Published in final edited form as:

Psychiatry Res. 2014 December 15; 220(0): 679–686. doi:10.1016/j.psychres.2014.08.001.

The differential diagnostic accuracy of the PTSD Checklist among men versus women in a community sample

Kelly S. Parker-Guilbert, Feea R. Leifker, Lauren M. Sippel, and Amy D. Marshall*

Department of Psychology, The Pennsylvania State University, University Park, PA 16802

Abstract

We evaluated the specific version of the PTSD Checklist (PCL-S) as a screening tool for the recruitment of community-residing men and women with diverse trauma experiences. We administered the PCL-S via telephone in the context of participant recruitment, as well as in a laboratory setting preceding administration of the Clinician Administered PTSD Scale (CAPS), the gold standard PTSD assessment tool. In the laboratory, the PCL-S performed reasonably well for men and women, yielding overall diagnostic efficiency (ODE) values (representing percentage of cases accurately identified) of 0.78 and 0.73, respectively, for our recommended cut-points of 42 for men and 49 for women. In contrast, as a recruitment tool, the PCL-S yielded an acceptable ODE of 0.79 for men at the recommended cut-point of 47, but only an ODE of 0.56 (representing diagnostic efficiency no greater than chance) for women at the recommended cut-point of 50. A recruitment cut-point of 57 for women yields a similarly modest ODE of 0.61, but with substantial cost to sensitivity. These findings suggest that use of the PCL-S to screen for PTSD among potential study participants may lead to gender biased study results, even when separate diagnostic cut-points for men and women are used.

Keywords

PCL; assessment; gender; trauma; recruitment; screening

1. Introduction

The development of a brief, effective tool to identify individuals with posttraumatic stress disorder (PTSD) is important in multiple ways. In clinical settings, PTSD screening instruments are commonly used to direct patients at elevated risk of having PTSD to further assessment, followed by the provision of resources and treatment, if indicated. In research settings, PTSD screening instruments are often used to identify potential study participants. In both settings, screening instruments are used to identify individuals with probable PTSD (preferably to be later confirmed using a more accurate, but more resource dependent,

© 2014 Elsevier Ireland Ltd. All rights reserved.

*Correspondence should be addressed to Amy D. Marshall, Department of Psychology, 140 Moore Building, Pennsylvania State University, University Park, PA 16802; Tel.: 814-863-1752; Fax: 814-863-7002, AmyMarshall@psu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

diagnostic interview), while also identifying individuals without PTSD to exclude those individuals from follow-up service provision or recruitment into the patient group in research studies. Incorrect identification of individuals with or without PTSD may unnecessarily deplete resources and/or lead to invalid study conclusions.

The PTSD Checklist (PCL; Weathers et al., 1993) is the most widely used PTSD screening instrument. Three versions of the PCL exist, with the specific version (PCL-S) aligning most closely with PTSD diagnostic criteria as it references a specified traumatic event and minimizes reporting of psychological distress that is not trauma-related. The PCL-S has demonstrated diagnostic accuracy in select, somewhat homogenous populations, such as veterans, medical patients, and sexual assault survivors (see McDonald and Calhoun, 2010). Knowing how the PCL-S functions among community-residing men and women unselected for trauma type is important, as such a sample provides more broadly generalizable knowledge.

The existence of gender differences in the prevalence, and potentially the etiology and presentation, of PTSD is becoming a recurring theme in the literature, yet little is known about gender's impact on PTSD screening (Brewin, 2005). Researchers have cautioned that the PCL's psychometric properties may differ across genders (Blanchard et al., 1996) and expressed surprise that little is known about how gender affects the PCL's performance (McDonald and Calhoun, 2010). Few researchers have examined gender differences in the PCL's performance. Freedy and colleagues (2010) examined gender differences in the PCL-C (civilian version, which does not anchor symptoms to a specific traumatic event) and found its performance was superior among men. King and colleagues (2013) found the PCL-M (military version, which anchors symptoms to stressful military experiences) predicted PTSD similarly among men and women. In another study, the PCL-M demonstrated slightly better (though statistically nonsignificant) performance among men than women (Yeager et al., 2007). Other PTSD screening instruments have been found to perform better among men than women (Prins et al., 2003), while additional studies have not revealed significant gender differences (Sheeran and Zimmerman, 2002; Calhoun et al., 2010). The dearth of statistically significant differences may be a function of differential gender differences across military and civilian populations, as well as unequal numbers of men and women within samples. Across most studies that report on this issue, group sizes differ by at least a ratio of 2:1.

The prospect that the PCL-S may be less accurate among women than men is concerning since the lifetime PTSD prevalence rate for women is at least twice that of men (McLean et al., 2011); thus the PCL-S may perform most poorly among the majority of individuals with PTSD. It has been suggested that PTSD manifests differently across the genders, such that it is more frequently characterized by comorbid symptoms of depression and generalized anxiety among women than men (Tolin and Foa, 2006). Additionally, depressive and anxiety disorders are more prevalent among women than men (Kessler, 2003). Therefore, compared to men, women's PCL-S scores may reflect a greater degree of generalized distress, as opposed to PTSD-specific distress (Armour et al., 2011). Indeed, screening instruments are typically less effective when discriminating between those with PTSD and those with similar symptoms (which may more frequently be the case among women) than

when discriminating between those with PTSD and healthy controls. However, few studies of PTSD screening instruments provide information on psychiatric symptoms among those without PTSD (McDonald and Calhoun, 2010), and none have examined this issue as potentially playing a role in the gender-based differential diagnostic accuracy of a PTSD screening instrument.

1.1 Methods for evaluating diagnostic accuracy

Various elements comprise an accurate screening instrument. Such an instrument possesses an optimal combination of sensitivity (i.e., ability to detect true cases) and specificity (i.e., ability to detect false cases). As sensitivity or specificity increases, the other decreases. The overall diagnostic efficiency (ODE) represents the percentage of cases accurately identified by the screening instrument. Related, positive predictive value (PPV) indicates the proportion of individuals who obtain a score above the designated cut-point on the PCL-S who truly have PTSD. Likewise, negative predictive value (NPV) indicates the proportion of individuals who obtain a score below the designated cut-point who truly do not have PTSD. PPV and NPV are naturally influenced by the sample's base rate of PTSD.

Optimal PCL-S cut-points vary by sample and purpose. For example, sensitivity may be more important than specificity in primary care settings where PTSD screening alerts health care professionals to possible PTSD-related difficulties, thus indicating a relatively lower cut-point. In contrast, specificity may be more important in research settings where false positives deplete limited resources without adding to the PTSD knowledge base. Similarly, specificity is especially important when the screening instrument is used without validation by a diagnostic interview, such that inclusion of false positives may yield misleading or invalid study conclusions. Thus, in some research settings, a higher cut-point may be optimal.

A PCL-S cut-point of 44 is generally accepted as indicative of probable PTSD. However, recommendation of this cut-point came from analyses of a small sample of primarily female motor vehicle accident or sexual assault survivors (Blanchard et al., 1996). Later studies suggest optimal PCL cut-points as low as 30 (e.g., Walker et al., 2002) and as high as 60 (Keen et al., 2008). The variability in cut-points is likely a function of study samples being selected based on trauma type, the intended use of the instrument (e.g., patient care, research screening), sample base rates of PTSD, and the gender composition of the sample examined.

1.2 Current study

Three goals existed for the current study. First, we aimed to provide a broadly generalizable estimate of the PCL-S's diagnostic accuracy for the purpose of research participant recruitment, something that has not been offered by prior literature. To do so, we examined the diagnostic accuracy of the PCL-S in a gender-balanced, mixed-trauma community sample in the context of study recruitment. To be comparable to prior studies examining the PCL's diagnostic accuracy outside of the context of participant recruitment, we also examined the PCL-S's diagnostic accuracy among the same sample when administered in a laboratory setting concurrent with the Clinician Administered PTSD Scale (CAPS). Second, we aimed to compare the PCL-S's diagnostic accuracy between men and women in a more

powerful way than prior studies. To do so, we utilized a sample comprised of equal numbers of men and women who were similar in demographic variables. We recruited such a sample by enrolling heterosexual cohabitating couples. Third, we aimed to explore the potential role of non-PTSD psychiatric symptoms in the diagnostic accuracy of the PCL-S among men and women. To do so, we measured general anxiety and depressive symptoms across men and women.

We hypothesized that the PCL-S would appear to be an effective screening instrument when examined across the total sample; however, diagnostic accuracy would be superior for men. We expected that this hypothesized gender difference would be partly due to more severe symptoms of depression and anxiety among women than men. Throughout, we take the perspective of researchers attempting to effectively recruit a community sample of individuals with PTSD. We aim to provide useful guidelines for researchers in a similar position.

2. Methods

2.1 Participants

The current study was conducted as part of a larger project examining the impact of PTSD on relationship functioning. Community participants were recruited using newspaper and internet advertisements (76%), postcards and/or flyers placed in businesses (20%) and an outpatient mental health clinic (4%). Recruitment efforts targeted married or cohabitating heterosexual couples in which at least one partner experienced a stressful life event and who resided in rural or semi-rural communities.

Couples were excluded from the study if neither partner met criteria for probable PTSD (i.e., PCL-S score above 44; $n = 122$ couples), they were no longer interested in the study ($n = 8$), partners' combined income exceeded \$100,000 per year and/or either partner had more than six years post-high school education ($n = 3$), or they ended their relationship ($n = 1$). Income and education restrictions were used to maintain a sample representative of rural communities.

Participants included 128 individuals from 64 couples. On average, participants were 37.06 ($SD = 12.72$) years of age, with an individual monthly income of \$1731.00 ($SD = \1522.00), and 14.3 ($SD = 2.32$) years of education. Most (68.7%) were employed and few (17%) were students. Participants identified their race/ethnicity as Caucasian (85.9%), African-American (6.3%), biracial/multiracial (3.9%), or Hispanic/Latino (3.9%). Most (72.7%) couples were married. All but three participants experienced an event that met DSM-IV Criterion A for PTSD. Participants' index traumas varied greatly; the most frequent traumas included motor vehicle accidents, intimate partner violence, and loved ones experiencing life-threatening events or sudden death.

2.2 Procedures

Couples contacted the lab, then each partner was screened for probable PTSD using the PCL-S (Weathers et al., 1993) over the telephone, as administered in previous studies (e.g., Blanchard et al., 2002). Participants identified their most stressful life event that currently

causes the most distress. Once a Criterion A trauma was identified (including reports of fear, helplessness, or horror), participants verbally completed the PCL-S in reference to that event.

Participants recruited into the study completed a lab session an average of 1.55 months ($SD = 1.58$) after the telephone screening. During the lab session, in addition to procedures not contained in the current report, participants completed a trauma questionnaire immediately before completing a paper version of the PCL-S. Clinical psychology doctoral students then administered a diagnostic interview for PTSD. Participants independently completed questionnaires regarding general psychiatric symptoms at home and mailed them to the lab.

2.3 Instruments

2.3.1. PTSD Checklist-Specific Version (PCL-S)—The PCL-S is a 17-item self-report measure of DSM-IV criteria B, C, and D PTSD symptoms. Respondents indicate how frequently they have been bothered by each PTSD symptom in reference to a specified event during the previous month. Responses are based on a 5-point Likert scale (1 = not at all; 2 = a little bit; 3 = moderately; 4 = quite a bit; 5 = extremely) and scores range from 17-85, with higher scores indicating greater symptom severity. The PCL-S is composed of the DSM-IV PTSD symptom clusters of re-experiencing (items 1-5), avoidance (items 6-12), and hyperarousal (items 13-17). It is most commonly scored using the total cut-point method (utilized in the current study), but may also be scored using the flowchart method, in which the presence of PTSD is determined by endorsement (i.e., > 2) of at least one re-experiencing, three avoidance, and two hyperarousal symptoms. The PCL-S has demonstrated good internal consistency and construct validity (e.g., Blanchard et al., 1996). During telephone administration, $\alpha = 0.91$ (0.94 for men, 0.78 for women). During in-lab administration, $\alpha = 0.93$ (0.94 for men, 0.88 for women).

2.3.2. Traumatic Life Events Questionnaire (TLEQ)—The TLEQ (Kubany et al., 2000) lists 22 types of potentially traumatic events and asks respondents to indicate if they experienced each event, and if so, how many times. The measure also queries fear, helplessness, horror, and what trauma currently causes the most distress. It has demonstrated adequate levels of test-retest reliability and good content validity (Kubany et al., 2000).

2.3.3. Clinician Administered PTSD Scale (CAPS)—The CAPS (Blake et al., 1995) is a semi-structured interview that assesses the frequency and intensity of each DSM-IV PTSD symptom using standard prompt questions and explicit, behaviorally-anchored rating scales that range from zero to four, with higher scores indicating greater frequency or intensity. The CAPS has demonstrated high interrater reliability and convergent validity with other PTSD measures (Weathers et al., 2001). We utilized the original scoring method (Blake et al., 1990) in which a symptom is considered present if the frequency is rated as one or higher and the intensity is rated as two or higher. This scoring method has demonstrated good internal consistency and test-retest reliability (Weathers et al., 1999). In the current sample, $\alpha = 0.94$ (0.95 for men, 0.90 for women). To determine interrater reliability, all recorded interviews were divided by gender and PTSD diagnosis (present or absent), then 10% of all interviews were randomly selected from these categories. The kappa

was 1.0 (100% agreement) for PTSD diagnosis. The intra-class correlation for total PTSD symptoms was 0.92.

2.3.4. Beck Depression Inventory-II (BDI-II)—The BDI-II (Beck et al., 1996) is a 21-item self-report measure of depressive symptoms. Items are rated on a four-point Likert scale, from zero to three (e.g., “I do not feel sad (0)” to “I am so sad or unhappy that I can't stand it (3)”). Total scores range from zero to 63, with higher scores indicating more severe depressive symptoms. In the current sample, $\alpha = 0.92$ (0.91 for men, 0.89 for women).

2.3.5. State-Trait Anxiety Inventory (STAI)—The STAI (Spielberger et al., 1970) is a 40-item self-report measure of anxiety proneness (i.e., trait anxiety) and anxiety in the moment (i.e., state anxiety). The current study utilized only the trait anxiety subscale. Items are rated on a four-point Likert scale from one (*almost never*) to four (*almost always*). In the current sample, $\alpha = 0.94$ (0.93 for men, 0.92 for women).

2.4 Data analysis

Gender differences in demographic characteristics, PCL-S scores, CAPS scores, and PTSD diagnoses were examined using Student's *t*-tests or χ^2 analyses, where appropriate. Due to the small number of individuals with racial minority status, gender differences in race were examined by dichotomous coding (i.e., White vs. non-White).

Complete data for the PCL-S and CAPS were provided by 97% of participants ($n = 124$), on whom analyses were conducted. Three participants did not experience a DSM-IV Criterion A trauma and one participant provided incomplete data during screening.

Husbands' and wives' CAPS scores were not significantly correlated ($r = -0.15$, *ns*), indicating that no meaningful within-couple dependency existed in the data, thus data could be analyzed by individual rather than by couple (Kenny et al., 2006).

The diagnostic accuracies of the telephone screening PCL-S and the in-lab PCL-S among the total sample and separately among men and women were tested against the CAPS using measures of sensitivity, specificity, ODE, PPV, and NPV. We examined these measures along continuous PCL-S cut-points ranging from 30 to 60, representing cut-points previously used in PCL-S validation studies. Given limited information provided by scores ranging from 30 to 39, we present results for scores ranging from 40 to 60. For the purpose of recruiting study participants, we emphasized specificity over sensitivity when identifying optimal cut-points. We also used a nonparametric receiver operating characteristic (ROC) analysis to plot the performance of the PCL-S in predicting a dichotomous outcome on the CAPS (i.e., presence or absence of PTSD) for both the telephone screening PCL-S and the in-lab PCL-S. We used the area under the curve (AUC) statistic to determine how well the PCL-S predicted the presence of PTSD as measured by the CAPS. AUC values range from 0.50 (known as the “line of no information,” such that the PCL's ability to detect PTSD is equal to chance) to 1.0 (where the PCL perfectly detects PTSD). Values ≥ 0.70 are considered acceptable, while values ≥ 0.80 are considered good.

Differences in anxiety and depressive symptoms between men and women were examined using *t*-tests among the full sample. Further, *t*-tests were used to examine differences in levels of anxiety and depression among women who screened as false positives (i.e., women who screened positive per the telephone PCL-S but did not meet diagnostic criteria for PTSD per the CAPS) versus women who screened as true negatives (i.e., women who screened negative for PTSD per the telephone PCL-S and did not meet diagnostic criteria for PTSD per the CAPS).

Effect size *d* was calculated and interpreted according to Cohen's (1988) guidelines (i.e., 0.20 = small, 0.50 = medium, 0.80 = large). All significance tests were two-tailed at $\alpha = 0.05$.

3. Results

Forty-seven participants (37.9%) met diagnostic criteria for current PTSD according to the CAPS, with more women than men meeting criteria for PTSD, $\chi^2(1, 124) = 6.95, p < .01$. PTSD base rates were 49.2% and 26.2% among women and men, respectively. The mean telephone screening PCL-S score across all participants was 44.83 ($SD = 14.44$; range 17-84), and the mean in-lab PCL-S score across all participants was 43.48 ($SD = 15.53$; range 17-77). The mean CAPS score was 37.03 ($SD = 24.30$; range: 0-102) among all participants and 60.28 ($SD = 15.36$; range: 31-102) among participants who met diagnostic criteria for PTSD according to the CAPS. Women had significantly higher telephone screening PCL-S scores ($M = 51.06$; $SD = 10.26$; range: 17-78) than men ($M = 38.39$; $SD = 15.35$; range: 17-84), $t(122) = -5.42, p < .001$. Women also had significantly higher in-lab PCL-S scores ($M = 49.47$; $SD = 13.76$; range: 19-76) than men ($M = 37.51$; $SD = 14.98$; range: 17-77), $t(126) = -4.71, p < .001$. Additionally, women had significantly higher CAPS scores ($M = 46.89$; $SD = 19.59$; range 0-93) than men ($M = 26.85$; $SD = 24.63$; range 0-102), $t(122) = -5.02, p < .001$. CAPS scores were significantly correlated with telephone screening PCL-S total scores ($r = 0.70, p < .001$) and in-lab PCL-S total scores ($r = 0.78, p < .001$).

Men and women did not differ significantly in age, race/ethnicity, student status, years of education, or average monthly income. Women reported more symptoms of anxiety ($t(96) = -4.79, p < .001, d = 0.98$) and depression ($t(111) = -4.87, p < .001, d = 0.92$) than men. Also, among participants who did not meet diagnostic criteria for PTSD according to the CAPS, women reported more symptoms of anxiety ($t(61) = -5.59, p < .001, d = 1.43$) and depression ($t(69) = -4.32, p < .001, d = 1.04$) than men. Further analyses indicated that women who screened as false positives for PTSD reported significantly more symptoms of anxiety ($t(28) = -2.30, p < .05, d = 0.87$) and depression ($t(30) = -2.19, p < .05, d = 0.80$) than women who screened as true negatives.

3.1. Diagnostic Accuracy of the PCL-S

The diagnostic accuracy of the telephone screening PCL-S for the total sample is shown in Table 1. We determined that a cut-point of 47 provided the best balance of sensitivity (0.77) and specificity (0.69). It correctly classified 72% of individuals (i.e., ODE). The PPV was

0.60, and the NPV was 0.83. See Figure 1, Panel A for the ROC curve for the overall sample ($AUC = 0.76$, $p < .001$; 95% CI = 0.67–0.84).

When analyzing the performance of the telephone screening PCL-S by gender (Table 1), the optimal cut-point among men was also 47. At this cut-point, 79% of men were correctly classified (sensitivity = 0.75, specificity = 0.87, PPV = 0.67, NPV = 0.91). See Figure 1, Panel B for the ROC curve for men ($AUC = 0.81$, $p < .001$; 95% CI = 0.68–0.94). Among women, the optimal cut-point was less clear. We decided that a cut-point of 50 provides moderate specificity without considerably compromising sensitivity (sensitivity = 0.52, specificity = 0.59, PPV = 0.55, NPV = 0.56). At this cut-point, 55% of women were correctly classified. Alternatively, if specificity is a substantial concern, a cut-point of 57, which correctly classified 61% of women, should be considered (sensitivity = 0.35, specificity = 0.78, PPV = 0.61, NPV = 0.39). See Figure 1, Panel C for the ROC curve for women ($AUC = 0.60$, $p = .16$; 95% CI = 0.46–0.74).¹

The diagnostic accuracy of the in-lab PCL-S for the total sample is shown in Table 2. We determined that a cut-point of 49 provided the best balance of sensitivity (0.70) and specificity (0.77). It correctly classified 76% of individuals. The PPV was 0.66 and the NPV was 0.82. See Figure 2, Panel A for the ROC curve for the overall sample ($AUC = 0.82$, $p < .001$; 95% CI = 0.75–0.89).

When analyzing the in-lab PCL-S results separated by gender (Table 2), the optimal cut-point for men was 42. At this cut-point, 78% of men were correctly classified (sensitivity = 0.75, specificity = 0.79, PPV = 0.55, NPV = 0.90). See Figure 2, Panel B for the ROC curve for men ($AUC = 0.87$, $p < .001$; 95% CI = 0.78–0.95). Among women, the optimal cut-point was 49. At this cut-point, 73% of women were correctly classified (sensitivity = 0.77, specificity = 0.70, PPV = 0.71, NPV = 0.77). See Figure 2, Panel B for the ROC curve for women ($AUC = 0.75$, $p = .001$; 95% CI = 0.63–0.87).

4. Discussion

The current study results indicate that the utility of the PCL-S for identifying potential PTSD study participants over the telephone is limited. Substantiating previous recommendations (McDonald and Calhoun, 2010), a minimum PCL-S cut-point of 47 appeared to be most appropriate when one's goal is to identify research study participants from a general community. This cut-point provided the most acceptable balance between sensitivity and specificity, though the specificity value was only 0.69, slightly below the minimum recommended value of 0.70. The PPV at this cut-point was also lower than is optimal (i.e., 0.60; recommended values are 0.70). Furthermore, the ROC analysis indicated that the PCL-S provided only acceptable discrimination of PTSD and no-PTSD cases over the telephone. These findings contrast with the in-lab results, which indicate a relatively more acceptable performance of the PCL-S, where the PPV suggests that scoring a 49 or above on the PCL-S is not necessarily indicative of PTSD, and the NPV suggests that

¹We also calculated the diagnostic accuracy of the telephone administered PCL-S using the flowchart scoring method. For men, the flowchart scoring method reduced both sensitivity and specificity compared to the optimal cut-point of 47. For women, this method increased sensitivity, but reduced specificity compared to the optimal cut-point of 50.

a negative diagnosis on the PCL-S at a cut-point of 49 provides a moderately accurate determination that PTSD is unlikely. The ROC curve for the in-lab PCL-S shows good discrimination. Thus, the shortcomings of the PCL-S are especially pronounced when it is used as a recruitment instrument.

The overall modest performance of the telephone screening PCL-S appears to be largely due to the current sample comprising 50% women and the PCL-S's clearly deficient performance among women. The ROC curve for women suggests that diagnostic efficiency was poor and the PCL-S did not perform recognizably better than chance. We believe this is a function of several related issues. First, because women without PTSD, according to the CAPS, exhibited substantially more symptoms of anxiety and depression than men without PTSD, the PCL-S had the challenging task of needing to distinguish PTSD from similar symptoms (i.e., general anxiety and depressive symptoms) more so among women than men. In fact, specificity values were lower for women than men, suggesting that the PCL-S may have captured more diffuse distress in women without PTSD than PTSD-specific symptoms. This conclusion is strengthened by the finding that women who were “false positives” reported significantly more anxiety and depression symptoms than women who were “true negative.” As the PCL-S performed worse for women during recruitment than in-lab, women's generalized distress may interfere more during recruitment; in contrast, the context of a laboratory study of PTSD may help anchor them to trauma-specific distress. It may be helpful to administer a more elaborate trauma questionnaire such as the TLEQ during the recruitment process to increase the likelihood of participants reporting trauma-specific distress.

Second, gender role norms (e.g., women are expected to be susceptible to the consequences of trauma while men are expected to be resilient) may contribute to gender differences in reporting of PTSD symptoms (Tolin and Foa, 2006). That is, men may be generally less likely to report PTSD symptoms, but the semi-anonymous nature of telephone screening may encourage them to openly report symptoms. In contrast, women, who tend to be more affiliative, may be especially likely to openly report PTSD symptoms in a laboratory setting (e.g., Ullman and Filipas, 2005).

Finally, the observed gender difference may be due to differences in the construct validity of PTSD. Armour and colleagues (2011), who found a greater degree of PTSD measurement error among women than men, raised the point that early research on the latent structure of PTSD was conducted with all-male samples such that gender differences may not be taken into account in the diagnostic conceptualization of PTSD. Additionally, PTSD symptoms may be more variable over time among women than men, as highlighted by Andrykowski and colleagues (2000), who found that many female breast cancer survivors exhibited large changes in PCL scores over time. Large symptom changes occurred in both directions, and did not adhere to a consistent pattern, suggesting a potential “reactivation” of PTSD symptoms at various points (e.g., Davidson et al., 1996). The impact of these issues on the diagnostic accuracy of PTSD screening instruments requires further exploration.

Overall, the PCL-S's modest performance may partly be a function of the high level of general life stress in the current sample. Potential study participants responded to

advertisements targeting individuals who experienced stressful events. Additionally, partners of individuals with PTSD often experience considerable psychological distress (e.g., Jordan et al., 1992). Nygaard and colleagues (2011) found that following a trauma, cohabitating individuals showed more similar PTSD symptoms than non-cohabitating individuals, and concluded that one's reactions to trauma may impact his/her partner's PTSD symptoms over time, though partners' PTSD symptoms may be more consistent with generalized distress than PTSD-specific symptoms (Renshaw et al., 2011). The degree of general distress present in a sample is not commonly considered in studies of the PCL-S's diagnostic accuracy. Therefore, when recruiting highly stressed study participants, the PCL-S may perform worse than prior literature suggests, particularly studies of samples with a relatively lower degree of stress (e.g., HMO patients; Walker et al., 2002). However, when attempting to efficiently recruit study participants with PTSD, it is often necessary to target individuals who have experienced stress in order to help individuals without PTSD self-select out. Consequently, the no-PTSD group will exhibit an elevated degree of psychological distress, thus decreasing the screening instrument's diagnostic accuracy.

When deciding which PCL-S cut-point to use for participant recruitment, the importance of sensitivity versus specificity must be determined and use of different cut-points for men and women should be considered. For example, if sensitivity takes priority over specificity, cut-points of 46 or 47 work well for men and women in this context. If specificity is prioritized over sensitivity, a cut-point of 50 (or 57 if specificity is a substantial priority) may be best for women. In contrast, among men, the slight increase in specificity may not be worth the substantial decrease in sensitivity; thus a cut-point of 47 remains best. Additionally, researchers using the PCL-S as a screening device may need to inform potential study participants that they may not be eligible to continue in the study once PTSD diagnostic status is confirmed in the lab. If such expectations are set with potential participants, a more abbreviated screening or lower cut-point may be possible, though further research on such methods would be needed. Without such considerations, the participant composition of the obtained sample (i.e., over-inclusion of women who do not have PTSD) may lead to gender-biased or inaccurate study results.

The results of the current study should be considered in light of its limitations. Most significantly, couples in which neither partner had a PCL-S score greater than 44 were not invited to attend a laboratory session, thus precluding collection of CAPS data and calculation of the PCL-S's diagnostic accuracy among these individuals. Therefore, we lack a representative sample of individuals without a trauma history and those without PTSD, which may deflate specificity values. As a result, our conclusions do not apply to the general population, but are specific to the context of recruiting study participants in a similar manner. However, we do not expect the current results to differ greatly from what would be found among the greater population since the high sensitivity of the PCL-S at a cut-point of 44 indicates that relatively few true positives were missed. If one aims to identify the PCL's diagnostic accuracy in the general community, the CAPS should be administered to all participants regardless of cut-score. Second, we included only individuals in cohabitating relationships, which, in addition to providing a highly stressed sample (as described above), limits generalizability of the study results. A benefit of this design, however, is that men and women had nearly identical demographic characteristics. This methodology also yielded an

equal gender distribution, providing a more valid test of gender differences. Third, participants were recruited strictly from rural or semi-rural communities. Although this work addresses a considerably under-studied population, we do not know if our conclusions generalize to recruitment of study participants from urban populations. Fourth, although the PCL-S has previously been used as a telephone administered recruitment device (e.g., Blanchard et al., 2002), it was not originally designed or tested in this context. Thus, our results may partly be a function of differences in telephone versus in-lab administrations. Fifth, with the publication of the DSM-5, there have been changes to the PTSD symptom structure; however, the current study's use of the DSM-IV criteria likely is still useful for informing gender differences in PTSD screening. Finally, participants completed the CAPS an average of 1.5 months after completing the recruitment screening PCL-S. Participants may have contacted us and completed the screening at a time of elevated PTSD symptoms. We believe the delay between the screening and lab session likely led to random error that decreased the accuracy of the PCL-S overall, but not in a systematic manner across participants. Moreover, this form of error is inherent to the process of recruiting study participants and our delay between recruitment and participation is not unusual for studies recruiting community participants. Thus, we do not believe this issue affects our overall conclusions.

In sum, the current study addressed a number of methodological limitations of the existing literature. Meanwhile, by considering gender-differentiated diagnostic accuracy, we identified a critical issue that needs to be taken into account by future researchers when recruiting participants with PTSD. Our hypothesized reasons for the PCL-S's differential diagnostic accuracy across men and women likely apply to other PTSD screening instruments as used in other contexts. We hope that our perspective on screening for PTSD in a research setting provides helpful guidance for other PTSD researchers and spurs additional research on efficient, accurate identification of PTSD across genders.

Acknowledgments

Dr. Marshall is supported by the National Institutes of Health's Building Interdisciplinary Research Careers in Women's Health (BIRCWH) program (K12 HD055882). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank Fiona Barwick, Kaitlyn Hanley, Lauren Szkodny, and numerous undergraduate research assistants for their helpful contributions to participant recruitment and data collection.

References

- Andrykowski MA, Cordova MJ, Mcgrath PC, Sloan DA, Kenady DA. Stability and change in posttraumatic stress disorder symptoms following breast cancer treatment: A 1-year follow-up. *Psycho-Oncology*. 2000; 9:69–78.10.1002/(SICI)1099-1611(200001/02)9:1 [PubMed: 10668061]
- Armour C, Elhai JD, Layne CM, Shevlin M, Durakovic-Belko E, Djapo N, Pynoos RS. Gender differences in the factor structure of posttraumatic stress disorder symptoms in war-exposed adolescents. *Journal of Anxiety Disorders*. 2011; 25:604–611.10.1016/j.janxdis.2011.01.010 [PubMed: 21377317]
- Beck, AT.; Steer, RA.; Brown, GK. *Manual for the Beck Depression Inventory-II*. Psychological Corporation; San Antonio, TX: 1996.
- Blake, DD.; Weathers, F.; Nagy, L.; Kaloupek, D.; Klauminzer, G.; Charney, D.; Keane, T. *National Center for Post-traumatic Stress Disorder, Behavioral Science Division- Boston VA*. Boston, MA: 1990. *Clinician Administered PTSD Scale (CAPS)*.

- Blake DD, Weathers FW, Nagy LM, Kaloupek DG. The development of a Clinician-Administered PTSD Scale. *Journal of Traumatic Stress*. 1995; 8:75–90.10.1002/jts.2490080106 [PubMed: 7712061]
- Blanchard EB, Hickling EJ, Galovski T, Veazey C. Emergency room vital signs and PTSD in a treatment seeking sample of motor vehicle accident survivors. *Journal of Traumatic Stress*. 2002; 15:199–204.10.1023/A:1015299126858 [PubMed: 12092911]
- Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA. Psychometric properties of the PTSD Checklist (PCL). *Behaviour Research and Therapy*. 1996; 34:669–673.10.1016/0005-7967(96)00033-2 [PubMed: 8870294]
- Brewin CR. Systematic review of screening instruments for adults at risk of PTSD. *Journal of Traumatic Stress*. 2005; 18:53–62.10.1002/jts.20007 [PubMed: 16281196]
- Calhoun PS, McDonald SD, Guerra VS, Eggleston AM, Beckham JC, Straits-Troster K. Clinical utility of the Primary Care-PTSD Screen among U.S. veterans who served since September 11, 2001. *Psychiatry Research*. 2010; 178:330–335.10.1016/j.psychres.2009.11.009 [PubMed: 20483463]
- Cohen, J. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum; Hillsdale, NJ: 1988.
- Davidson, J.; Foa, EB.; Blank, AS.; Brett, EA.; Fairbank, J.; Green, BL.; Herman, JL.; Keane, TM.; Kilpatrick, DL.; March, JS.; McNally, RJ.; Pitman, RK.; Resnick, HS.; Rothbaum, BO. Posttraumatic stress disorder. In: Widiger, TA.; Frances, AJ.; Pincus, HA.; Ross, R.; First, MB.; Davis, WW., editors. *DSM-IV Sourcebook*. Vol. 2. American Psychiatric Press; Washington DC: 1996.
- Freedly JR, Steenkamp MM, Magruder KM, Yeager DE, Zoller JS, Hueston WJ, Carek PJ. Posttraumatic stress disorder screening test performance in civilian primary care. *Family Practice*. 2010; 27:615–624.10.1093/fampra/cmz049 [PubMed: 20622049]
- Jordan BK, Marmar CR, Fairbank JA, Schlenger WE, Kulka RA, Hough RL, Weiss DS. Problems in families of male Vietnam veterans with posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*. 1992; 60(6):916–926.10.1037/0022-006X.60.6.916 [PubMed: 1460153]
- Keen SM, Kutter CJ, Niles BL, Krinsley KE. Psychometric properties of PTSD Checklist in sample of male veterans. *Journal of Rehabilitation Research & Development*. 2008; 45:465–474.10.1682/JRRD.2007.09.0138 [PubMed: 18629754]
- Kenny, DA.; Kashy, DA.; Cook, DL. *Dyadic Data Analysis*. Guilford; New York, NY: 2006.
- Kessler RC. Epidemiology of women and depression. *Journal of Affective Disorders*. 2003; 74:5–13.10.1016/S0165-0327(02)00426-3 [PubMed: 12646294]
- King MW, Street AE, Gradus JL, Vogt DS, Resick PA. Gender differences in posttraumatic stress symptoms among OEF/OIF veterans: An item response theory analysis. *Journal of Traumatic Stress*. 2013; 26:175–183.10.1002/jts.21802 [PubMed: 23526678]
- Kubany ES, Leisen MB, Kaplan AS, Watson SB, Haynes SN, Owens JA. Development and preliminary validation of a brief broad-spectrum measure of trauma exposure: The Traumatic Life Events Questionnaire. *Psychological Assessment*. 2000; 12:210–224.10.1037/1040-3590.12.2.210 [PubMed: 10887767]
- McDonald SD, Calhoun PS. The diagnostic accuracy of the PTSD Checklist: A critical review. *Clinical Psychology Review*. 2010; 30:976–987.10.1016/j.cpr.2010.06.012 [PubMed: 20705376]
- McLean CP, Asnaani A, Litz BT, Hofmann SG. Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity, and burden of illness. *Journal of Psychiatric Research*. 2011; 45:1027–1035.10.1016/j.jpsychires.2011.03.006 [PubMed: 21439576]
- Nygaard E, Wentzel-Larsen T, Hussain A, Heir T. Family structure and Posttraumatic stress reactions: A longitudinal study using multilevel analyses. *BMC Psychiatry*. 2011; 11:195–204.10.1186/1471-244X-11-195 [PubMed: 22171549]
- Prins A, Ouimette P, Kimerling R, Cameron RP, Hugelshofer DS, Shaw-Hegwer J, Sheikh JI. The primary care PTSD screen (PC-PTSD): Development and operating characteristics. *Primary Care Psychiatry*. 2003; 9:8–14.10.1185/135525703125002380
- Renshaw KD, Allen ES, Rhoades GK, Blais RK, Markman HJ, Stanley SM. Distress in spouses of service members with symptoms of combat-related PTSD: Secondary traumatic stress or general

- psychological distress? *Journal of Family Psychology*. 2011; 25:461–469.10.1037/a0023994 [PubMed: 21639635]
- Sheeran T, Zimmerman M. Screening for posttraumatic stress disorder in a general psychiatric outpatient setting. *Journal of Consulting and Clinical Psychology*. 2002; 70:961–966.10.1037/0022-006X.70.4.961 [PubMed: 12182279]
- Spielberger, CD.; Gorsuch, RL.; Lushene, RE. *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press; Palo Alto, CA: 1970.
- Tolin DF, Foa EB. Sex differences in trauma and posttraumatic stress disorder: A quantitative review of 25 years of research. *Psychological Bulletin*. 2006; 132:959–992.10.1037/1942-9681.S.1.37 [PubMed: 17073529]
- Ullman SE, Filipas HH. Gender differences in social reactions to abuse disclosures, post-abuse coping, and PTSD of child sexual abuse survivors. *Child Abuse and Neglect*. 2005; 29(7):767–782.10.1016/j.chiabu.2005.01.005 [PubMed: 16051351]
- Walker EA, Newman E, Dobie DJ, Ciechanowski P, Katon W. Validation of the PTSD Checklist in an HMO sample of women. *General Hospital Psychiatry*. 2002; 24:375–380.10.1016/S0163-8343(02)00203-7 [PubMed: 12490338]
- Weathers FW, Keane TM, Davidson JRT. Clinician Administered PTSD scale: A review of the first ten years of research. *Depression and Anxiety*. 2001; 13:132–156.10.1002/da.1029 [PubMed: 11387733]
- Weathers, FW.; Litz, BT.; Herman, DS.; Huska, JA.; Keane, TM. The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility; Paper presented at the 9th Annual Conference of the ISTSS; San Antonio, TX. 1993.
- Weathers FW, Ruscio AM, Keane TM. Psychometric properties of nine scoring rules for the Clinician-Administered Posttraumatic Stress Disorder Scale. *Psychological Assessment*. 1999; 11(2):124–133.10.1037/1040-3590.11.2.124
- Yeager DE, Magruder KM, Knapp RG, Nicholas JS, Frueh BC. Performance characteristics of the posttraumatic stress disorder checklist and SPAN in Veterans, Affairs primary care settings. *General Hospital Psychiatry*. 2007; 4:294–301.10.1016/j.genhosppsych.2007.03.00 [PubMed: 17591505]

The PCL-S performs better in lab than as a participant recruitment device.

The PCL-S has acceptable performance as a recruitment tool among men.

The PCL-S does not perform acceptably as a recruitment tool among women.

Deficient performance among women may be due to women's generalized distress.

Figure 1.

Receiver Operating Characteristic (ROC) curve for telephone recruitment PCL-S scores versus CAPS diagnosis of PTSD in the total sample (Panel A), men (Panel B), and women (Panel C). The straight diagonal line represents the line of no information, while the curved line indicates the strength of the PCL-S's performance in accurately identifying PTSD.

Figure 2.

Receiver Operating Characteristic (ROC) curve for in-lab PCL-S scores versus CAPS diagnosis of PTSD in the total sample (Panel A), men (Panel B), and women (Panel C). The straight diagonal line represents the line of no information, while the curved line indicates the strength of the PCL-S's performance in accurately identifying PTSD.

Table 1
Diagnostic Characteristics of the Telephone Recruitment PCL-S by Cut-point

Cut-point	Sensitivity	Specificity	PPV	NPV	ODE
40	0.87 (0.75/0.94)	0.47 (0.73/0.09)	0.50 (0.50/0.50)	0.86 (0.89/0.60)	0.68 (0.70/0.55)
41	0.87 (0.75/0.94)	0.49 (0.76/0.13)	0.51 (0.52/0.51)	0.86 (0.89/0.67)	0.69 (0.71/0.59)
42	0.87 (0.75/0.94)	0.52 (0.80/0.13)	0.53 (0.57/0.51)	0.87 (0.90/0.67)	0.70 (0.74/0.59)
43	0.87 (0.75/0.94)	0.52 (0.80/0.13)	0.53 (0.57/0.51)	0.87 (0.90/0.67)	0.70 (0.74/0.59)
44	0.87 (0.75/0.94)	0.53 (0.82/0.13)	0.53 (0.60/0.51)	0.87 (0.90/0.67)	0.70 (0.75/0.59)
45	0.85 (0.75/0.90)	0.58 (0.84/0.22)	0.56 (0.63/0.53)	0.87 (0.90/0.70)	0.72 (0.77/0.62)
46	0.82 (0.75/0.87)	0.64 (0.84/0.34)	0.58 (0.63/0.56)	0.86 (0.90/0.73)	0.72 (0.77/0.65)
47	0.77 (0.75/0.77)	0.69 (0.87/0.44)	0.60 (0.67/0.57)	0.83 (0.91/0.67)	0.72 (0.79/0.62)
48	0.60 (0.63/0.58)	0.74 (0.87/0.56)	0.58 (0.63/0.56)	0.75 (0.87/0.58)	0.67 (0.75/0.57)
49	0.57 (0.63/0.55)	0.75 (0.89/0.56)	0.59 (0.67/0.55)	0.74 (0.87/0.56)	0.67 (0.77/0.56)
50	0.53 (0.56/0.52)	0.77 (0.89/0.59)	0.58 (0.64/0.55)	0.73 (0.85/0.56)	0.67 (0.75/0.56)
51	0.53 (0.56/0.52)	0.78 (0.91/0.59)	0.60 (0.69/0.56)	0.73 (0.85/0.56)	0.67 (0.77/0.56)
52	0.49 (0.56/0.45)	0.81 (0.91/0.66)	0.61 (0.69/0.56)	0.72 (0.85/0.55)	0.67 (0.77/0.56)
53	0.49 (0.56/0.45)	0.81 (0.91/0.66)	0.61 (0.69/0.56)	0.72 (0.85/0.55)	0.67 (0.77/0.56)
54	0.45 (0.50/0.42)	0.81 (0.91/0.66)	0.58 (0.67/0.54)	0.70 (0.84/0.54)	0.64 (0.76/0.54)
55	0.40 (0.44/0.39)	0.83 (0.91/0.72)	0.59 (0.63/0.57)	0.70 (0.82/0.43)	0.65 (0.73/0.56)
56	0.36 (0.38/0.35)	0.87 (0.96/0.75)	0.63 (0.75/0.58)	0.69 (0.81/0.42)	0.66 (0.78/0.57)
57	0.34 (0.31/0.35)	0.90 (0.98/0.78)	0.67 (0.83/0.61)	0.69 (0.80/0.39)	0.68 (0.82/0.61)
58	0.28 (0.31/0.24)	0.90 (0.98/0.78)	0.62 (0.83/0.53)	0.67 (0.80/0.52)	0.65 (0.82/0.53)
59	0.28 (0.31/0.26)	0.92 (0.98/0.84)	0.68 (0.83/0.62)	0.68 (0.80/0.54)	0.68 (0.82/0.58)
60	0.21 (0.31/0.16)	0.92 (0.98/0.84)	0.63 (0.83/0.50)	0.66 (0.80/0.51)	0.65 (0.80/0.51)

Notes. Values for the total sample are presented first, then men's values and women's values, respectively, are included in parentheses; PPV = positive predictive value; NPV = negative predictive value; ODE = overall diagnostic efficiency.

Table 2
Diagnostic Characteristics of the In-Lab PCL-S by Cut-point

Cut-point	Sensitivity	Specificity	PPV	NPV	ODE
40	0.87(0.75/0.95)	0.62(0.77/0.39)	0.57(0.52/0.59)	0.89(0.90/0.87)	0.71(0.77/0.66)
41	0.87(0.75/0.94)	0.63(0.77/0.42)	0.58(0.52/0.60)	0.89(0.90/0.88)	0.72(0.77/0.67)
42	0.87(0.75/0.94)	0.65(0.79/0.45)	0.59(0.55/0.62)	0.90(0.90/0.88)	0.73(0.78/0.69)
43	0.83(0.63/0.94)	0.69(0.81/0.52)	0.61(0.53/0.64)	0.88(0.87/0.89)	0.74(0.77/0.72)
44	0.81(0.63/0.90)	0.70(0.83/0.52)	0.61(0.56/0.64)	0.86(0.87/0.85)	0.74(0.78/0.70)
45	0.79(0.64/0.87)	0.70(0.83/0.52)	0.61(0.56/0.63)	0.85(0.87/0.81)	0.74(0.78/0.69)
46	0.72(0.56/0.81)	0.73(0.85/0.55)	0.61(0.56/0.63)	0.82(0.85/0.75)	0.73(0.78/0.67)
47	0.70(0.56/0.77)	0.75(0.85/0.61)	0.62(0.56/0.65)	0.81(0.85/0.74)	0.73(0.78/0.69)
48	0.70(0.56/0.77)	0.77(0.85/0.64)	0.63(0.56/0.67)	0.82(0.85/0.75)	0.74(0.78/0.70)
49	0.70(0.56/0.77)	0.79(0.85/0.70)	0.66(0.56/0.71)	0.82(0.85/0.77)	0.76(0.78/0.73)
50	0.70(0.56/0.77)	0.79(0.85/0.70)	0.66(0.56/0.71)	0.82(0.85/0.77)	0.76(0.78/0.73)
51	0.66(0.56/0.71)	0.80(0.88/0.70)	0.66(0.60/0.69)	0.80(0.86/0.72)	0.75(0.77/0.70)
52	0.64(0.56/0.68)	0.81(0.88/0.73)	0.67(0.60/0.70)	0.80(0.82/0.71)	0.75(0.77/0.70)
53	0.62(0.56/0.65)	0.81(0.88/0.73)	0.66(0.60/0.69)	0.79(0.82/0.69)	0.74(0.77/0.69)
54	0.57(0.56/0.58)	0.81(0.88/0.73)	0.64(0.60/0.67)	0.77(0.82/0.75)	0.73(0.77/0.66)
55	0.57(0.56/0.58)	0.85(0.90/0.79)	0.69(0.64/0.72)	0.78(0.86/0.67)	0.75(0.77/0.69)
56	0.55(0.56/0.55)	0.88(0.92/0.82)	0.72(0.69/0.74)	0.77(0.87/0.66)	0.76(0.83/0.69)
57	0.43(0.50/0.39)	0.88(0.92/0.82)	0.67(0.67/0.67)	0.72(0.85/0.59)	0.71(0.81/0.61)
58	0.38(0.44/0.35)	0.89(0.92/0.85)	0.67(0.64/0.69)	0.71(0.83/0.58)	0.79(0.80/0.61)
59	0.32(0.31/0.32)	0.89(0.92/0.85)	0.63(0.56/0.67)	0.69(0.80/0.57)	0.68(0.77/0.59)
60	0.28(0.25/0.29)	0.89(0.92/0.85)	0.59(0.50/0.64)	0.68(0.79/0.56)	0.66(0.75/0.58)

Notes. Values for the total sample are presented first, then men's values and women's values, respectively, are included in parentheses; PPV = positive predictive value; NPV = negative predictive value; ODE = overall diagnostic efficiency.