Received: 26 March 2019 | Revised: 21 May 2019 | Accepted: 10 June 2019

DOI: 10.1002/da.22942

ANXIETY AND DEPRESSION ASSOCIATION OF AMERICA
WILEY

RESEARCH ARTICLE

Developing an optimal short-form of the PTSD Checklist for DSM-5 (PCL-5)

Kelly L. Zuromski[1,2] | Berk Ustun[3] | Irving Hwang[2] | Terence M. Keane[4,5] | Brian P. Marx[4,5] | Murray B. Stein[6,7,8] | Robert J. Ursano[9] | Ronald C. Kessler[2]

[1]Department of Psychology, Harvard University, Cambridge, Massachusetts

[2]Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts

[3]Center for Research on Computation and Society, John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts

[4]National Center for PTSD, Behavioral Science Division, VA Boston Healthcare System, Boston, Massachusetts

[5]Department of Psychiatry, Boston University School of Medicine, Massachusetts

[6]Department of Psychiatry, University of California San Diego, La Jolla, California

[7]Department of Family Medicine and Public Health, University of California San Diego, La Jolla, California

[8]VA San Diego Healthcare System, San Diego, California

[9]Department of Psychiatry, Center for the Study of Traumatic Stress, Uniformed Services University of the Health Sciences, Bethesda, Maryland

Correspondence
Ronald C. Kessler, Department of Health Care Policy, Harvard Medical School, Boston, MA 02115.
Email: Kessler@hcp.med.harvard.edu

Funding information
Military Suicide Research Consortium, Grant/Award Numbers: W81XWH-16-2-0004, W81XWH-16-2-0003; U.S. Department of Defense, Grant/Award Number: HU0001-15-2-0004

Abstract

Background: Although several short-forms of the posttraumatic stress disorder (PTSD) Checklist (PCL) exist, all were developed using heuristic methods. This report presents the results of analyses designed to create an optimal short-form PCL for DSM-5 (PCL-5) using both machine learning and conventional scale development methods.

Methods: The short-form scales were developed using independent datasets collected by the Army Study to Assess Risk and Resilience among Service members. We began by using a training dataset (n = 8,917) to fit short-form scales with between 1 and 8 items using different statistical methods (exploratory factor analysis, stepwise logistic regression, and a new machine learning method to find an optimal integer-scored short-form scale) to predict dichotomous PTSD diagnoses determined using the full PCL-5. A smaller subset of best short-form scales was then evaluated in an independent validation sample (n = 11,728) to select one optimal short-form scale based on multiple operating characteristics (area under curve [AUC], calibration, sensitivity, specificity, net benefit).

Results: Inspection of AUCs in the training sample and replication in the validation sample led to a focus on 4-item integer-scored short-form scales selected with stepwise regression. Brier scores in the validation sample showed that a number of these scales had comparable calibration (0.015–0.032) and AUC (0.984–0.994), but that one had consistently highest net benefit across a plausible range of decision thresholds.

Conclusions: The recommended 4-item integer-scored short-form PCL-5 generates diagnoses that closely parallel those of the full PCL-5, making it well-suited for screening.

KEYWORDS
diagnosis, military personnel, psychological tests/psychometrics, trauma- and stressor-related disorders

1 | INTRODUCTION

Posttraumatic stress disorder (PTSD) is a commonly occurring and seriously impairing disorder (Koenen et al., 2017) with a low treatment rate (Thornicroft et al., 2018). Given that screening is effective in detecting PTSD (Warner, Warner, Appenzeller, & Hoge, 2013), several validated screening scales have been developed for this purpose (Gates et al., 2012; Parker-Guilbert, Moshier, Marx, & Keane, 2018; Wisco, Marx, & Keane, 2012). The PTSD Checklist (PCL; F. Weathers, Litz, Herman, Huska, & Keane, 1993; F. W. Weathers et al., 2013) is one of the most widely used of these scales

[Correction added on 09 August 2019, after first online publication: The "department" for affiliation #4 has been changed]

(Elhai, Gray, Kashdan, & Franklin, 2005; Hoge, Riviere, Wilk, Herrell, & Weathers, 2014). The DSM-5 version of the PCL (PCL-5) assesses each of the 20 DSM-5 (American Psychiatric Association, 2013) criteria B-E symptoms of PTSD and is recommended for screening and monitoring PTSD symptoms throughout treatment in the Department of Veterans Affairs and Department of Defense (VA/DoD; Department of Veterans Affairs, 2017).

Although the PCL-5 has excellent psychometric properties (Blevins, Weathers, Davis, Witte, & Domino, 2015; Bovin et al., 2016; Keane et al., 2014; Wortmann et al., 2016), one weakness is the scale's length (5–10 min completion time; National Center for PTSD), which is problematic given that VA/DoD also recommend screening for many other psychiatric disorders (U.S. Department of Veterans Affairs). To reduce respondent burden, several short-form (2–6 item) versions of the DSM-IV PCL (Bliese et al., 2008; Lang & Stein, 2005) and PCL-5 (Price, Szafranski, van Stolk-Cooke, & Gros, 2016) have been created along with a computer-adaptive version of the PCL-5 (Finkelman et al., 2017, 2018). These short-forms are limited, though, either because they were developed using heuristic methods, or in the case of computer-adaptive testing, cannot be used with paper and pencil administration. Furthermore, research on the comparative performance of the different short-form PCLs is limited (Tiet, Schutte, & Leyva, 2013), creating uncertainty about the optimal number and content of items (Bressler, Erford, & Dean, 2018).

We carried out a secondary analysis of the Army Study to Assess Risk and Resilience among Servicemembers (Army STARRS; Ursano et al., 2014) to develop an optimal short-form PCL-5 using machine learning methods and conventional statistical methods like those used to develop earlier short-forms. Scale development and validation were based on separate subsamples of respondents. The results of these analyses are reported in this paper.

2 | MATERIALS AND METHODS

2.1 | Samples

Army STARRS was a 2009–2015 epidemiological-neurobiological study of risk-protective factors for suicidal behaviors among U.S. Army soldiers (Ursano et al., 2014). We used data from several Army STARRS surveys to create two independent samples for analysis: One in which our models were developed (Training Sample) and the other in which these models were tested (Validation Sample).

We used data from the Army STARRS Pre-Post Deployment Study (PPDS) for model development. The PPDS was a four-wave panel survey of three Brigade Combat Teams initially surveyed before deployment to Afghanistan (T0; October 2011–February 2012; $n = 8{,}558$), then shortly after returning from Afghanistan (T1; September 2012–February 2013), 1–2 months later (T2; October 2012–March 2013), and 9–15 months later (T3; June 2013–May 2014). Because PCL-5 only became available for PPDS T2-T3, these waves were our training sample ($n = 8{,}365$ in T2 and $n = 552$ in T3 but not T2).

The validation sample consisted of respondents to the Army STARRS Longitudinal Survey (LS), an ongoing follow-up study of Army STARRS survey respondents, who were not in PPDS T2-T3 ($n = 11{,}728$; including $n = 6{,}280$ ever-deployed and $n = 5{,}448$ never-deployed). The two Army STARRS surveys in this segment of STARRS-LS included (a) The New Soldier Study (NSS; January 2011–November 2012) of new soldiers interviewed within 48 hr of reporting for Basic Combat Training ($n = 39{,}132$); and (b) the All Army Study (January 2011–March 2013) of active duty soldiers not in basic training nor deployed to a combat theatre ($n = 24{,}894$).

The recruitment and consent procedures for all these surveys, which are discussed in more detail elsewhere (Heeringa et al., 2013; Kessler, Colpe et al., 2013), were approved by the Human Subjects Committees of all Army STARRS collaborating organizations.

2.2 | Measures

2.2.1 | PCL-5

The PCL-5 includes 20 questions to evaluate the presence and severity of the 20 DSM-5 Criteria B-E symptoms of PTSD over the past month (0 = not at all to 4 = extremely). Probable clinical diagnoses of DSM-5 PTSD were assigned based on PCL-5 responses using four PTSD diagnostic thresholds validated against DSM-IV PCL cutoffs in prior work (e.g., Hoge et al., 2014): One threshold based on DSM-5 scoring (i.e., at least one PCL-5 item for Criteria B and C and two for Criteria D and E endorsed at a score of 2 = moderately or higher) and three thresholds based on total PCL scores ≥28, ≥32, and ≥38. We aimed to create short-form PCL-5 scales that would reproduce each of these diagnoses derived from the full PCL-5 using responses to a subset of the 20 questions.

2.2.2 | Psychopathological correlates

We evaluated the convergent and discriminant validity of our short-form measure compared to the full PCL-5 by comparing their associations with known correlates that have been examined in prior psychometric work on the PCL-5 (Bovin et al., 2016) in the validation sample. The correlates considered were measures of DSM-IV major depressive episode, generalized anxiety disorder, panic disorder, and intermittent explosive disorder in the 30 days before the survey based on the self-administered version of the Composite International Diagnostic Interview Screening Scales (CIDI-SC; Kessler, Calabrese et al., 2013). Good concordance exists between CIDI-SC diagnoses and diagnoses based on blinded clinical reappraisal interviews with the Structured Clinical Interviews for DSM-IV (Kessler, Santiago et al., 2013). Suicide ideation in the 30 days before the LS1 survey was assessed with a modified version of the Columbia Suicidal Severity Rating Scale (Posner et al., 2011) that asked about lifetime history of active (i.e., "Did you ever in your life have thoughts of killing yourself?") and passive (i.e., "Did you ever wish you were dead or would go to sleep and never wake up?") ideation and recency in the 30 days before the survey to create a single dichotomous variable of presence/absence of recent suicide ideation.

2.2.3 | Sociodemographic correlates

We also compared associations of diagnoses based on our final short-form PCL-5 and full PCL-5 with several socio-demographic variables, including sex, low education (no education beyond high school graduation or GED), junior enlisted rank (E1-E4), and history of multiple combat deployments (2 vs. 0–1), all assessed with administrative records, and self-reported minority status (Non-Hispanic Black or Hispanic).

2.3 | Analysis methods

We created short-form PCL-5 scales using five statistical methods: Three methods that aimed to produce the same integer scoring system as the full PCL-5 (which can be scored without a computer or a calculator) and two methods that used weighted scoring.

The first integer-scored method used Risk-calibrated Supersparse Linear Integer Model (RiskSLIM; Ustun & Rudin, 2017), which is a machine learning algorithm to efficiently find the best-fitting logistic regression model that has small integer weights and obeys custom constraints. RiskSLIM optimized prediction of dichotomized PTSD diagnostic outcomes in the full PCL-5 (see Measures) from responses to between one and eight PCL-5 questions. Similar to prior work (Ustun & Rudin, 2017; Ustun et al., 2017), each model was required to obey constraints so that it would use a fixed number of questions (1–8) and produce a positive integer-valued score that was monotonic across response levels. One possible RiskSLIM integer scoring of the 0–4 PCL-5 response categories is 0,1,1,1,1. This is equivalent to dichotomous yes–no scoring, as in the Primary Care PTSD Screen for DSM-5 (PC-PTSD-5), a short screening scale often used in VA settings rather than a short-form PCL-5 (Prins et al., 2016).

In addition to RiskSLIM, we used two other statistical methods, each generating both integer-scored and weighted short-form scales. The first was forward stepwise logistic regression to select between one and eight items to predict the same dichotomous PTSD diagnostic outcomes as in RiskSLIM. We summed the 0–4 responses to the selected items to create integer-scored versions and created the weighted versions by multiplying the regression coefficients by the 0–4 responses, summing, and transforming the logit to create predicted probabilities of the diagnostic outcome. The second statistical method was to select between one and eight items based on strength of loadings in a unidimensional exploratory factor analysis of all PCL-5 questions. Integer-scored and weighted versions were created as in the stepwise scales by summing the 0–4 response scores (integer-scored) and estimating logistic regression equations to generate weighted versions with logit-transformed predicted probabilities.

We considered 160 short-form scales ($5 \times 8 \times 4$): Each scale was built using one of the five statistical methods, included between 1 and 8 PCL items, and was designed to predict each of the four dichotomous diagnostic outcomes defined by the full PCL-5. In particular, we considered the area under the receiver operating characteristic (ROC) curve (AUC), which reflects the probability that a randomly selected case on the dichotomous diagnostic outcome will have a higher short-form score than a randomly selected noncase.

We used inspection of the AUCs across models to narrow the range of short-form scales in the validation sample (Cortez & Mohri, 2004). We then evaluated the operating characteristics of each remaining scale using the following standard calibration and performance metrics:

1. Brier score: The mean-squared difference between predicted probabilities of case designations and observed designations based on the full PCL-5 to assess calibration,
2. Sensitivity (SN): The proportion of respondents defined as cases by the full PCL-5 that are classified correctly at being cases on the short-form scale,
3. Specificity (SP): The proportion of respondents defined as noncases by the full PCL-5 that are classified correctly as being noncases on the short-form scale),
4. Positive predictive value (PPV): The proportion of respondents at or above a given screening threshold on the short-form scale that are defined as cases by the full PCL-5,
5. Net benefit (NB): The number of true positives at or above the screening threshold minus the discounted number of false positives at or above the threshold, where the discount rate is defined as PPV/(1−PPV) at the threshold for each logically possible threshold on each scale.

Although seldom included in evaluations of screening scales, NB provides more intuitive and clinically useful information than SN, SP, and PPV in comparing scales because it accounts for between-clinician variation in the relative valuations of correctly detecting a true positive and correctly excluding a true negative (Van Calster et al., 2018). NB is typically evaluated through decision curves (Vickers & Elkin, 2006), which plot the minimum PPV the clinician would require to designate a patient as screening positive (*x*-axis), and the NB of the screening scale at that threshold (*y*-axis). Comparing decision curves for different screening scales shows the range of PPV over which each scale is optimal and the magnitude of this benefit.

The validation sample data were weighted when we calculated short-form scale operating characteristics to adjust for the over-sampling in LS1 of respondents who reported mental disorders or suicidality in their baseline survey.

3 | RESULTS

3.1 | Sociodemographic distribution of the samples

The unweighted sociodemographic distributions in the training sample and validation sample (including ever-deployed and never-deployed subsamples) were 6.3–24.3% female, 69.0–83.1% with no education beyond high school, 23.6–28.4% non-Hispanic Black or Hispanic, and 34.9–82.3% junior enlisted rank (Table 1). The much higher proportion of respondents with junior enlisted rank in the never-deployed validation sample (82.3%) than other samples (34.9–50.1%) reflects the high proportion of validation sample

TABLE 1 Unweighted sociodemographic and Army career characteristic distribution in the training and validation samples

	Training sample		Validation sample					
			Total		Ever deployed		Never deployed	
	%	(*SE*)	%	(*SE*)	%	(*SE*)	%	(*SE*)
Female	6.3	(0.3)	18.0	(0.4)	12.6	(0.4)	24.3	(0.6)
Low education (no college)	83.1	(0.4)	74.2	(0.4)	69.0	(0.6)	80.3	(0.5)
Minority status (non-Hispanic Black/Hispanic)	23.6	(0.4)	26.0	(0.4)	23.9	(0.5)	28.4	(0.6)
Junior enlisted rank (E1-E4)	50.1	(0.5)	56.9	(0.5)	34.9	(0.6)	82.3	(0.5)
History of multiple combat deployments	49.5	(0.5)	27.4	(0.4)	51.2	(0.6)	0.0	(0.0)
(n)	(8,917)		(11,728)		(6,280)		(5,448)	

Note: The training sample consisted of all T2 and T3 respondents to the Army STARRS Pre-Post Deployment Study. The validation sample consisted of all participants in the STARRS Longitudinal Study T1 survey who were not in the training sample. See the text for more detail on the samples.
Abbreviations: *SE*, standard error; STARRS, Study to Assess Risk and Resilience among Servicemembers.

respondents from the NSS, virtually none of whom (other than the few who were in another branch of service before their recent Army enlistment) previously deployed. Roughly half of the training and ever-deployed validation samples (49.5–51.2%) had a history of multiple combat deployments.

3.2 | Thirty-day prevalence estimates of DSM-5 PTSD based on the full PCL-5

Unweighted 30-day prevalence estimates of DSM-5 PTSD, determined by applying the aforementioned four diagnostic thresholds to the full PCL-5, were consistently highest in the ever-deployed validation sample (12.8–17.9%), lowest in the training sample (5.2–9.2%), and intermediate in the never-deployed validation sample (7.8–11.6%; Table 2). Prevalence estimates within sample were consistently highest using the liberal PCL-5 ≥28 scoring rule (9.2–17.9%), lowest using the conservative ≥38 scoring rule (5.2–12.8%), and intermediate using the ≥32 (7.0–15.9%) and DSM-5 Criteria B-E (6.2–15.6%) scoring rules.

3.3 | The PCL-5 items selected for the short-form scales

Given that integer-scored and weighted versions of the short-form scales have the same items, we considered a total of 96 ($8 \times 3 \times 4$) different short-form item sets: Each contained between one and eight items, created using one of three different statistical methods to select the subset of item (RiskSLIM, stepwise regression, factor analysis), and used to predict one of four different dichotomous PTSD outcomes.

Inspection of items in each set shows that those based on based on factor analysis were different from those based on RiskSLIM and stepwise regression (see Tables S1–S4). For example, the RiskSLIM and stepwise sets for the scales with six items (the minimum number required to determine PTSD diagnostic status based on DSM-5 diagnostic rules) included an average of two items from Criterion B (intrusive symptoms, compared to one required in DSM-5), one from Criterion C (avoidance, compared with at least one required in DSM-5), two from Criterion D (negative alterations in cognition and mood, compared with at least two required in DSM-5), and one from Criterion E (alterations in arousal and reactivity, compared with at least two required in DSM-5). In contrast, the factor analysis set included four symptoms from Criterion B, one symptom each from Criteria C and D, and none from Criterion E. These differences occurred because RiskSLIM and stepwise regression both select items to optimize explained variance in the outcomes, leading to selection of minimally redundant items, whereas factor analysis optimizes part-whole associations among the items, leading to selection of items with maximum redundancy.

The implications of these differences can be seen by inspecting AUCs in the training sample (Figure 1a–d). Four observations are noteworthy. First, short-form scales built using RiskSLIM and stepwise

TABLE 2 Thirty-day DSM-5 PTSD prevalence estimates based on responses to the full PCL-5 using four diagnostic thresholds in the unweighted training and validation samples

	DSM-5 Criteria B-E		PCL-5 ≥ 28+		PCL-5 ≥ 32+		PCL-5 ≥ 38+		
	%	(*SE*)	%	(*SE*)	%	(*SE*)	%	(*SE*)	
Training sample	6.2	(0.3)	9.2	(0.3)	7.0	(0.3)	5.2	(0.2)	(8,917)
Validation sample									
Total	13.1	(0.3)	15.0	(0.3)	13.2	(0.3)	10.5	(0.3)	(11,728)
Ever deployed	15.6	(0.5)	17.9	(0.5)	15.9	(0.5)	12.8	(0.4)	(6,280)
Never deployed	10.1	(0.4)	11.6	(0.4)	9.9	(0.4)	7.8	(0.4)	(5,448)

Abbreviations: PCL-5, DSM-5 version of the PCL; PTSD, posttraumatic stress disorder; SE, standard error.

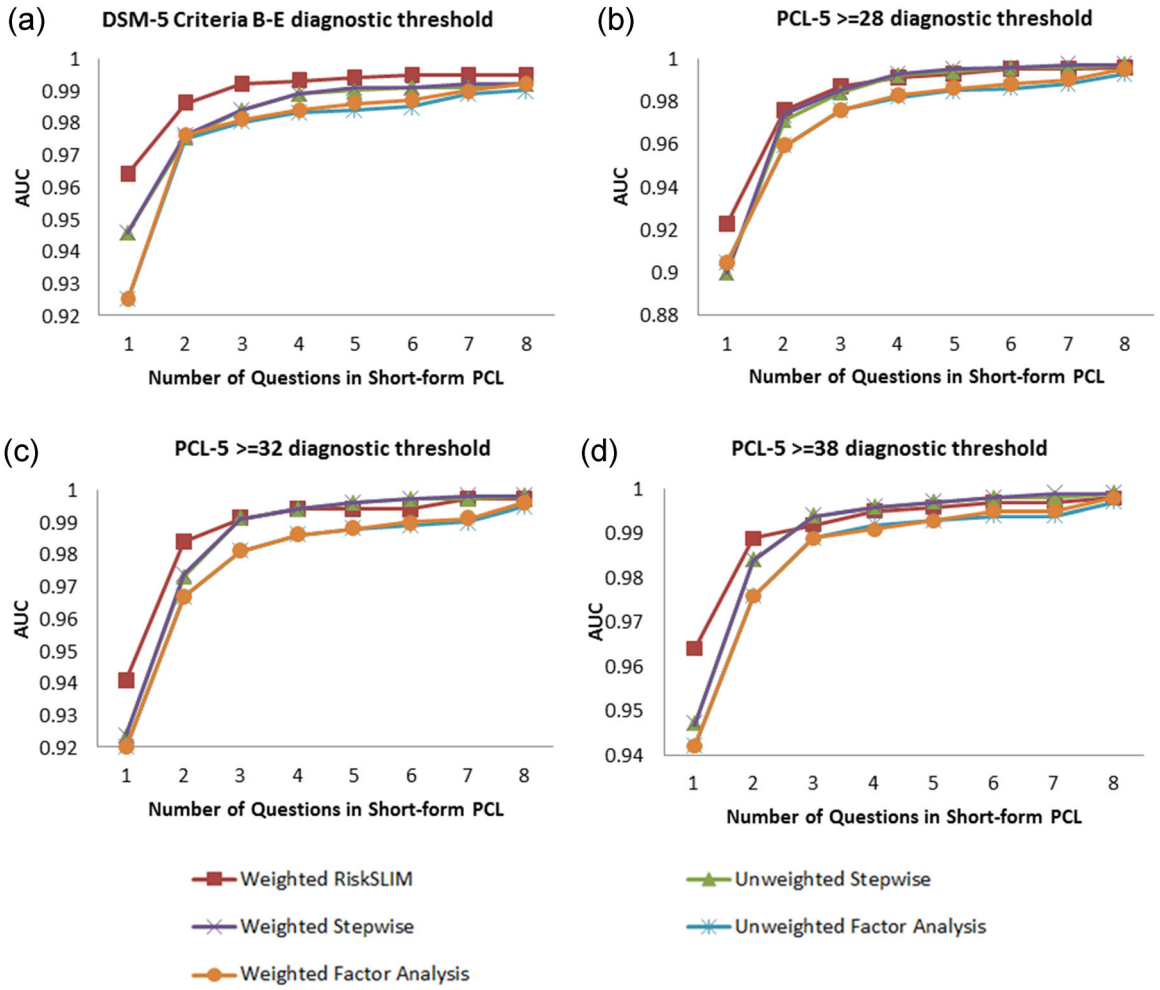

FIGURE 1 (a–d) AUCs of all short-form PCL-5 scales in the training sample (n = 8,917). Short-form PCL scales have between one and eight items and were created using three statistical methods (RiskSLIM, stepwise regression, and factor analysis), including both weighted and unweighted versions of stepwise and factor analysis. Each scale was used to predict dichotomous PTSD outcomes (shown in each panel, a–d), which were assigned based on full PCL-5 responses using four diagnostic thresholds (DSM-5 Criteria B-E, and total PCL-5 scores ≥28, ≥32, and ≥38). PCL-5, DSM-5 version of the PCL; PTSD, posttraumatic stress disorder; RiskSLIM, Risk-calibrated Supersparse Linear Integer Model

regression consistently outperformed those built via factor analysis. Second, although the AUCs continued to rise as number of questions increased, the marginal gain in performance of including a question became negligible after four questions, given that the AUC either approached or exceeded 0.99 for all scales predicting all diagnostic outcomes. Third, although we would expect scales built with weighted stepwise regression to outperform those built with unweighted stepwise regression (as the weights capture differences in relative importance of questions), the two methods yielded similar values of AUC (differences only in the third decimal place; see Table S5). Fourth, although we would expect performance of scales based on RiskSLIM to be better than performance of unweighted stepwise regression because the optimal integer scoring in RiskSLIM allows question-specific nonlinearities to be detected, these differences were small. The latter two observations tell us that optimal weights were similar across questions and that the original PCL linear scoring assumption was consistent with optimal scoring across response categories.

3.4 | Validation of short-form PCL-5 scales based on unweighted stepwise regression

3.4.1 | Narrowing the focus to four-item short-form scales

On the basis of the aforementioned results, we focused further analysis on the integer-scored short-form PCL-5 scales built with stepwise regression. We considered scales with between four and six items given that the incremental benefit of including more than six items was minimal. We expanded the analysis to consider 144 associations: Each of 12 integer-scored short-form scales (4- to 6-item scales selected to predict four different dichotomous PCL-5 diagnostic outcomes in the training sample) with the same outcomes in the validation sample and subsamples. AUCs of all 12 scales either approached or exceeded 0.99 predicting all outcomes in the validation sample and subsamples (see Table S6). We consequently focused subsequent analyses on the 4-item scales.

3.4.2 | Operating characteristics at clinically useful screening thresholds

Brier scores of all 4-item scales were consistently low in the total validation sample (0.019–0.028) and subsamples (0.015–0.032), indicating good calibration of all scales (see Table S7). Inspection of ROC curves was of little help in distinguishing among the different 4-item scales, as none was consistently higher than the others (see Supplemental Figures S1a-d, S2a-2d, S3a-3d) and all had excellent performance. For example, when SP was fixed at 0.9, SN was consistently greater than 0.9 in predicting each outcome.

Stronger discrimination between 4-item scales was found when examining NB. We focused on PPV in the range 0.25–0.75, although we examined the full range of PPV, based on the assumptions that: (a) Clinicians would not want to carry out further evaluations with more than three false positives for every one true positive (PPV = 0.25), noting that the vast majority of true positives would be screened in across scales and samples at that level of PPV (SN = 0.92–0.98) and (b) clinicians would not want to require more than three true positives for every one false positive (PPV = 0.75), noting that such a stringent rule would miss 20–30% of true cases across scales and samples. The decision curves in the total sample (Figures 2a–d) showed that the 4-item short-form scale designed to optimize prediction of the most liberal outcome (i.e., PCL-5 ≥ 28) had marginally higher NB than the other 4-item scales when PPV was in the specified range for three of the four diagnostic outcomes and equivalent to the other 4-item scales for the other outcome (the DSM-5 Criteria B-E outcome). This pattern was more pronounced in the never-deployed subsample (Figure S4a–d), whereas all 4-item short-form scales had equivalent NB in the 0.25–0.75 PPV range in the ever-deployed subsample (Figure S5a–d). On the basis of these results, we selected the 4-item short-form scale designed to optimize prediction of the most liberal outcome (i.e., PCL-5 ≥28) as our recommended scale (Appendix Table 1). We note that even outside this PPV range (<0.25 and >0.75), this pattern of results remains the same.

3.4.3 | Characteristics of the optimal short-form scale

The optimal 4-item short-form PCL-5 scale includes one item assessing each DSM-5 Criteria B-E: B3 (suddenly feeling or acting as if the stressful experience were actually happening again), C2 (avoidance of external reminders of the stressful experience), D6 (distant or cutoff from other people), and E1 (irritable or aggressive behavior). We do not recommend a single diagnostic threshold for this 0–16 integer-scored scale, as the appropriate threshold will depend on whether the user wants to use a conservative (PCL-5 ≥ 38), liberal (PCL-5 ≥ 28), or intermediate (PCL-5 ≥ 32 or DSM-5 Criteria B-E) definition of PTSD as well as the relative value to the user of correctly detecting true positives versus correctly excluding true negatives. However, full information in online supplemental materials (Tables S8–S10) allows users to select the appropriate threshold based on these considerations.

3.4.4 | Comparing correlates of diagnoses based on full PCL-5 and short-form scales

We compared sociodemographic and psychopathological correlates of PTSD diagnoses based on our recommended 4-item short-form PCL-5 scale with those of diagnoses based on the full PCL-5 in the validation sample (Table 3). Thresholds in the short-form scale were selected to make prevalence estimates equivalent to those using the full PCL-5. Odds ratios of correlates with the two diagnoses were very similar for all correlates across all diagnostic scoring systems.

4 | DISCUSSION

In this study, we sought to develop a short-form of the PCL with two goals in mind: (a) Building a clinically useful brief PTSD screener to reduce respondent burden and (b) improving upon statistical methods used to create such a screener, given existing short-form PCLs were created using heuristic methods. To do so, we investigated empirically which PCL-5 items should be used in an optimal short-form version of the scale. Comparing several statistical methods, we found that regression-based short-form PCL-5 scales outperform factor analysis-based short-form scales but that the advantages of weighting (either unrestricted with logistic regression or restricted integer-score weighting with RiskSLIM) are minimal. The latter result indicates that the optimal logistic regression weights are very similar across PCL-5 questions and that the 0–4 scoring assumption is consistent with optimal scoring. One implication of the latter finding is that 0–4 scoring is superior to the 0–1 scoring used in the PC-PTSD-5. We also found that performance does not improve meaningfully with the addition of more than four items, leading us to recommend a 4-item short-form scale. This short-form PCL generates diagnoses that closely parallel those of the full PCL-5 and demonstrates similar psychometric properties (e.g., convergent and discriminant validity), making it well-suited for screening.

It is important to note that this study is not an attempt to ascertain which symptoms do or do not belong in the PTSD diagnostic criteria. Our results should not be interpreted as speaking to this question. Given the very strong associations among DSM-5 Criteria B-E symptoms of PTSD and the strong psychometric properties of the PCL-5, numerous 4-item short-form PCL-5 scales could be created that have operating characteristics close to those of our recommended short-form scale. The four items in our recommended scale are somewhat better than these others, though, in being the minimally redundant set of the 20 PCL-5 items distinguishing cases from noncases according to previously identified PCL-5 PTSD diagnostic thresholds (Hoge et al., 2014). This differs from the content-driven item selection methods used in other PTSD screeners (e.g., the PC-PTSD-5; Prins et al., 2016). As in any stepwise regression scheme, the optimal items included on our short-form should be interpreted broadly as capturing the variance due to all scale items with which they are correlated rather than representing unique effects of specific symptoms. Like other PTSD screeners (i.e., PC-PTSD-5 and the 4-item PCL-5 developed by Price et al., 2016), however, our final short-form includes items assessing for at least one symptom from each DSM-5 PTSD criterion, though the individual items are mostly

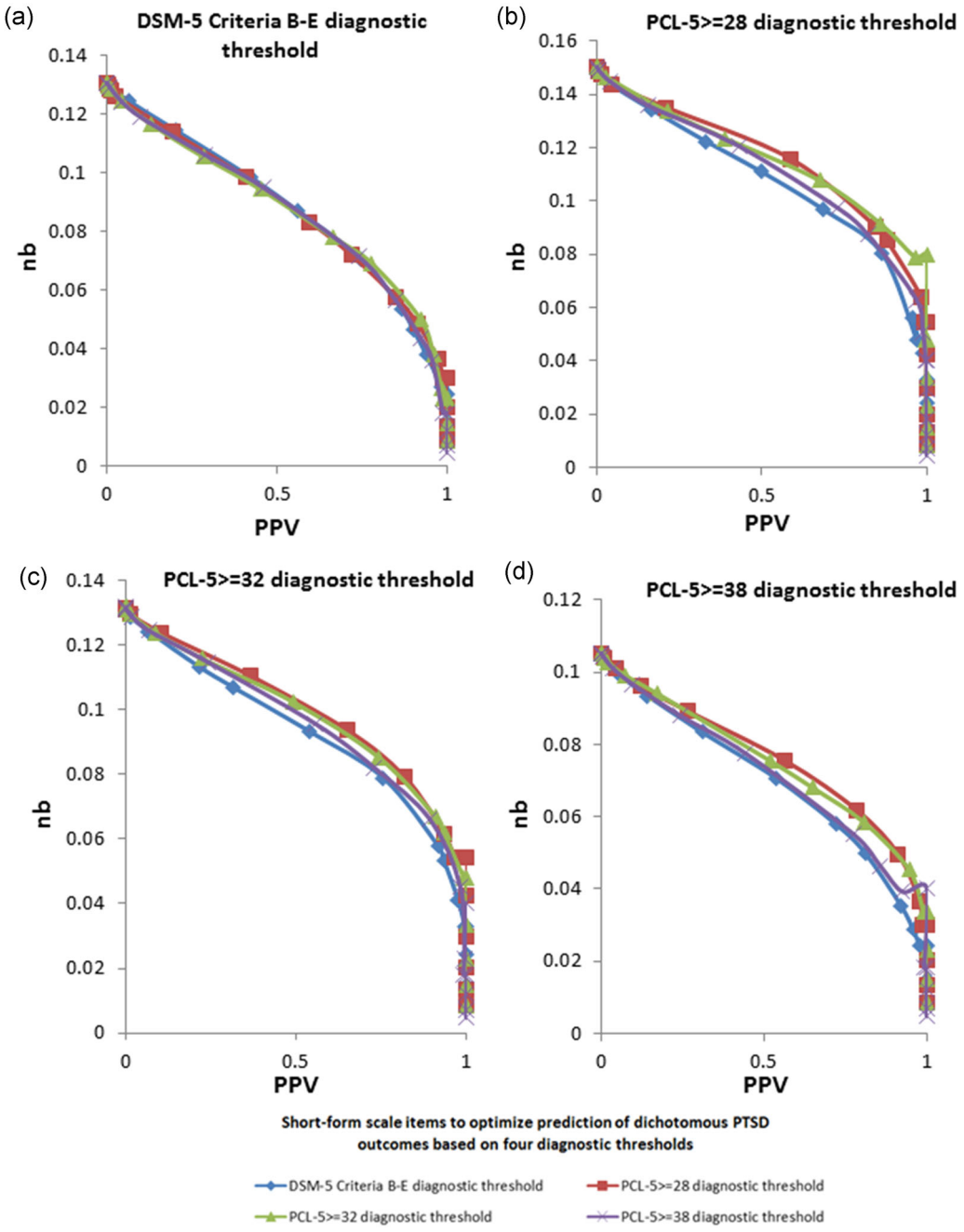

FIGURE 2 (a–d) Decision curves for all unweighted stepwise 4-item short-form scales predicting PTSD outcomes in the total validation sample (n = 11,728). The items included on each 4-item scale were optimized in the training sample to predict dichotomous PTSD outcomes assigned based on full PCL-5 responses using four diagnostic thresholds (DSM-5 Criteria B-E, and total PCL-5 scores ≥28, ≥32, and ≥38). We then used each of these 4-item unweighted stepwise short-form scales to predict these same dichotomous PTSD outcomes in the validation sample, as shown in each panel of the figure (a–d). PCL-5, DSM-5 version of the PCL; PTSD, posttraumatic stress disorder

different (e.g., only one overlapping item between our 4-item short-form and Price et al.'s).

Screening scales should not be used to render clinical diagnoses (McDonald & Calhoun, 2010) but rather to focus attention on individuals most likely to warrant clinical evaluation. As shown in the supplemental materials, our recommended 4-item short-form scale would be well-suited to screen for PTSD in contexts where administration of the full PCL is not possible. At a threshold of 5+, for example, the scale would detect virtually all cases defined by the full PCL-5 as meeting DSM-5 criteria (SN = 0.976) while screening in only a small proportion of PCL-5 noncases (1−SP = 0.066). At a threshold of 6+, the scale would detect an even higher proportion of cases using the conservative PCL-5 ≥ 38 threshold (SN = 0.982) with an even lower false positive rate (1−SP = 0.059).

Our findings should be interpreted in the context of several limitations. First, although our samples were large, they consisted

TABLE 3 Bivariate associations of socio-demographics, Army career variables, and 30-day psychopathology with DSM-5 PTSD based on the final 4-item short-form PCL-5 scale[a] and the full PCL-5, assigned using four diagnostic thresholds in the unweighted validation sample (n = 11,728)[b]

	PTSD diagnostic threshold															
	DSM-5 Criteria B-E				PCL-5 ≥ 28				PCL-5 ≥ 32				PCL-5 ≥ 38			
	Short-form		Full PCL-5		Short-form		Full PCL-5		Short-form		Full PCL-5		Short-form		Full PCL-5	
	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)
I. Sociodemographics and Army Career																
Female	1.5[c]	(1.4–1.8)	1.5[c]	(1.3–1.7)	1.5[c]	(1.3–1.7)	1.5[c]	(1.3–1.7)	1.5[c]	(1.3–1.7)	1.5[c]	(1.3–1.7)	1.5[c]	(1.3–1.7)	1.5[c]	(1.3–1.8)
Low education	1.7[c]	(1.5–1.9)	1.8[c]	(1.5–2.0)	1.7[c]	(1.5–1.9)	1.6[c]	(1.5–1.9)	1.7[c]	(1.5–1.9)	1.7[c]	(1.5–2.0)	1.6[c]	(1.4–1.8)	1.6[c]	(1.4–1.9)
Low family income	1.5[c]	(1.3–1.7)	1.4[c]	(1.2–1.5)	1.4[c]	(1.3–1.6)	1.3[c]	(1.2–1.5)	1.4[c]	(1.3–1.6)	1.4[c]	(1.2–1.5)	1.5[c]	(1.4–1.7)	1.5[c]	(1.3–1.7)
Minority status	1.3[c]	(1.2–1.5)	1.2[c]	(1.1–1.4)	1.3[c]	(1.1–1.4)	1.2[c]	(1.1–1.3)	1.3[c]	(1.1–1.4)	1.3[c]	(1.1–1.4)	1.3[c]	(1.2–1.5)	1.3[c]	(1.1–1.5)
Junior enlisted rank (E1-E4)	1.2[c]	(1.1–1.3)	1.2[c]	(1.1–1.4)	1.2[c]	(1.0–1.3)	1.2	(1.0–1.3)	1.2	(1.0–1.3)	1.1[c]	(1.0–1.2)	1.2[c]	(1.0–1.3)	1.1	(1.0–1.2)
History of multiple combat deployments	1.3[c]	(1.2–1.5)	1.4[c]	(1.2–1.6)	1.4[c]	(1.3–1.6)	1.4[c]	(1.3–1.6)	1.4[c]	(1.3–1.6)	1.5[c]	(1.3–1.6)	1.4[c]	(1.3–1.6)	1.5[c]	(1.4–1.7)
II. Psychopathology																
Major depressive episode	39.4[c]	(34.3–45.3)	35.3[c]	(30.8–40.5)	35.5[c]	(31.0–40.6)	42.5[c]	(37.0–48.8)	35.5[c]	(31.0–40.6)	42.1[c]	(36.6–48.4)	44.3[c]	(38.2–41.4)	46.4[c]	(39.8–53.9)
Generalized anxiety disorder	41.9[c]	(36.3–48.3)	39.1[c]	(34.0–45.0)	39.7[c]	(34.5–45.7)	45.7[c]	(39.6–52.8)	39.7[c]	(34.5–45.7)	47.7[c]	(41.3–55.2)	48.3[c]	(41.5–56.1)	56.5[c]	(48.4–66.0)
Panic disorder	14.5[c]	(11.7–17.8)	16.1[c]	(13.0–20.0)	16.3[c]	(13.1–20.4)	18.8[c]	(15.0–23.6)	16.3[c]	(13.1–20.4)	18.8[c]	(15.1–23.4)	13.5[c]	(11.0–16.5)	15.3[c]	(12.4–18.8)
Intermittent explosive disorder	10.4[c]	(9.1–11.8)	9.4[c]	(8.3–10.7)	10.5[c]	(9.2–11.8)	10.2[c]	(9.0–11.6)	10.5[c]	(9.2–11.9)	10.1[c]	(8.9–11.5)	11.1[c]	(9.8–12.7)	10.7[c]	(9.3–12.2)
Suicide ideation	9.9[c]	(8.4–11.5)	8.6[c]	(7.4–10.0)	9.8[c]	(8.4–11.4)	10.7[c]	(9.2–12.5)	9.8[c]	(8.4–11.4)	10.9[c]	(9.3–12.7)	10.2[c]	(8.7–11.9)	10.2[c]	(8.7–11.9)

Abbreviations: CI, confidence interval; OR, odds ratio; PCL-5, DSM-5 version of the PCL; PTSD, posttraumatic stress disorder.

[a]The final 4-item short-form PCL scale was optimized to predict PCL ≥ 28. See text for more detail.

[b]PTSD diagnoses using the short-form scale were dichotomized at the threshold with the lowest McNemar test (i.e., with a prevalence estimate closest to the prevalence estimate in the full PCL-5) for each of the four PTSD outcomes.

[c]Significant at the 0.05 level, two-sided test.

entirely of U.S. Army soldiers and recently-separated Veterans. It would be useful to evaluate our recommended short-form scale in other populations, including civilian populations, given that past research has highlighted population-specific variation in PCL operating characteristics (Wilkins, Lang, & Norman, 2011). Such differences may be due to exposure to different traumatic event types between populations (e.g., experience of military-specific traumatic events such as combat) or time since event exposure. These factors may affect likelihood of experiencing a given PTSD symptom, which may necessitate development of additional short-form PCL-5 scales that are population-specific. Second, we did not evaluate the test-retest reliability of our recommended scale. This would be useful given the use of short-form scales for symptom tracking as part of measurement-based care (Fortney et al., 2017). Third, we did not have access to clinical interviews to validate PTSD diagnoses, instead using probable diagnoses based on the full PCL-5 as the outcomes. Although diagnoses based on the PCL have been shown to correlate highly with diagnoses based on blinded clinical interviews, including the "gold standard" Clinician-Administered PTSD Scale (Keen, Kutter, Niles, & Krinsley, 2008), additional testing of our short-form scale in predicting interview-based PTSD diagnoses would be useful.

5 | CONCLUSIONS

With the increased emphasis on screening for common mental disorders, the development and use of psychometrically sound and efficient screening tools is critical. To this end, we derived short-form PCL-5 scales using several statistical methods and found that the optimal one is a 4-item scale created using stepwise regression. Instead of a single diagnostic threshold, we offer clinicians the opportunity to select cutoffs on this short-form scale based on clinical setting and judgment using the detailed information provided in our Supplemental Materials. Given its brevity and excellent operating characteristics, this short-form PCL-5 could have great utility for case-finding in a variety of settings, particularly where screening time is a concern.

ACKNOWLEDGEMENTS

This study was supported, in part, by the STARRS-LS study. STARRS-LS is sponsored and funded by the Department of Defense (Grant number HU0001–15-2–0004). Dr. Zuromski was additionally supported by the Military Suicide Research Consortium, an effort supported by the Office of the Assistant Secretary of Defense for Health Affairs (Award No. W81XWH-16-2-0003; W81XWH-16-2-0004). The contents are solely the responsibility of the authors and do not necessarily represent the views of the Veteran's Health Administration, Military Suicide Research Consortium, Department of the Army, or Department of Defense. Although a draft of this manuscript was submitted to the Army and NIMH for review and comment prior to before submission, this was with the understanding that comments would be no more than advisory.

CONFLICT OF INTERESTS

In the past 3 years, Dr. Kessler received support for his epidemiological studies from Sanofi Aventis; was a consultant for Johnson & Johnson Wellness and Prevention, Sage Pharmaceuticals, Shire, Takeda; and served on an advisory board for the Johnson & Johnson Services Inc. Lake Nona Life Project. Kessler is a co-owner of DataStat, Inc., a market research firm that carries out healthcare research. Dr. Stein has been a consultant for the past 3 years for Actelion, Alkermes, Aptinyx, Bionomics, Dart Neuroscience, Healthcare Management Technologies, Janssen, Neurocrine Biosciences, Oxeia Biopharmaceuticals, Pfizer, and Resilience Therapeutics. Dr. Stein has stock options in Oxeia Biopharmaceticals. The remaining authors declare that there are no conflict of interests.

DATA ACCESSIBILITY

The Army STARRS datasets, including the NSS, generated and analyzed during the current study are available in the Interuniversity Consortium for Political and Social research (ICPSR) repository at the University of Michigan, https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/35197.

GROUP INFORMATION

The Army STARRS Team consists of Coprincipal Investigators Robert J. Ursano, MD (Uniformed Services University of the Health Sciences) and Murray B. Stein, MD, MPH (University of California San Diego and VA San Diego Healthcare System)

Site Principal Investigators: Steven Heeringa, PhD (University of Michigan), James Wagner, PhD (University of Michigan), and Ronald C. Kessler, PhD (Harvard Medical School)

Army liaison/consultant: Kenneth Cox, MD, MPH (US Army Public Health Center)

Other team members: Pablo A. Aliaga, MA (Uniformed Services University of the Health Sciences); David M. Benedek, MD (Uniformed Services University of the Health Sciences); Laura Campbell-Sills, PhD (University of California San Diego); Carol S. Fullerton, PhD (Uniformed Services University of the Health Sciences); Nancy Gebler, MA (University of Michigan); Robert K. Gifford, PhD (Uniformed Services University of the Health Sciences); Meredith House, BA (University of Michigan); Paul E. Hurwitz, MPH (Uniformed Services University of the Health Sciences); Sonia Jain, PhD (University of California San Diego); Tzu-Cheng Kao, PhD (Uniformed Services University of the Health Sciences); Lisa Lewandowski-Romps, PhD (University of Michigan); Holly H. Mash, PhD (Uniformed Services University of the Health Sciences); James E. McCarroll, PhD, MPH (Uniformed Services University of the Health Sciences); James A. Naifeh, PhD (Uniformed Services University of the Health Sciences); Tsz Hin Hinz Ng, MPH (Uniformed Services University of the Health Sciences); Matthew K. Nock, PhD (Harvard University); Nancy A. Sampson, BA (Harvard Medical School); Patcho Santiago, MD, MPH (Uniformed Services University of the Health Sciences); Gary H. Wynn, MD (Uniformed Services University of the Health Sciences); and Alan M. Zaslavsky, PhD (Harvard Medical School).

ORCID

Kelly L. Zuromski http://orcid.org/0000-0002-4798-898X
Robert J. Ursano http://orcid.org/0000-0002-1861-9173
Ronald C. Kessler http://orcid.org/0000-0003-4831-2305

REFERENCES

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Publishing.

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*, 489–498. https://doi.org/10.1002/jts.22059

Bliese, P. D., Wright, K. M., Adler, A. B., Cabrera, O., Castro, C. A., & Hoge, C. W. (2008). Validating the primary care posttraumatic stress disorder screen and the posttraumatic stress disorder checklist with soldiers returning from combat. *Journal of Consulting and Clinical Psychology*, *76*, 272–281. https://doi.org/10.1037/0022-006x.76.2.272

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*, 1379–1391. https://doi.org/10.1037/pas0000254

Bressler, R., Erford, B. T., & Dean, S. (2018). A systematic review of the Posttraumatic Stress Disorder Checklist (PCL). *Journal of Counseling & Development*, *96*, 167–186. https://doi.org/10.1002/jcad.12190

Van Calster, B., Wynants, L., Verbeek, J. F. M., Verbakel, J. Y., Christodoulou, E., Vickers, A. J., … Steyerberg, E. W. (2018). Reporting and interpreting decision curve analysis: A guide for investigators. *European Urology*, *74*, 796–804. https://doi.org/10.1016/j.eururo.2018.08.038

Cortez, C., & Mohri, M. (2004). AUC optimization vs. error rate minimization. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems 17: Proceedings of the 2004 Conference*. Cambridge, MA: MIT Press.

Department of Veterans Affairs. (2017). VA/DoD clinical practice guideline for the management of posttraumatic stress disorder and acute stress disorder. (Version 3). Retrieved from https://www.healthquality.va.gov/guidelines/MH/ptsd/

Elhai, J. D., Gray, M. J., Kashdan, T. B., & Franklin, C. L. (2005). Which instruments are most commonly used to assess traumatic event exposure and posttraumatic effects?: A survey of traumatic stress professionals. *Journal of Traumatic Stress*, *18*, 541–545. https://doi.org/10.1002/jts.20062

Finkelman, M. D., Lowe, S. R., Kim, W., Gruebner, O., Smits, N., & Galea, S. (2017). Customized computer-based administration of the PCL-5 for the efficient assessment of PTSD: A proof-of-principle study. *Psychological Trauma*, *9*, 379–389. https://doi.org/10.1037/tra0000226

Finkelman, M. D., Lowe, S. R., Kim, W., Gruebner, O., Smits, N., & Galea, S. (2018). Item ordering and computerized classification tests with cluster-based scoring: An investigation of the countdown method. *Psychological Assessment*, *30*, 204–219. https://doi.org/10.1037/pas0000470

Fortney, J. C., Unutzer, 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*, 179–188. https://doi.org/10.1176/appi.ps.201500439

Gates, M. A., Holowka, D. W., Vasterling, J. J., Keane, T. M., Marx, B. P., & Rosen, R. C. (2012). Posttraumatic stress disorder in veterans and military personnel: Epidemiology, screening, and case recognition. *Psychological Services*, *9*, 361–382. https://doi.org/10.1037/a0027649

Heeringa, S. G., Gebler, N., Colpe, L. J., Fullerton, C. S., Hwang, I., Kessler, R. C., … Ursano, R. J. (2013). Field procedures in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *International Journal of Methods in Psychiatric Research*, *22*, 276–287. https://doi.org/10.1002/mpr.1400

Hirschel, M. J., & Schulenberg, S. E. (2010). On the viability of PTSD Checklist (PCL) short form use: Analyses from Mississippi Gulf Coast Hurricane Katrina survivors. *Psychological Assessment*, *22*, 460–464. https://doi.org/10.1037/a0018336

Hoge, C. W., Riviere, L. A., Wilk, J. E., Herrell, R. K., & Weathers, F. W. (2014). The prevalence of post-traumatic stress disorder (PTSD) in US combat soldiers: A head-to-head comparison of DSM-5 versus DSM-IV-TR symptom criteria with the PTSD Checklist. *The Lancet Psychiatry*, *1*, 269–277. https://doi.org/10.1016/s2215-0366(14)70235-4

Keane, T. M., Rubin, A., Lachowicz, M., Brief, D. J., Enggasser, J. L., Roy, M., … Rosenbloom, D. (2014). Temporal stability of DSM-5 posttraumatic stress disorder criteria in a problem-drinking sample. *Psychological Assessment*, *26*, 1138–1145. https://doi.org/10.1037/a0037133

Keen, S. M., Kutter, C. J., Niles, B. L., & Krinsley, K. E. (2008). Psychometric properties of PTSD Checklist in sample of male veterans. *Journal of Rehabilitation Research & Development*, *45*, 465–474. https://doi.org/10.1682/jrrd.2007.09.0138

Kessler, R. C., Calabrese, J. R., Farley, P. A., Gruber, M. J., Jewell, M. A., Katon, W., … Wittchen, H. U. (2013). Composite International Diagnostic Interview Screening Scales for DSM-IV anxiety and mood disorders. *Psychological Medicine*, *43*, 1625–1637. https://doi.org/10.1017/s0033291712002334

Kessler, R. C., Colpe, L. J., Fullerton, C. S., Gebler, N., Naifeh, J. A., Nock, M. K., … Heeringa, S. G. (2013). Design of the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *International Journal of Methods in Psychiatric Research*, *22*, 267–275. https://doi.org/10.1002/mpr.1401

Kessler, R. C., Santiago, P. N., Colpe, L. J., Dempsey, C. L., First, M. B., Heeringa, S. G., … Ursano, R. J. (2013). Clinical reappraisal of the Composite International Diagnostic Interview Screening Scales (CIDI-SC) in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *International Journal of Methods in Psychiatric Research*, *22*, 303–321. https://doi.org/10.1002/mpr.1398

Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., … Kessler, R. C. (2017). Posttraumatic stress disorder in the World Mental Health surveys. *Psychological Medicine*, *47*, 2260–2274. https://doi.org/10.1017/s0033291717000708

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*, 585–594. https://doi.org/10.1016/j.brat.2004.04.005

McDonald, S. D., & Calhoun, P. S. (2010). The diagnostic accuracy of the PTSD Checklist: A critical review. *Clinical Psychology Review*, *30*, 976–987. https://doi.org/10.1016/j.cpr.2010.06.012

National Center for PTSD. Using the PTSD Checklist for DSM-5 (PCL-5). Retrieved from https://www.ptsd.va.gov/professional/assessment/documents/using-PCL5.pdf

Parker-Guilbert, K., Moshier, S. J., Marx, B. P., & Keane, T. M. (2018). Measures of PTSD symptom severity. In C. R. Nemeroff, & C. B. Marmar (Eds.), *Post-traumatic stress disorder*. New York, NY: Oxford University Press.

Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., & Mann, J. J. (2011). The Columbia-Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry*, *168*, 1266–1277. https://doi.org/10.1176/appi.ajp.2011.10111704

Price, M., Szafranski, D. D., van Stolk-Cooke, K., & Gros, D. F. (2016). Investigation of abbreviated 4 and 8 item versions of the PTSD

Checklist 5. *Psychiatry Research*, *239*, 124–130. https://doi.org/10.1016/j.psychres.2016.03.014

Prins, A., Bovin, M. J., Smolenski, D. J., Marx, B. P., Kimerling, R., Jenkins-Guarnieri, M. A., … Tiet, Q. Q. (2016). The primary care PTSD screen for DSM-5 (PC-PTSD-5): Development and evaluation within a veteran primary care sample. *Journal of General Internal Medicine*, *31*, 1206–1211. https://doi.org/10.1007/s11606-016-3703-5

Steele, N. M., Benassi, H. P., Chesney, C. J., Nicholson, C., & Fogarty, G. J. (2014). Evaluating the merits of using brief measures of PTSD or general mental health measures in two-stage PTSD screening. *Military Medicine*, *179*, 1497–1502. https://doi.org/10.7205/milmed-d-14-00183

Thornicroft, G., Evans-Lacko, S., Koenen, K. C., Kovess-Masfety, V., Williams, D. R., & Kessler, R. C. (2018). Patterns of treatment and barriers to care in posttraumatic stress disorder. In E. J. Bromet, E. G. Karam, K. C. Koenen, & D. J. Stein (Eds.), *Trauma and posttraumatic stress disorder: Global perspectives from the WHO World Mental Health surveys* (pp. 137–152). New York: Cambridge University Press.

Tiet, Q. Q., Schutte, K. K., & Leyva, Y. E. (2013). Diagnostic accuracy of brief PTSD screening instruments in military veterans. *Journal of Substance Abuse Treatment*, *45*, 134–142. https://doi.org/10.1016/j.jsat.2013.01.010

U.S. Department of Veterans Affairs. VA/DoD Clinical Practice Guidelines. Retrieved from https://www.healthquality.va.gov/

Ursano, R. J., Colpe, L. J., Heeringa, S. G., Kessler, R. C., Schoenbaum, M., & Stein, M. B. (2014). The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Psychiatry*, *77*, 107–119. https://doi.org/10.1521/psyc.2014.77.2.107

Ustun, B., Adler, L. A., Rudin, C., Faraone, S. V., Spencer, T. J., Berglund, P., … Kessler, R. C. (2017). The World Health Organization adult attention-deficit/hyperactivity disorder self-report screening scale for DSM-5. *JAMA Psychiatry*, *74*, 520–526. https://doi.org/10.1001/jamapsychiatry.2017.0298

Ustun, B., & Rudin, C. (2017). Learning optimized risk scores on large-scale datasets. *axriv.org*, axriv: 1610.00168. Retreived from http://arxiv.org/abs/1610.00168

Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making*, *26*, 565–574. https://doi.org/10.1177/0272989x06295361

Warner, C. H., Warner, C. M., Appenzeller, G. N., & Hoge, C. W. (2013). Identifying and managing posttraumatic stress disorder. *American Family Physician*, *88*, 827–834.

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

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

Wilkins, K. C., Lang, A. J., & Norman, S. B. (2011). Synthesis of the psychometric properties of the PTSD Checklist (PCL) military, civilian, and specific versions. *Depression and Anxiety*, *28*, 596–606. https://doi.org/10.1002/da.20837.

Wisco, B. E., Marx, B. P., & Keane, T. M. (2012). Screening, diagnosis, and treatment of post-traumatic stress disorder. *Military Medicine*, *177*, 7–13.

Wortmann, J. H., Jordan, A. H., Weathers, F. W., Resick, P. A., Dondanville, K. A., Hall-Clark, B., … Litz, B. T. (2016). Psychometric analysis of the PTSD Checklist-5 (PCL-5) among treatment-seeking military service members. *Psychological Assessment*, *28*, 1392–1403. https://doi.org/10.1037/pas0000260

Zhang, Y., & Yang, Y. (2015). Cross-validation for selecting a model selection procedure. *Journal of Econometrics*, *187*, 95–112. https://doi.org/10.1016/j.jeconom.2015.02.006

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zuromski KL, Ustun B, Hwang I, et al. Developing an optimal short-form of the PTSD Checklist for DSM-5 (PCL-5). *Depress Anxiety*. 2019;36:790–800. https://doi.org/10.1002/da.22942

APPENDIX: SHORT-FORM PCL-5[1]

Instructions: Below is a list of problems that people sometimes have in response to a very stressful experience. Please read each problem carefully and then circle one of the numbers to the right to indicate how much you have been bothered by that problem in the past month.

In the past month, how much were you bothered by:

	Not at all	A little bit	Moderately	Quite a bit	Extremely
1. Suddenly feeling or acting as if the stressful experience were actually happening again (as if you were actually back there reliving it)?	0	1	2	3	4
2. Avoiding external reminders of the stressful experience (for example, people, places, conversations, activities, objects, or situations)?	0	1	2	3	4
3. Feeling distant or cut off from other people?	0	1	2	3	4
4. Irritable behavior, angry outbursts, or acting aggressively?	0	1	2	3	4

[1]The final short-form PCL-5 scale was created using unweighted stepwise regression optimized to predict PTSD diagnoses using the PCL-5 ≥ 28 threshold