 ($SD = 17.23$), which was robust in sensitivity analyses. Finally, for categorical anchors, the analytic group that included only individuals who reported minimal improvement on VR-12 Item 9 had a mean $PCL\text{-}5_{\Delta 0\text{–}6}$ of 11.7 ($SD = 18.3$), which was slightly higher than in the sensitivity analysis that removed outliers ($PCL\text{-}5_{\Delta 0\text{–}6}$: $M = 9.9$, $SD = 10.1$). Among individuals who reported minimal detectable improvement on RAS Item 14 between baseline and 6 months, the mean $PCL\text{-}5_{\Delta 0\text{–}6}$ was 11.7 ($SD = 15.2$) and slightly smaller in the sensitivity analysis ($PCL\text{-}5_{\Delta 0\text{–}6}$: $M = 10.3$, $SD = 10.3$).

Exploratory analyses to derive PCL-5 change scores not indicative of MID suggested that participants whose scores reflected change that was categorized as less than minimally important or reported no change on the MCS (i.e., scores of -2.5 to 2.5) had a mean $PCL\text{-}5_{\Delta 0\text{–}6}$ score of 5.1 ($SD = 13.3$), which increased in the sensitivity analysis ($PCL\text{-}5_{\Delta 0\text{–}6}$: $M = 7.5$, $SD = 10.0$).

DISCUSSION

In this secondary data analysis, we estimated MID metrics for the PCL-5 among adult, nonveteran primary care patients using multiple distribution- and anchor-based metrics. The results indicated PCL-5 change scores between 9 and 12 are likely to reflect reliable change (i.e., not due to measurement error) and to be indicative of at least minimal detectable patient-reported symptom improvement. We also found evidence of excellent test–retest reliability for the PCL-5. These results can help guide both clinicians and researchers using the PCL-5.

To our knowledge, we are reporting the first estimate of test–retest reliability of the PCL-5 with an adult nonveteran sample of primary care patients. We found evidence of excellent test–retest reliability (.94) in the shortest time lag (0–3 days), consistent with the test–retest reliability estimate Weathers et al. (1993) found when developing the initial version of the PCL (i.e., reflecting *DSM-IV* symptom

TABLE 3 Distribution-Based Estimates of PTSD Checklist for *DSM-5* minimal important difference (MID)

Method	Equation	SD_B ($N = 971$)	$SD_{\Delta 0-6}$ ($n = 680$)
Medium effect size/0.5*SD	.5*SD	8.9	9.2
1.96*SEM	1.96*SEM	8.5	8.8
Coefficient of repeatability/MDC/SRD	$1.96*\sqrt{(2)}$ *SEM = 2.77*SEM	12.0	12.5
RCI > 1.96	$1.96*S_{DIFF}$	12.0	12.5

Note: PTSD = posttraumatic stress disorder; *DSM-5* = *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.); SEM = standard error of measurement; S_{DIFF} = standard error of measurement of the difference score ($\sqrt{2 * SEM^2}$); RCI = reliable change index; MDC = minimal detectable change; SRD = smallest real difference.

TABLE 4 Anchor-based estimates of PTSD Checklist for *DSM-5* minimal important difference

Method	n	PCL-$5_{\Delta 0-6}$
$MCS_{\Delta 6-0}$	122	9.8
VR-12 item $9_{\Delta 0-6}$	73	11.7
RAS item $14_{\Delta 0-6}$	69	11.7

Note: PTSD = posttraumatic stress disorder; *DSM-5* = *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.); VR-12 = Veterans RAND 12–item Health Survey; MCS = VR-12 Mental Health Component Score; RAS = Recovery Assessment Scale.

criteria) with a male veteran sample, (i.e., .96 with a 2–3-day lag). Our retest estimates and 95% confidence intervals were nearly identical across three approaches.

Previous work by Blevins et al. (2015) demonstrated a Pearson's r PCL-5 test–retest estimate of .82 ($M_{lag} = 6$ days) with an undergraduate sample, similar to our 4–7-day lag estimate (.84–.86, $M_{lag} = 6$ days). Bovin et al. (2016) also reported evidence of good test–retest reliability (Pearson's $r = .86$) when veterans completed two PCL-5 administrations within 30 days ($M_{lag} = 28$ days). Likewise, Roberts et al. (2021) reported a similar Pearson's r test–retest reliability estimate of .84 with a 30-day lag ($M_{lag} = 20$ days). Compared with the shortest time lag (0–3 days), longer time lags were associated with lower test–retest reliability, although the estimates still ranged from adequate to good. Bovin et al. (2016) also reported lower test–retest reliability ($r = .82$) for participants who completed two PCL-5s outside of the 30-day time lag ($M_{lag} = 34$ days) than for those who completed the two PCL-5s within 30 days ($r = .86$; $M_{lag} = 28$ days). Overall, despite the sample diversity, estimates of PCL-5 test–retest reliability were similar to those reported in previous studies.

If test–retest estimates are downwardly biased, using SEM-derived distribution-based metrics of MID will artificially increase MID estimates. For example, with the test–retest estimate of .79, MID based on RCI would increase from 12 to 22.5. Using the original PCL test–retest reliability estimate of .96 derived from a veteran sample with a short test–retest lag (Weathers et al., 1993), Marx et al.'s (2022) RCI-based estimates would decrease from 15 and 18 to 7 and 9, respectively.

Poor test–retest reliability, reflecting a high degree of measurement error, can jeopardize the internal validity of research findings and lead to inaccurate assessments of clinical change. When test–retest estimates are underestimated (i.e., downwardly biased), this could halt an important line of research, prevent an effective treatment from being implemented, or lead to the discontinuation of an effective treatment. Conversely, the overestimation of test–retest estimates would yield biased MID estimates, which could lead to expending resources on treatments that lead to insignificant clinical improvement, the continuation of an ineffective treatment regimen, and/or premature discontinuation of treatment. More research is needed to identify optimal time lags for mental health-related PROs.

The distribution-based estimates based on a medium effect size and 95% confidence intervals ranged from 8.8 to 12.0, and anchor-based estimates fell into this range as well (i.e., 9.7–11.7). Taken together, this suggests that PCL-5 change scores between 9 and 12 may be considered reliable and at least of minimal clinical importance based on the sample estimates. PCL-5 change scores less than 5 likely do not reflect reliable or clinically important change. Indeed, our SEM for this sample was 4.33, suggesting that any change at or below this score may be due to measurement error. Although SEM is still sometimes used as a distribution-based MID metric (Wyrwich, Tierney, et al., 1999), most currently accepted distribution-based formulas (e.g., coefficient of repeatability) yield estimates larger than SEM. Further, patients who reported no change in mental health–related quality of life had a mean PCL-5 change score of 5.1–7.5. More work is needed to determine whether score changes between 5 and 9 are reliable and/or minimally important.

Previous researchers have noted concerns about using test–retest reliability estimates with data from people from populations considered "not stable" (see Reeve et al., 2013, p.1895) and recommend using Cronbach's alpha rather than test–retest reliability (Wyrwich, Nienaber, et al., 1999). At baseline, our sample was approximately 3

standard deviations below the national MCS score average ($M = 22.0$, $SD = 10.0$) and above the PCL-5 threshold for probable PTSD ($M = 48.0$; $SD = 17.7$). Nevertheless, our test–retest estimates were excellent with shorter lags, and our estimates with longer lags were comparable to other test–retest estimates observed in less clinically severe samples. Our data suggest that assessing test–retest reliability with a clinical sample does not necessarily "make the PRO measure look unreliable" (Reeve et al., 2013, p. 1895). Although we refrain from making recommendations regarding which type of reliability to use when calculating SEM for MID metrics, we note that the difference was negligible in this sample (i.e., .93 vs. .94 for Cronbach's alpha and test–retest, respectively). However, this would have made a significant difference in Marx et al.'s (2022) findings (i.e., .94–.95 vs. .84 for Cronbach's alpha and test–retest estimate, respectively).

For PTSD trialists, we recommend calculating sample-specific reliability and/or using estimates representative of the sample. In the absence of data that can be used to calculate sample-derived MID estimates, these results can inform power and sample size calculations for clinical trials. Given the nuanced considerations of power calculations, we refrain from offering specific recommendations, though trialists may want to consider our MID range as an indication of reliable and at least partial response.

PCL score change is commonly used by clinicians as part of evidence-based treatment for PTSD to monitor treatment response, determine whether a patient may need additional or fewer sessions, serve as an indicator of treatment problems, (e.g., working on the wrong index traumatic event), and guide clinical case consultation and supervision (Galovski et al., 2012; Hembree et al., 2003; Monson et al., 2018; Sloan et al., 2022). The weekly measurement of PTSD using the PCL-5 is an active part of prolonged exposure, cognitive processing therapy, and written exposure therapy (Resick et al., 2017; Rothbaum et al., 2007; Sloan & Marx, 2019). Similarly, in measurement-based pharmacotherapy, decisions to augment or change medication are often informed by patient responses to measures like the PCL-5.

Clinical guidelines from the National Center for PTSD have not yet been updated for the PCL-5 and currently suggest using PCL for *DSM-IV* indicators of reliable change (i.e., 5–10 points; National Center for PTSD, 2022). Our analyses support the recommendation to interpret PCL-5 change scores less than 5 as likely not reliable. However, some of our analyses indicated scores lower than 8.5 may also not be reliable. Although replication is necessary, our results indicate that a PCL-5 score change of 9–12 points could be considered at least a partial response. Nierenberg and DeCecco (2001) describe a partial response as "minimally improved". Based on these findings, clinicians may want to see a larger change in the PCL-5 before determining whether there is a treatment effect. Depending on the treatment and the time since the start of treatment, this may prompt a decision about whether to continue the current treatment, alter the approach (e.g., modification, augmentation, increased dose), end treatment, or switch to another treatment, which comes with the risk of losing the small gains achieved.

Although several metrics are considered appropriate to establish individual-level MID, caution should be applied. Our view is aligned with Norman et al. (2003), who aptly stated the following regarding the use of 0.5*SD as a distribution-based metric:

> It would be inappropriate for this to be viewed as a fixed benchmark, like the α of 0.05 for statistical significance, but it would not be inappropriate to consider this as an approximate rule of thumb in the absence of more specific information. (p. 590)

We recommend using these PCL-5 change score ranges to facilitate conversations about symptoms with patients rather than as strict guidelines for assessing treatment effectiveness. Further, individuals with more severe symptoms may require higher degrees of change for partial response or MID. Therefore, the patient's report of PTSD symptom change, treatment goals, functional changes, and treatment preferences should guide all clinical decision-making regardless of PCL-5 change scores (Fortney et al., 2017).

The present study represents the first examinations of PCL-5 MID and test–retest reliability with a primarily nonveteran (i.e., 95% or more of participants in subsamples) adult primary care sample in the United States. Other strengths include the use of multiple distribution- and anchor-based methods for calculating MID, multiple methods for calculating test–retest reliability, and multiple time lags. PCL-5 data for test–retest analyses were collected from patients for clinical purposes (i.e., MBC), so our approach may have more ecological validity than research survey–derived test–retest estimates. To our knowledge, the current study included the largest sample size and used the shortest time lag to estimate PCL-5 test–retest reliability compared with all previous studies.

Because this secondary data analysis was completed with data from primary care patients participating in a clinical trial to receive evidence-based treatment, our results may not generalize to other populations. Given that the baseline VR-12 MCS mean score was approximately 3 standard deviations below the general population mean

in the present sample, the sample may be more clinically severe or complex than non-FQHC primary care populations. Limitations of our anchor-based analyses include not using well-established anchor measures and not establishing the importance of change (i.e., overreliance on face validity, use of smallest change measured by response options, and empirical logits). Other limitations include not testing anchors' sensitivity to change and using a longer time lag than recommended, particularly for the VR-12 analyses. See Wang et al. (2022) for anchor-based measure selection and reporting recommendations. Future research should estimate MID using the recommended single-item PRO measure approach, in addition to others, which assesses symptom change on a 5-point response scale with the following response options: *much worse*, *a little worse*, *the same/no change*, *a little better*, and *much better* (see King, 2011, for more recommendations). Response options could be further refined using human-centered methodologies.

Our sample was seeking and/or receiving treatment, so it is possible test–retest estimates from the longer lags are downwardly biased from symptom improvement due to the PCL-5 scores' sensitivity to change. However, our estimates were consistent with previous research that used similar lags in samples of individuals who were not seeking and/or receiving treatment. Alternatively, our 0–3 day-time lag may have been too short and did not reach the 100-participant sample size as recommended by Kennedy (2022).

Unlike other MID studies, we focused exclusively on patients who reported minimal improvement and not minimal deterioration. Although this may be a limitation, research suggests MIDs can differ for improvement versus deterioration (King et al., 2019). Future research should examine whether the PCL-5 MID is similar for patients who report minimal improvement and deterioration. Independent replications are also warranted, especially for anchor-based MID estimates, which were derived with smaller sample sizes.

Future research should prioritize identifying optimal MID metrics and/or metric combinations. Given the debate regarding the use of MID at the between- versus within-person level (e.g., Wells et al., 2001), more work is needed to determine whether MID-derived change scores function equivalently at both levels. One new approach to estimating individual-level MID involves using moderated nonlinear factor analysis (see Morgan-Lopez et al., 2022), which may be more statistically appropriate. Additional work is necessary to determine the impact of baseline symptom severity on metrics of MID and test–retest reliability for the PCL-5. Further psychometric evaluation of the PCL-5 with a nonveteran sample, including factor structure, convergent validity, and criterion validity, is also still needed for evidence of validity for use with this population.

We aimed to examine metrics of MID and test–retest reliability for the PCL-5 with a sample of adult primary care patients in the United States who screened positive for PTSD. Our analyses indicated a PCL-5 change score of 9–12 was reliable and associated with at least minimal patient-reported improvement. We found evidence that PCL-5 change scores of 5.1–7.5 were associated with no change when using the anchor-based approach with the VR-12 MCS. We also found evidence of excellent test–retest reliability of PCL-5 scores ($r = .94$) when examining reliability with a time lag of 0–3 days. Reliability estimates decreased as time lags increased, although the estimates were consistent across retest methods. These findings can help clinicians, researchers, and policymakers interpret the findings of studies reporting PCL-5 outcomes and contribute new information to help establish clinical and research guidelines for PCL-5 use.

OPEN PRACTICES STATEMENT

The pragmatic trial was preregistered with ClinicalTrials.gov (Identifier: NCT02738944).

We did not preregister this secondary analysis. Survey data for these analyses are available via the Patient-Centered Outcomes Research Institute (PCORI) public data repository. Data for these analyses retrieved from the clinical registry are not included in the public data repository with other trial data. Deidentified data used in these analyses are available upon request with a signed data-sharing agreement. Materials and analysis code for this study are available by emailing the corresponding author at bblancha@uw.edu.

AUTHOR NOTE

Brittany E. Blanchard and Morgan Johnson were co–first authors. Morgan Johnson is now at Merck (Upper Gwynedd, Pennsylvania, USA). Shiyu Chen is now at the Fred Hutchinson Cancer Center (Seattle, Washington, USA).

This work was supported by grants from the Patient-Centered Outcomes Research Institute (PCS-1406-19295). Brittany E. Blanchard is supported by University of Washington's Institute of Translational Health KL2 Program through the National Center for Advancing Translational Sciences of the National Institutes of Health (KL2TR002317).

ORCID

Brittany E. Blanchard https://orcid.org/0000-0002-1750-8463
David E. Reed II https://orcid.org/0000-0003-0109-5999
Brian P. Marx https://orcid.org/0000-0001-8988-1796
John C. Fortney https://orcid.org/0000-0001-9409-8894

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Beaton, D. E., Bombardier, C., Katz, J. N., Wright, J. G., Wells, G., Boers, M., Strand, V., & Shea, B. (2001). Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal Clinically Important Difference. *The Journal of Rheumatology*, *28*(2), 400–405.

Beckerman, H., Roebroeck, M. E., Lankhorst, G. J., Becher, J. G., Bezemer, P. D., & Verbeek, A. L. M. (2001). Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research*, *10*(7), 571–578. https://doi.org/10.1023/A:1013138911638

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. *Journal of Traumatic Stress*, *28*(6), 489–498. https://doi.org/10.1002/jts.22059

Borsboom, D., & Molenaar, D. (2015). Psychometrics. In J. D. Wright (Ed)., *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 418–422). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.43079-5

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. *Psychological Assessment*, *28*(11), 1379–1391. https://doi.org/10.1037/pas0000254

Corrigan, P. W., Salzer, M., Ralph, R. O., Sangster, Y., & Keck, L. (2004). Examining the factor structure of the Recovery Assessment Scale. *Schizophrenia Bulletin*, *30*(4), 1035–1041. https://doi.org/10.1093/oxfordjournals.schbul.a007118

de Vet, H. C., Terwee, C. B., Ostelo, R. W., Beckerman, H., Knol, D. L., & Bouter, L. M. (2006). Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes*, *4*, 1–5. https://doi.org/10.1186/1477-7525-4-54

de Vet, H. C. W., Terwee, C., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide*. Cambridge University Press. https://assets.cambridge.org/97805211/18200/frontmatter/9780521118200_frontmatter.pdf

Economic Research Service. (2020). *2010 Rural–Urban Commuting Area (RUCA) Codes*. U.S. Department of Agriculture. https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/

Forkus, S. R., Raudales, A. M., Rafiuddin, H. S., Weiss, N. H., Messman, B. A., & Contractor, A. A. (2023). The Posttraumatic Stress Disorder (PTSD) Checklist for *DSM-5*: A systematic review of existing psychometric evidence. *Clinical Psychology: Science and Practice*, *30*(1), 110–121. https://doi.org/10.1037/cps0000111

Fortney, J. C., Bauer, A. M., Cerimele, J. M., Pyne, J. M., Pfeiffer, P., Heagerty, P. J., … & Unützer, J. (2021). Comparison of teleintegrated care and telereferral care for treating complex psychiatric disorders in primary care: a pragmatic randomized comparative effectiveness trial. *JAMA Psychiatry*, *78*(11), 1189–1199. https://doi.org/10.1001/jamapsychiatry.2021.2318

Fortney, J. C., Heagerty, P. J., Bauer, A. M., Cerimele, J. M., Kaysen, D., Pfeiffer, P. N., … & Unützer, J. (2020). Study to promote innovation in rural integrated telepsychiatry (SPIRIT): Rationale and design of a randomized comparative effectiveness trial of managing complex psychiatric disorders in rural primary care clinics. *Contemporary Clinical Trials*, *90*, 105873. https://doi.org/10.1016/j.cct.2019.105873

Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2017). A tipping point for measurement-based care. *Psychiatric Services*, *68*(2), 179–188. https://doi.org/10.1176/appi.ps.201500439

Galovski, T. E., Blain, L. M., Mott, J. M., Elwood, L., & Houle, T. (2012). Manualized therapy for PTSD: Flexing the structure of cognitive processing therapy. *Journal of Consulting and Clinical Psychology*, *80*(6), 968–981. https://doi.org/10.1037/a0030600

Gobin, R. L., Mackintosh, M. A., Willis, E., Allard, C. B., Kloezeman, K., & Morland, L. A. (2018). Predictors of differential PTSD treatment outcomes between veteran and civilian women after cognitive processing therapy. *Psychological Trauma: Theory, Research, Practice, and Policy*, *10*(2), 173–182. https://doi.org/10.1037/tra0000266

Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases*, *40*(2), 171–178. https://doi.org/10.1016/0021-9681(87)90069-5

Han, B., Wong, E. C., Mao, Z., Meredith, L. S., Cassells, A., & Tobin, J. N. (2016). Validation of a brief PTSD screener for underserved patients in federally qualified health centers. *General Hospital Psychiatry*, *38*, 84–88. https://doi.org/10.1016/j.genhosppsych.2015.07.009

Hankinson, S. E., Manson, J. E., Spiegelman, D., Willett, W. C., Longcope, C., & Speizer, F. E. (1995). Reproducibility of plasma hormone levels in postmenopausal women over a 2–3-year period. *Cancer Epidemiology, Biomarkers and Prevention*, *4*, 649–654.

Hays, R. D., & Peipert, J. D. (2021). Between-group minimally important change versus individual treatment responders. *Quality of Life Research*, *30*(10), 2765–2772. https://doi.org/10.1007/s11136-021-02897-z

Hembree, E. A., Rauch, S. A., & Foa, E. B. (2003). Beyond the manual: The insider's guide to prolonged exposure therapy for PTSD. *Cognitive and Behavioral Practice*, *10*(1), 22–30. https://doi.org/10.1016/S1077-7229(03)80005-6

Hertzmark, E., & Spiegelman, D. (2010). The SAS ICC9 macro. https://ysph.yale.edu/cmips/research/software/icc9_340204_284_47911_v2.pdf

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*(1), 12–19. https://doi.org/10.1037//0022-006x.59.1.12

Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status: ascertaining the minimal clinically important difference.

Controlled Clinical Trials, *10*(4), 407–415. https://doi.org/10.1016/0197-2456(89)90005-6

Jones, D., Kazis, L., Lee, A., Rogers, W., Skinner, K., Cassar, L., Wilson, N., & Hendricks, A. (2001). Health status assessments using the Veterans SF-12 and SF-36: Methods for evaluating outcomes in the Veterans Health Administration. *The Journal of Ambulatory Care Management*, *24*(3), 68–86. https://doi.org/10.1097/00004479-200107000-00011

Kazis, L. E., Selim, A., Rogers, W., Ren, X. S., Lee, A., & Miller, D. R. (2006). Dissemination of methods and results from the Veterans Health Study: Final comments and implications for future monitoring strategies within and outside the Veterans Healthcare System. *The Journal of Ambulatory Care Management*, *29*(4), 310–319. https://doi.org/10.1097/00004479-200610000-00007

Kennedy, I. (2022). Sample size determination in test–retest and Cronbach alpha reliability estimates. *British Journal of Contemporary Education*, *2*(1), 17–29. https://doi.org/10.52589/BJCE-FY266HK9

King, M. T. (2011). A point of minimal important difference (MID): A critique of terminology and methods. *Expert Review of Pharmacoeconomics & Outcomes Research*, *11*(2), 171–184. https://doi.org/10.1586/erp.11.9

King, M. T., Dueck, A. C., & Revicki, D. A. (2019). Can methods developed for interpreting group-level patient-reported outcome data be applied to individual patient management? *Medical Care*, *57*(Suppl5 1), S38–S45. https://doi.org/10.1097/MLR.0000000000001111

Lang, A. J., & Stein, M. B. (2005). An abbreviated PTSD checklist for use as a screening instrument in primary care. *Behaviour Research and Therapy*, *43*(5), 585–594. https://doi.org/10.1016/j.brat.2004.04.005

Lathan, E. C., Selwyn, C. N., & Langhinrichsen-Rohling, J. (2021). The "3 Es" of trauma-informed care in a federally qualified health center: Traumatic event-and experience-related predictors of physical and mental health effects among female patients. *Journal of Community Psychology*, *49*(2), 703–724. https://doi.org/10.1002/jcop.22488

Lehavot, K., Katon, J. G., Chen, J. A., Fortney, J. C., & Simpson, T. L. (2018). Post-traumatic stress disorder by gender and veteran status. *American Journal of Preventive Medicine*, *54*(1), e1–e9. https://doi.org/10.1016/j.amepre.2017.09.008

Marx, B. P., Lee, D. J., Norman, S. B., Bovin, M. J., Sloan, D. M., Weathers, F. W., Keane, T. M., & Schnurr, P. P. (2022). Reliable and clinically significant change in the Clinician-Administered PTSD Scale for *DSM-5* and PTSD Checklist for *DSM-5* among male veterans. *Psychological Assessment*, *34*(2), 197–203. http://doi.org/10.1037/pas0001098

Monson, C. M., Shields, N., Suvak, M. K., Lane, J. E., Shnaider, P., Landy, M. S., Wagner, A. C., Sijeric, I., Masina, T., Wanklyn, S. G., & Stirman, S. W. (2018). A randomized controlled effectiveness trial of training strategies in cognitive processing therapy for posttraumatic stress disorder: Impact on patient outcomes. *Behaviour Research and Therapy*, *110*, 31–40. https://doi.org/10.1016/j.brat.2018.08.007

Morgan-Lopez, A. A., Saavedra, L. M., Ramirez, D. D., Smith, L. M., & Yaros, A. C. (2022). Adapting the multilevel model for estimation of the reliable change index (RCI) with multiple timepoints and multiple sources of error. *International Journal of Methods in Psychiatric Research*, *31*(2), Article e1906. https://doi.org/10.1002/mpr.1906

National Center for PTSD. (2022). *PTSD Checklist for DSM-5 (PCL-5)*. https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp

National Center for PTSD. (2023). *How common is PTSD in adults?* https://www.ptsd.va.gov/understand/common/common_adults.asp

Nierenberg, A. A., & DeCecco, L. M. (2001). Definitions of antidepressant treatment response, remission, nonresponse, partial response, and other relevant outcomes: A focus on treatment-resistant depression. *Journal of Clinical Psychiatry*, *62*(Suppl1), 5–9.

Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care*, *41*, 582–592. https://doi.org/10.1097/01.MLR.0000062554.74615.4C

Qin, S., Nelson, L., McLeod, L., Eremenco, S., & Coons, S. J. (2019). Assessing test–retest reliability of patient-reported outcome measures using intraclass correlation coefficients: Recommendations for selecting and documenting the analytical formula. *Quality of Life Research*, *28*(4), 1029–1033. https://doi.org/10.1007/s11136-018-2076-0

Reeve, B. B., Wyrwich, K. W., Wu, A. W., Velikova, G., Terwee, C. B., Snyder, C. F., Schwartz, C., Revicki, D. A., Moinpour, C. M., McLeod, L. D., Lyons, J. C., Lenderking, W. R., Hinds, P. S., Hays, R. D., Greenhalgh, J., Gershon, R., Feeny, D., Fayers, P. M., Cella, D., … Butt, Z. (2013). ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Quality of Life Research*, *22*(8), 1889–1905. https://doi.org/10.1007/s11136-012-0344-y

Resick, P. A., Monson, C. M., & Chard, K. M. (2017). *Cognitive processing therapy for PTSD: A comprehensive manual*. Guilford Publications.

Revicki, D. A., Erickson, P. A., Sloan, J. A., Dueck, A., Guess, H., Santanello, N. C., & Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group (2007). Interpreting and reporting results based on patient-reported outcomes. *Value in Health*, *10*(Suppl2), S116–S124. https://doi.org/10.1111/j.1524-4733.2007.00274.x

Roberts, N. P., Kitchiner, N. J., Lewis, C. E., Downes, A. J., & Bisson, J. I. (2021). Psychometric properties of the PTSD Checklist for *DSM-5* in a sample of trauma-exposed mental health service users. *European Journal of Psychotraumatology*, *12*(1), Article 1863578. https://doi.org/10.1080/20008198.2020.1863578

Rothbaum, B., Foa, E., & Hembree, E. (2007). *Reclaiming your life from a traumatic experience: A prolonged exposure treatment program workbook*. Oxford University Press. https://doi.org/10.1093/med-psych/9780190926892.001.0001

Sedaghat, A. R. (2019). Understanding the minimal clinically important difference (MCID) of patient-reported outcome measures. *Otolaryngology–Head and Neck Surgery*, *161*(4), 551–560. https://doi.org/10.1177/0194599819852604

Sloan, D. M., & Marx, B. P. (2019). *Written exposure therapy for PTSD: A brief treatment approach for mental health professionals*. American Psychological Association. https://doi.org/10.1037/0000139-000

Sloan, D. M., Thompson-Hollands, J., Hayes, A. M., Lee, D. J., Alpert, E., & Marx, B. P. (2022). Sudden gains in two trauma-focused treatments for posttraumatic stress disorder. *Behavior Therapy*, *53*(2), 255–266. https://doi.org/10.1016/j.beth.2021.08.003

Spottswood, M., Davydow, D. S., & Huang, H. (2017). The prevalence of posttraumatic stress disorder in primary care: A systematic review. *Harvard Review of Psychiatry*, *25*(4), 159–169. https://doi.org/10.1097/HRP.0000000000000136

Stefanovics, E. A., Rosenheck, R. A., Jones, K. M., Huang, G., & Krystal, J. H. (2018). Minimal clinically important differences (MCID) in assessing outcomes of post-traumatic stress disorder. *Psychiatric Quarterly*, *89*(1), 141–155. https://doi.org/10.1007/s11126-017-9522-y

Terwee, C. B., Terluin, B., Knol, D. L., & de Vet, H. C. W. (2011). Combining clinical relevance and statistical significance for evaluating quality of life changes in the individual patient. *Journal of Clinical Epidemiology*, *64*(12), 1465–1467. https://doi.org/10.1016/j.jclinepi.2011.06.015

Turner, D., Schünemann, H. J., Griffith, L. E., Beaton, D. E., Griffiths, A. M., Critch, J. N., & Guyatt, G. H. (2010). The minimal detectable change cannot reliably replace the minimal important difference. *Journal of Clinical Epidemiology*, *63*(1), 28–36. https://doi.org/10.1016/j.jclinepi.2009.01.024

United States Census Bureau. (2016). *Poverty thresholds: 2016*. https://www.census.gov/data/tables/2016/demo/supplemental-poverty-measure/poverty-thresholds.html

Unützer, J., Chan, Y. F., Hafer, E., Knaster, J., Shields, A., Powers, D., & Veith, R. C. (2012). Quality improvement with pay-for-performance incentives in integrated behavioral health care. *American Journal of Public Health*, *102*(6), e41–e45. https://doi.org/10.2105/AJPH.2011.300555

Unützer, J., Choi, Y., Cook, I. A., & Oishi, S. (2002). Clinical computing: A web-based data management system to improve care for depression in a multicenter clinical trial. *Psychiatric Services*, *53*(6), 671–678. https://doi.org/10.1176/ps.53.6.671

Wang, Y., Devji, T., Qasim, A., Hao, Q., Wong, V., Bhatt, M., Prasad, M., Wang, Y., Noori, A., Xiao, Y., Ghadimi, M., Lozano, L. E. C., Phillips, M. R., Carrasco-Labra, A., King, M., Terluin, B., Terwee, C., Walsh, M., Furukawa, T. A., & Guyatt, G. H. (2022). A systematic survey identified methodological issues in studies estimating anchor-based minimal important differences in patient-reported outcomes. *Journal of Clinical Epidemiology*, *142*, 144–151. https://doi.org/10.1016/j.jclinepi.2021.10.028

Weathers, F., Litz, B., Herman, D., Huska, J., & Keane, T. (1993, October). *The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility* [Conference presentation]. Annual meeting of the International Society for Traumatic Stress Studies, San Antonio, TX, USA.

Weathers, F. W., Litz, B. T., Keane, T. M., Palmier, P. A., Marx, B. P., & Schnurr, P. P. (2013). *The PTSD Checklist for DSM-5 (PCL-5)*. https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp

Wells, G., Beaton, D., Shea, B., Boers, M., Simon, L., Strand, V., Brooks, P., & Tugwell, P. (2001). Minimal clinically important differences: Review of methods. *The Journal of Rheumatology*, *28*(2), 406–412. https://pubmed.ncbi.nlm.nih.gov/11246688/

Wright, A., Hannon, J., Hegedus, E. J., & Kavchak, A. E. (2012). Clinimetrics corner: A closer look at the minimal clinically important difference (MCID). *Journal of Manual & Manipulative Therapy*, *20*(3), 160–166. https://doi.org/10.1179/2042618612Y.0000000001

Wyrwich, K. W., Nienaber, N. A., Tierney, W. M., & Wolinsky, F. D. (1999). Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Medical Care*, *37*(5), 469–478. https://doi.org/10.1097/00005650-199905000-00006

Wyrwich, K. W., Tierney, W. M., & Wolinsky, F. D. (1999). Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *Journal of Clinical Epidemiology*, *52*(9), 861–873. https://doi.org/10.1016/S0895-4356(99)00071-2

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Blanchard, B. E., Johnson, M., Campbell, S. B., Reed, D. E., Chen, S., Heagerty, P. J., Marx, B. P., Kaysen, D., & Fortney, J. C. (2023). Minimal important difference metrics and test–retest reliability of the PTSD Checklist for *DSM*-5 with a primary care sample. *Journal of Traumatic Stress*, *36,* 1102–1114. https://doi.org/10.1002/jts.22975