APPENDIX A. SEARCH STRATEGIES

I. Ovid MEDLINE

Search Strategy:

1 exp mass screening/ or screen*.mp.

2 exp Stress Disorders, Post-Traumatic/

3 (posttraumatic stress or posttraumatic stress disorder* or post-traumatic stress disorder* or ptsd).mp.

4 combat disorder*.mp. or exp Combat Disorders/)

5 or/2-4

6 1 and 5

7 limit 6 to (english language and humans and yr="1981 -Current")

II. PILOTS Database

Search textword "screen*"

With Limits:
English language
1980 -2012

And Descriptor categories:
"self report instruments, adults" or "self report instruments", "ptsd assessment instruments", "dissociation assessment instruments", "acute stress disorder assessment instruments", "assessment instruments" or "assessment"

Return to Contents

APPENDIX B. STUDY SELECTION AND DATA EXTRACTION FORM

Title of Study										**Check if Background paper** ☐	
Journal					**First Author**				**Year 2004**	**Inclusion Eligibility? Y N**	
Screening Tool	**PCL** (version)	**PDS**	**Penn Inv**	**IES**	**DTI**	**DES**	**Miss. Scale**	**SPAN**	**IDCL**	**PC-PTSD**	**Other**

Base Rate of PTSD: **Response Rates:** Screening Sample: Interview Sample:

Scoring Stats	**Cut-Point(s)**	**Sensitivity (%)**	**Specificity (%)**	**PPV**	**NPV**	**+ LR**	**- LR**	**ROCd'**	**ROC c-stat (AUC)**	**Other Outcomes**

Diag-nostic Measure			**Administration**	**Notes:** (e.g., study design different from single cohort, Info on ease of administration, unique scoring method, etc.)
		Clinician-Administered PTSD Scale (**CAPS**)		
		MINI International Neuropsychiatric Interview	Face-to-Face	
		Comprehensive international diagnostic interview (**CIDI**)	Telephone	
		Diagnostic Interview Schedule (**DIS**)		
		The Structured Clinical Interview for DSM-IV Disorders (**SCID-I**).	Telehealth	

Special Samples

Women ☐ Medical ☐ Age 60+ ☐ TBI ☐ Minorities ☐ Sexual trauma ☐
Veterans ☐ Substance Abuse ☐ Vietnam only ☐ OEF/OIF ☐ Other (note) ☐

Psychometric Notes

+LR= sensitivity/(1-specificity) **-LR= (**1-sensitivity)/specificity
Good test= LR+ of at least 2.0 and LR- of 0.5 or less.

		Diagnostic (Gold Standard) Test		
		Positive	Negative	
Screening Test	Positive:	a (true positive)	b (false positive)	PPV=a/(a+b)
	Negative:	c (false negative)	d (true negative)	NPV=d/(c+d)
		Sens=a/(a+c)	Spec=d/(b+d)	

QUADAS Eval

Sample broadly representative of those who would be screened? **Y N**
Time b/w screen and dx interview: _____ **months**
Is it considered "concurrent"? **Y N**
Did whole/ random sample have dx interview?
Whole Random (%?)
Random Group Different from Total Group? **Y N**
Were diagnostic interviews conducted blindly? **Y N**
Was screen cut score confirmed on a separate or split sample? **Y N**
Was relevant clinical data available for interpretation of screen (like would be in clinic)? **Y N**
Were *reasons for* withdrawals or refusals in the study explained? **Y N**

Return to Contents

APPENDIX C. LEVELS OF EVIDENCE

Based on Criteria for the Rational Clinical Examination Series (Simel 2008)[7]

Level I Evidence

Independent, blind comparison of sign or symptom results with a "gold standard" of anatomy, physiology, diagnosis, or prognosis among a large number of consecutive patients suspected of having the target condition.

Independent: neither the test result nor the gold standard result are used to select patients for the study.

Blind: test and gold standard each applied and interpreted without knowledge of the result of the other.

Gold Standard: the results of biopsy, angiography, autopsy, xray, sonogram, physiologic study, follow-up, therapeutic response, etc. that establish the true anatomy, physiology, diagnosis or outcome of the target condition.

Target Condition: the anatomic or physiologic state, disease, syndrome, prognosis or therapeutic response that the sign or symptom is designed to identify.

Large Numbers: sufficient numbers of patients to have narrow confidence limits on the resulting sensitivity, specificity, or likelihood ratio.

Level II Evidence

Independent, blind comparison of sign or symptom results with a "gold standard" among a small number of consecutive patients suspected of having the target condition.

Small Number: insufficient numbers of patients to have narrow confidence limits on the resulting sensitivity, specificity, or likelihood ratio. (N.B. You should note that the definition of "small" is relative and depends on the size of all extant studies. For example, if you have several studies of many hundreds of patients, then a study of only 80 patients might be considered small.)

Level III Evidence

Independent, blind comparison of signs and symptoms with a "gold standard" among non-consecutive patients suspected of having the target condition. The short-coming here is restricting the study sample to a subset of patients who both underwent and generated definitive results on both the sign or symptom and the application of the gold standard. The results over-estimate accuracy.

Level IV Evidence

Non-independent comparison of signs and symptoms with a "gold standard" among "grab" samples of patients who obviously have the target condition plus, perhaps, normal individuals. In addition to the selection bias of Level III, these studies restrict their samples to the obvious, "black or white" presentations (sometimes even selected on the basis of their gold standard result) that don't need a clinical examination (other than pattern recognition), and exclude the

"shades of gray" that comprise the clinical spectrum of early as well as late, mild as well as severe, and other but commonly confused conditions. The results greatly over-estimate accuracy.

Level V Evidence

Non-independent comparisons of signs and symptoms with a standard of uncertain validity (which may even "incorporate" the sign or symptom result in its definition) among "grab" samples of patients plus, perhaps, normals. In addition to the biases of Level IV, these studies often include the sign or symptom result as part of a "lead standard," resulting in a self-fulfilling prophesy. The results extravagantly over-estimate accuracy.

Return to Contents

APPENDIX D. PEER REVIEW COMMENTS/AUTHOR RESPONSES

REVIEWER COMMENT	RESPONSE
1. Are the objectives, scope, and methods for this review clearly described?	
Yes. 1. This is an excellent and comprehensive review with a wealth of very useful information. 2. The objective of the report appears to be a literature synthesis of the feasibility and diagnostic accuracy of PTSD screening tools for primary care settings. This could be slightly clarified in the introduction, rather than the broad statement on literature on screening tools in general, since issues of screening effectiveness and clinical efficiency appear to be beyond the scope of the report. 3. For the most part, the key questions are clearly stated, but could avoid using "etc." in KQ #1 and 2, and instead clearly state the specific characteristics reviewed, and for KQ#2, list the specific psychometric properties of interest. I am not sure that the implementability issue fits better in KQ#2 than it would in KQ#1 or as a separate question, since that information is reviewed separately on page 18, and does not map to the levels of evidence framework used to evaluate the question of diagnostic accuracy (psychometric properties and utility) in KQ#2. 4. Explanation and application of levels of evidence need to be clearer, especially in the discrimination of levels II and III: a. The description of the shortcoming for level III, "patients who both underwent and generated definitive results on both the sign and symptom and the application of the gold standard" is not clear. I am wondering if this is an allusion to the verification bias where follow-up or administration of one part of the testing protocol is dependent of results from a prior part of the testing protocol (e.g. administering the gold standard first, and then the screen to all cases but only a sample of controls, as described in Simel). I'm also wondering if this may be an editing glitch, since this text is repeated in the summary for level IV, and this kind of non-independence would be more of a Level IV issue. b. The key element that can take a study from Level II to Level III is the use of non-consecutive patients that are selected on the basis of some factor other than eligibility for screening that would result in a non-representative sample and introduce bias. Such results do not reliably over estimate accuracy, as stated on p. 48. The effect of the bias will be due to how the sample was selected and the ways in which they differ from the target population. Examples cited in the STARD guidelines include: exclusion of patients with comorbid conditions or symptoms that could adversely affect test accuracy but would likely be present in the target population; studies in specialty settings where the spectrum of symptom expression is narrowed; or just non-consecutive and non-random selection of the sample. I would then assume that pronounced violations of sampling assumptions, such as case control studies, would be graded at Level IV.	1. Thank you 2. We have modified the statement of the objective of the review. 3. We have modified KQ1 and 2 as suggested. We agree that the implementation processes of screening would have best been covered in a separate question and have done so to improve clarity of the findings. 4. The descriptions of levels of evidence were taken from instructions for preparing a Rational Clinical Examination article. 4a. We agree that the shortcoming for Level III is verification bias – selection of patients for verification rather than inclusion of consecutive patients. We also agree that the section of text was mistakenly repeated in the summary 4b. We agree that selection bias is one of the main differences between Level II and Level III studies. We have clarified our application of these ratings in Appendix F.
5. The discussion of each screen under KQ#2 could be more complete and detailed. Not all psychometric properties included in the articles are consistently reported, including key indicators of diagnostic accuracy such as likelihood ratios and (if provided) post-test odds of a positive test. If only sensitivity and specificity are reported, it is important to include the prevalence of PTSD in the sample. This may be a minor issue, since most (but not all) of this information is in Table 5, but it is not clear why some specific statistics are pulled out in the text and that the type of statistics discussed are not completely consistent across measures, so the reader does not get a clear critique of the state of the evidence for each screener.	5. We have now made the text more consistent throughout.
Yes	Thank you

Return to Contents

REVIEWER COMMENT	RESPONSE
Yes and No Yes, the objectives of the review are clearly described through the three key questions: Question 1: What tools are used to screen for PTSD in primary care settings, and what are their characteristics (length, format, etc)? Question 2: What are the psychometric properties and utility of the screening (operating characteristics) and their implementability (ease of administration) in primary care clinics? Question 3: Do the psychometric properties and utility of each of the screening tools differ according to age, gender, race/ethnicity, substance abuse or other comorbidities? Yes, the scope of the review is on screening tools used and validated in primary care. No, the methods for the review are not always logical, accurate or clearly described 1. Study selection a. Rationale for why studies outside of the US were excluded was not provided. Discussion of how Veterans in VA primary care may differ from civilians in primary care was not addressed. Perhaps there are reasons why screening practices/recommendations might differ in VA versus civilian primary care. Greater rationale for the inclusion/exclusion of studies seems warranted.	1. We have addressed these points in the report. 1a. We included only studies done in the United States because of greater relevance to the care of US Veterans. There were no studies that compared screen efficiency or effectiveness across both Veteran and non-Veteran samples. It may be that a given screen performs better in one population vs. another or for PTSD associated with one type of trauma vs. another; however, given the absence of evidence this would be purely speculative on our part. Available evidence suggests that PTSD is under-recognized in non-Veteran primary care settings (c.f. Graves, 2011) suggesting that, from a healthcare system perspective, screening for PTSD might also facilitate further mental health evaluation and treatment among non-Veterans assuming available mental health resources. As to whether screening practices/recommendations do or should differ in Veteran vs. non-Veteran primary care settings is a matter of policy and resource availability not screen characteristics and so is beyond the scope of this review. We have clarified the rationale for inclusion/exclusion of studies.
b. Why were screens included that did not include PTSD items (e.g., GAD-7)? Was study selection based on administration of a PTSD gold standard in a primary care setting? If yes, other non-PTSD screens may need to be considered in the review (e.g., GHQ)	1b. We state that we included screens for multiple psychiatric disorders or multiple anxiety disorders if there was a study that investigated the ability of the screen to identify PTSD in a primary care setting. No other screens identified in our literature search process were eligible for inclusion.
c. If studies with fewer than 50 participants were excluded, why was the Lange et al, (2003) study included? There were only 49 women interviewed with the gold standard interview.	1c. We excluded studies with fewer than 50 patients in the screening population.
d. There appears to be an assumption that gold standards are equivalent. This may not be an accurate assumption. Furthermore, it seems important to recognize that there are different scoring algorithms within gold standards. For example, there are at least 9 different scoring rules for the CAPS and the selection of one over another will surely impact diagnostic accuracy. Granted, scoring rules are rarely presented in studies, but the importance of this should not be overlooked.	1d. We identified the gold standard diagnostic tool used in each study and noted where scoring for the gold standard differed from the scoring method described in Table 1. We agree that different gold standard instruments or scoring rules could alter the findings in a given study. As the reviewer notes, scoring rules are rarely presented in studies, as was true in the vast majority of studies included in this review. While we do not think that variation in gold standard instrumentation or scoring would appreciably alter the overall findings of the review, we have included a statement of that possibility in our limitations section.

Return to Contents

REVIEWER COMMENT	RESPONSE
2. Screen/study description a. The PC-PTSD does not include a stem that asks about traumatic events. This is inaccurately reflected in the description of the measure: "Respondents are asked about symptoms experienced in response to a traumatic event in the past month" (p. 13)	2a. This has been clarified.
b. The SPAN was not validated with a primary care sample in the original Meltzer-Brody study. It was "developed in a psychiatry clinic for the purpose of detecting PTSD in psychiatric populations with PTSD prevalence around 50%" (p14). Yes, it was argued that it could be used in settings with a lower prevalence, like primary care, and yes, it was tested in primary care setting in the Yeager et al., study, but it was not developed/validated in primary care.	2b. We include the SPAN because there was a study that tested it in primary care setting as noted above.
c. The review correctly recognizes that there are three different versions of the PCL, and three different scoring options (p.14). All three versions of the PCL are represented in the studies reviewed, and information on scoring algorithms is often missing. The review treats the PCL as a single screen and does not mention how scoring options may impact diagnostic accuracy. This seems problematic for the accuracy and validity of the review.	2c. We have clarified which version of the PCL was used in each study. However, while there are different versions of the PCL and different scoring approaches to the instrument (e.g., symptoms/symptom cluster, total score, etc.), we believe that the importance of these differences is greatly attenuated when the PCL is used as a screening tool rather than as a diagnostic tool, a tool to assess symptom change in treatment, or as a means to estimate population prevalence rates (see Wilkins et al., 2011). Because the function of a screening tool is to identify individuals in need of further evaluation, all PTSD screening tools have lower discriminability than one would expect from a diagnostic tool. The more relevant scoring issue is cut-score, and we made efforts to include information about multiple cut-scores when studies provided that information. Accordingly, we do not feel that the accuracy or the validity of the review has been compromised.
d. As previously mentioned, it is unclear why the GAD-7/GAD-2 is included in the review. The screen does not include any PTSD items.	2d. As stated above, we included screens if there was a study that investigated the ability of the screen to identify PTSD in a primary care setting. Although the GAD-7 or GAD-2 may not be specific to PTSD, whether it performs better or worse than a PTSD-specific screen was an empirical question we thought worth considering given an appropriate gold standard and study design.
3. Table 3: summary of screens used in primary care a. It is not clear which study was used to report on test-retest reliability	3a. References have been added to Table (see footnotes).
b. Although scoring may be the same for briefer versions of the PCL, test-retest reliability cannot be assumed to be the same.	3b. We have noted this on Table 3.
c. Should internal consistency be presented as well?	3c. Internal consistency has been noted on Table 3 where reported (see footnotes).
Yes and No Some things are clearly described, but further justification is needed for the decisions the authors chose to make, e.g., to include studies of non-Veterans given the target audience of this report. The absence of this content makes it difficult to judge the statement on p. 30 that there is no information as to whether a given screen performs better in Veteran or non-Veteran samples. The absence of such information may be of limited relevance.	We have clarified the inclusion and exclusion criteria. Our literature search yielded no studies comparing the performance of screening tests in Veteran and non-Veteran samples. We have now highlighted results of studies in Veterans in the discussion to make it more relevant for the target audience of the report.

Return to Contents

REVIEWER COMMENT	RESPONSE
Yes. As stated on page 1, the premise of screening for PTSD is "to facilitate mental health treatment engagement 1) earlier in the course of the illness and 2) to engage patients in treatment who might otherwise not be identified…" For this purpose, the report undertakes to identify PTSD screeners for primary care (pc) settings and evaluate them, using the published literature. Three questions were formulated, which address evidence on the utility (and relative utility) of available scales. The questions and the methodology to answer them are perhaps too narrowly formulated. This is especially the case when one becomes aware of the fact that the studies that have evaluated PTSD screeners in pc have not evaluated the impact of screenings in engaging mental health workers more effectively, in terms of reaching patients who would not be identified. As a result, the report is a technical evaluation of the studies that evaluated PTSD screeners in pc: their design, analysis, etc. The lion share of the work—the evaluation of screening (by any means) for mental health delivery, and the outcome in terms of improving health--- remains to be done.	Thank you. We agree, that there is important work that remains to be done involving the impact of screen use on the delivery of mental health care and on health outcomes. We included this in our recommendations.
2. Is there any indication of bias in our synthesis of the evidence?	
No. I do not see any evidence of bias.	Thank you
No	Thank you
Yes and No A. Not sure about bias, but there are some problematic statements about the PC-PTSD and PCL. 1. Appendix E: Evidence Tables (Prins et al., 2003) a. The PC-PTSD was evaluated in one VA Health Care Facility, not two different VA's in California. b. The CAPS was administered in person, not over the phone c. As noted in the Evidence Table, the use of blind interviews was "not reported". The assumption was made, however, that interviewers were not blind (versus not reported), and the study was given a level IV rating. Although not clear from the original study, interviewers were indeed blind. Perhaps "not-reported" findings can be followed-up rather than assumed to be negative. 2. Freedy et al., 2010 a. Similar to Prins et al., 2003 -- It is assumed that interviewers were not blind to the screen results because they were administered on the same day as the diagnostic interview. But, what was the order of administration? Did interviewers know how to interpret screen results (cutoff scores for screens)? 3. PCL a. The PCL version used in the Yeager et all study is not clear. In the study, the PCL is described as "a series of 17 questions about symptoms or signs of PTSD resulting from military experiences taking place within the past month". This suggests that the PCL-M was used. b. The PCL version used in the Prins et al., 2003 study is also not clear. However, a correction to the article was published with clear reference to the PCL-S (Prins & Ouimette, 2004, Primary Care Psychiatry, 9, p151). The review also states that 124 "woman" [sic], were screened and interviewed. That is incorrect, 167 participants completed both the PCL-S and the PC-PTSD.	1a. This has been corrected. 1b. Thank you for clarifying this. 1c. Thank you for providing this additional information. Given this clarification, we have now determined that this study should have a rating of Level III. 2a. We assumed that interviews were not blind not because of their timing relative to administration of the screen, but rather because non-blind evaluations may be biased (similar to RCTs), and so the absence of a clear statement indicating that diagnostic interviews were conducted blindly in most cases means that they were not. However, as suggested by this reviewer, we sent an email to Dr. Freedy requesting further information, but have not received a response in the more than one month since the email was sent. 3a. We have clarified that no version was specified in this study. 3b. We have clarified that the PCL-S was used in this study. We have replaced the data from the original paper with the data presented in the Corrigendum.
No. The report gives no indication of bias in any of the decision or text.	Thank you.
No	Thank you.

Return to Contents

REVIEWER COMMENT	RESPONSE
3. Are there any published or unpublished studies that we may have overlooked?	
Yes. There is some evidence that the PC-PTSD performs adequately in VA substance use populations (p. 37, item 3). See Kimerling et al., (2006) Addictive Behaviors 31(11).	We are familiar with the Kimerling (2006) study but did not include it in this review because the study sample was that of patients who were receiving substance abuse treatment and not those presenting in primary care clinics.
No. Question whether it was necessary to include studies done on MH population and instruments that are not specific screens for PTSD – specifically the GAD-7	As noted previously, we included screens if there was a study that investigated the ability of the screen to identify PTSD in a primary care setting.
Yes For excellent reviews of the PCL, including the importance of spectrum effects (e.g., age, race, etc), bias, and prevalence, please see: 1. McDonald, S.D. & Calhoun, P.S. (2010). The diagnostic accuracy of the PTSD Checklist: A critical review. Clinical Psychology Review. doi:10:1016/j.cpr.2010.06.012. 2. 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. doi: 10.1002/da.20837.	Thank you for sharing these references. These reviews provide excellent background information on the PCL but do not focus on studies conducted in primary care.
Yes The report is so comprehensive that I think it will surprise readers in its presentation of studies they may not know of. However, it could be even more complete in several respects: 1. There is a corrigendum to the Prins et al. 2003 study that reports critically important information about the PCL. There were significant errors in the 2003 report due to a software problem regarding the handling of missing data. The data reported on the PCL need to be based on the 2004 correction. 2. A paper by Calhoun and colleagues (2010) comparing the SPAN and the PC-PTSD may have been overlooked. 3. In meta-analysis it is common to ask authors for data needed to include the paper in the analysis. Was there any attempt to contact investigators for information that could have allowed an excluded paper to be included? If not, I recommend that the authors use this strategy if it possibly could yield additional studies to include in the review	Thank you. We have addressed your concerns. 1. We have updated the report based on the Corrigendum. 2. We reviewed this excellent paper but it did not meet our inclusion criteria. Subjects in that study were part of the Mid-Atlantic MIRECC post-deployment registry and consisted of Veterans who served in the military after to September 11, 2001. According to the authors, "Eligible Veterans were recruited through mailings, advertisements, and clinician referrals". As such, it was not eligible for this review. 3. We did not exclude studies because of missing information. As noted in the Literature Flow (Figure 1) studies were excluded if the study setting, population, or purpose did not meet our inclusion criteria.
No. No overlooked study on screening scales in primary care.	Thank you.
4. Please write any additional suggestions or comments below. If applicable, please indicate the page and line numbers from the draft report.	
Future directions #6 is an important point, and the authors may want to specifically refer to the need for studies of screening effectiveness in VA.	Thank you for this suggestion. We have now made our recommendations more specific.
P. 1: first paragraph: I don't think the screening is meant to "Identify PTSD," or to facilitate treatment engagement so much as to identify Veterans who need further evaluation and possibly treatment for PTSD. Similar issue in the more detailed paragraph near end of page 5. Screening is not necessarily correlated with reducing delays for treatment – in fact, in VA the typical concern from PTSD teams is that PC refers too many patients because of a positive screen, thus tying up the resources needed to reduce access delays (though screening can lead to earlier diagnosis and an opportunity for intervention earlier in the course of an individual's illness). These issues do receive some discussion in the "clinical consideration" paragraph on page 38.	Thank you for this feedback. We clarified the statement on page to indicate that screens are intended to facilitate detection of a condition (in this case PTSD), not to identify it directly. We agree that screening is not correlated with treatment; however, the purpose of screening programs is to increase the rate of treatment, particularly among those early in the course of the illness as you note. The concern you raise about too many patients having positive screens and the effect of this on limited clinical resources is an important one. This suggests that from a clinical standpoint the screen used by VA is too sensitive as it is currently employed; however, altering th as clear policy implicati

Return to Contents

REVIEWER COMMENT	RESPONSE
1. "Screening tools that focus on evaluating traumatic experiences are not likely to be clinically useful given the high population prevalence of traumatic events and the much lower conditional probability of developing PTSD (IOM, Breslau, Wang)" p.36 a. True, but the diagnostic precision of screens that include a trauma probe versus those that don't has not been empirically established. Perhaps inclusion of a trauma exposure question will decrease the number of false positives in primary care. Future research could compare screens with and without a trauma probe. b. Does this statement suggest that screening for military sexual trauma is not warranted? 2. "Very short screens (i.e., one or two items) performed less well than longer screens with positive likelihood ratios less than 3.0, making them less clinically useful" p. 3 PLUS, "Screens not specific to PTSD but for which there was a study that evaluated the ability of the screen to detect PTSD performed less well than those that focused on the detection of PTSD exclusively" p.37 a. Combined, these argue against the use of the SIPS or multi-purpose screens with only 1 or 2 items relevant to PTSD. b. So, moderate to longer PTSD screens seem to be better but the threshold for acceptable length is not clear. If "successful screening programs utilize instruments that are simple, valid, precise, and acceptable both clinically and socially" (p. 1), the remaining PTSD screens should be evaluated along these dimensions. For example, future research needs to determine preference and ease of administration based on number of items, reading level, response format, etc 3. "However, there were no high quality studies examining the performance of the PC-PTSD in a primary care setting" p.37 a. Perhaps Freedy, Prins, and Gore can be contacted for clarification on the QUADAS ratings, and subsequent changes made to level of evidence.	1a. The statement that you reference was meant to clarify the scope of the review. On the other hand, we agree that whether screen performance would be improved with inclusion of a traumatic exposure item(s) is a worthy empirical question. 1b. No. It simply clarifies the scope of the review. 2a. We agree with the reviewer's conclusion that the available evidence suggests that screens longer than 2 items perform better. 2b. We did not find any information that any of the screening tools used in the studies cited in this review were unacceptable to patients or administrators. The longest screening tool (27 items) was reported to take patients only 10 minutes to complete, suggesting that none of the screens would be administratively burdensome. However, given the absence of comparative information about patient or provider preferences regarding screening tools, further research would be needed to make definitive statements about these issues. 3a. We have updated the information from one of the studies mentioned and adjusted the quality assessment. We contacted the author of another study for clarification but did not receive a response. We did not find anything requiring clarification in the third study.
The report has the potential to be an important guide for both practice and research. It is well done is so many respects but it could be enhanced by additions to the text and tables. It also needs to be cleaned for typos, some of which are important (e.g., on p. 20 the paragraph on the PCL says in one place that there were 2 studies and in another that there were 3, Table 4 shows 3, and the paragraph mentions an additional study by Kimerling (2006) that does not appear in the table). Specific recommendations are as follows: 1. More detail is needed about how the quality assessment ratings were determined. Although detail is provide in the Appendices, I could not make the crosswalk between the QUADAS evaluation questions in Appendix B, Appendix C, the 5 criteria listed for each study in Appendix E, and the level of evidence rating. In fact, I don't see the clear connection between the QUADAS criteria and the QUADAS questions in Appendix B. For example, in QUADAS, representativeness is about whether the full range of patients to whom a test would be applied was included in the sample. It appears that sample representativeness—and not spectrum inclusiveness—was more important in evaluating studies for the report. The fact that a study had one site is mentioned in a couple places, even though this is not relevant to evaluating quality according to the QUADAS or RSES systems. Also, in some cases the problem appears to be missing information. RCES level 1 evidence requires that neither the test result nor gold standard was used to select patients. Yeager's study, which was rated at the highest level, is mentioned as being a random sample of participants from 4 sites, whereas Andrykowski's study is described as "women in remission from breast cancer." Note that there is a typo in Appendix C and elsewhere in the text: it should be "Rational" not "Rationale" Clinical Examination Series.	Thank you. We have corrected the typos and clarified the additional studies cited in the paragraph on the PC-PTSD. 1. Thank you for pointing this out. We have now included an additional table in our appendices (Appendix F) that clarifies the relationship between the individual QUADAS ratings and the overall Level of Evidence ratings. Now that we have included the crosswalk table in the report, we hope that study ratings have been clarified. The Andrykowski study was rated as a Level IV because the diagnostic interviews were not conducted blind to patient screening status. We have corrected the typo.

Return to Contents

REVIEWER COMMENT	RESPONSE
2. Figure 1 was illegible when the document was printed, even though it appeared fine on the screen. Also, I recommend providing the N for each reason the 122 excluded studies ruled out. 3. In Table 3 it would help to know the test-retest interval for each study. 4. Table 4 should specify the gold-standard measure used for each study and if relevant how it was scored, e.g., the "1/2" rule for the CAPS. 5. For Key Question 2, the amount of detail in the text about studies varies unsystematically. For example, there was no information on p. 20 about the sample used in the Prins 2003 study and a lot of information on p. 22 about the sample used in the Dobie 2002 study. 6. On p. 21 only 2 SPAN studies were discussed but the Table 4 lists 3. Freedy 2010 is excluded. 7. Caution is needed regarding the inferences that are drawn when relevant information is missing. For example, and perhaps most notably, on p. 22, the report says that it is unclear whether CAPS interviewers were blind to PCL scores in the Prins 2003 study but elsewhere the report specifically states that lack of blinding was a major flaw of this study. Lack of information about blinding is not the same thing as lack of blinding. Regardless, things like this are so important that it is worth asking authors for missing information. 8. Table 5 is difficult to read. The use of shading to indicate different screens does not provide enough clarity or distinctiveness. For example, the authors could use a separate leftmost column to indicate the screen, with the study information in a column to the right: Screen Author/Year Cutpoints Breslau Freedy 2010 xx Kimerling 2006 xx PC-PTSD Freedy 2010 xx Gore 2008 xx 9. Given that the report includes studies of both Veterans and non-Veterans, can any more be said about whether findings might generalize from one population to the other? 10. Given that the PC-PTSD is currently used for screening in both VA and DoD settings, can any more be said other than a recommendation for a study comparing it with other screening instruments? 11. I recommend rewording recommendation 3 on p. 38. There is plenty of evidence about how screening tools work in the presence of other comorbidities because comorbidity is the rule rather than the exception in PTSD. What is missing is information about whether there is differential performance as a function of comorbidity. 12. The relevance of recommendation 4 is unclear or perhaps is not clearly worded. There is evidence about depression and anxiety screening in Veterans.	2. We have added the number of studies for each exclusion reason. 3. We have added this information to Table 3 (see footnotes). 4. We have added the gold standard measure to Table 4. Studies did not typically report how the measure was scored. 5. We have reviewed this and standardized the amount of text. 6. We have added Freedy 2010 to the discussion of the SPAN studies. 7. As noted above, we have obtained information from one author and another author did not respond to our inquiry. 8. Thank you for the suggestion. We have modified the table. 9. A comment about Veterans vs. non-Veterans has been added to the discussion. 10. Our primary recommendation is for VA (and DoD) to evaluate whether use of the screen has improved health outcomes for Veterans and to examine the impact of its use on the healthcare system. 11. Thank you for the suggestion. We have reworded the statement and clarified our point. 12. We agree that this point needs rewording as well, and incorporated the intended point elsewhere.
It would be of interest to have a review of the literature on screening among Veterans of other countries. Can we learn anything from this literature? Can we learn anything from DoD screening?	We chose not to include DoD studies because screening among active duty service members is complicated by limited confidentiality, potential deleterious effect of mental health diagnoses on military careers, and greater levels of stigma related to mental health conditions compared to that seen in non-active duty populations.
5. Are there any clinical performance measures, programs, quality improvement measures, patient care services, or conferences that will be directly affected by this report? If so, please provide detail.	
Not at this time. PC-PTSD followed by PCL when indicated is current measure and this report is unlikely to affect that.	Thank you
1. It seems like data from the PCMHI office may be able to address the impact of PTSD screening on referrals to co-located care or specialty care (i.e., access to care measure). And, with the new OEF4 performance measure, it might be possible to look at screening and engagement with treatment (8 sessions within 14 weeks). 2. DSM5 is around the corner. The content validity and predictive validity of PTSD screens will need to be evaluated against these new diagnostic criteria.	1. We agree that evaluating the impact of screening implementation on service utilization is an important area that should be explored. 2. We agree and ha g DSM-5 modification

Return to Contents

REVIEWER COMMENT	RESPONSE
The performance measure for PTSD screening is simply an indicator of whether screening has occurred, so I think this answer is no.	Thank you
6. Please provide any recommendations on how this report can be revised to more directly address or assist implementation needs.	
1. Perhaps more focused statements can be made about how the review can inform policy, guide services, support performance measures, and direct future research. For example: a. Although additional research is needed on what screen is best for detecting PTSD in VA primary care, there are good reasons to screen for PTSD in VA (see guidelines propose by US Preventive Services Task Force). b. Currently, if a patient screens positive for PTSD, CPRS presents certain follow-up options/services. Indeed, the clinical reminder is not "resolved" until an option is selected. The report would be strengthened by addressing these options and perhaps making recommendations for additional ones. c. As previously noted, the relationship between PTSD screening and access to care, and type of care would enhance implementation needs. 2. For future research, more specific examples of what should be done is needed. For example, a. Which screens (moderate and longer screens?) should be compared in VA primary care clinics and on what dimensions (ease of administration, diagnostic accuracy)? b. How should the impact of spectrum effects be analyzed? Comparing AUC's may not be the best approach.	1a. To our knowledge the USPSTF does not currently recommend routine PTSD screening. However, VA has significant clinical and political impetus for conducting routine PTSD screens on Veterans who use VA services. 1b. Although the requirement to institute a particular clinical reminder may be a result of national VA policy, how the clinical reminders are implemented varies across VISNs, Consequently, it would be less helpful to make specific recommendations about how the performance measure should be resolved. 1c. We agree. 2a. We do not recommend any particular screening tool since all have their limitations. Specific recommendations for future research are delineated. b. If what the reviewer means by "spectrum effects" is subsyndromal PTSD, then we agree that this would have implications for the criterion of a study. Comparisons of screen AUC's across studies requires a comparable outcome criterion.
With the formal adoption of DSM-5 in May 2013, the relevance of the data based on DSM-IV are unclear. Data obtained from DSM-IV versions may not generalize to DSM-5 versions if and when such data would be available. The authors need to address this issue more directly and incorporate it into their recommendations.	Agreed. We have now included comments about the relevance of the review with respect to DSM-5.

Return to Contents

APPENDIX E. EVIDENCE TABLE

Author, Year Screen	Gold Standard	Screen Sample I. Age, gender, special population II. Response Rate	Interview Sample I. Age, gender, special population II. Response rate	QUADAS Item Ratings I. Representativeness II. Quality of Gold Standard III. Concurrent IV. % Interviewed V. Blind Interviews RCE Level of Evidence
Freedy 2010[11] **Breslau**	CAPS	I. Not reported; required to be ≥18 years old, English speaking, no gross cognitive impairment, medically stable II. 774 of 3728 approached in clinic consented (20.8%); 519 of 774 consented were contacted (67%); telephone interviews done in 411 (11% of those approached in clinic, 53% of those consented, 79% of those contacted for interview)	I. 53% 18-44 years old, 19% 45-54 years old, 19% 55-65 years old, 7% 66-75 years old, 1.2% ≥76 years 83% women 65% white, 32% African American, 3% other 45% married 24% high school education or less II. 100% of those screened	I. No (significant differences in gender and race from clinic population during recruitment period) II. Fair (telephone, experienced survey interviewers) III. Yes IV. 79% of those who were contacted for interview V. No **Level of Evidence: IV**
Kimerling 2006[13] **Breslau**	CAPS	I. Veterans; other screen sample characteristics NR II. 237 of 258 approached (92%) were eligible and completed Breslau scale	I. Veterans Mean age = 52 years (range 22 to 85) 61% women 68% white, 18% African American, 5% Hispanic, 5% Asian/Pacific Islander, 1% Native American, 3% other 44% married 59% employed II. 57% returned for interview (significantly higher percentage of women in participants vs. non-participants)	I. Yes II. Good (in person, trained psychologists) III. Yes, approximately 1 month IV. 57% of those who consented, completed Breslau scale and were eligible V. Yes **Level of Evidence: III**
Freedy 2010[11] **PC-PTSD**	*See above*			
Gore 2008[10] **PC-PTSD**	PSS-I	I. 21% <30 years; 24% 31-34 years old; 31% 41-50 years old; 16% 51-60 years old; 8% ≥61 years 60% male Recruited from 3 military health system primary care clinics in Washington, DC area (service members, retirees, and family members) II. estimated 87.4% (3234 of approximately 3700 approached) NOTE: participants first administered SIPS; subgroup participated in 2nd phase of study (PC-PTSD and structured clinical interview); unclear if all invited to participate in 2nd phase	I. 24% <30 years, 23% 31-34 years old, 31% 41-50 years old, 18% 51-60 years old, 4% ≥61 years 61% male II. 93% of those who consented to interview (213/229); 6.6% of those screened (213/3234)	I. Unclear; oversampled patients who responded “bothered a little” and “bothered a lot” to single screening question II. Fair (unclear if in-person or telephone, trained mental health professionals) III. Yes IV. 6.6% of those screened

Return to Contents

Author, Year Screen	Gold Standard	Screen Sample I. Age, gender, special population II. Response Rate	Interview Sample I. Age, gender, special population II. Response rate	QUADAS Item Ratings I. Representativeness II. Quality of Gold Standard III. Concurrent IV. % Interviewed V. Blind Interviews RCE Level of Evidence
Prins 2003[9] **PC-PTSD**	CAPS	I. Not reported; recruited from general medical and women's health clinics at a VA facility in California; required to have no gross cognitive impairment and English speaking II. Number approached for screening not reported	I. Mean age = 52 years* 34.0% male 33% married 43% unemployed 27% high school education or less II. 50% of those who completed screening (167/335); participants repeated the PC-PTSD at the interview NOTE: all screened individuals invited to participate in interview	I. VA sample from 1 VA in California with 50% response rate II. Good (in-person, trained psychologists) III. Yes IV. 50% V. Yes **Level of Evidence: III**
Gore 2008[10] **SIPS**	*See above*			
Freedy 2010[11] **SPAN**	*See above*			
Meltzer-Brody 2004[55] **SPAN**	MINI	I. Mean age = 34 years 100% female 43% white, 49% African-American 30% (n=88/292) reported a traumatic event and completed the SPAN II. 76% (292/384 approached)	I. Mean age = 35 years 52% white, 41% African American II. 11% of total sample (32/292) of total sample; 36% of those with trauma who were invited for interview (32/88)	I. Women presenting for annual exam at ob/gyn clinic; n=32 completed interview II. Good (in-person, psychiatrist) III. Not reported IV. 11% of total sample V. Yes **Level of Evidence: III**
Yeager 2007[8] **SPAN**	CAPS	I. Group 1 - Veterans (male & female) II. 74.1% (888/1198) I. Group 2 - Female Veterans (oversample) II. 69.2% (191/276)	I. 79% male 63% white II. Group 1 - 82% of those who completed screen (728/888) or 61% of those approached (728/1198) Group 2 – 68% of those who completed screen (130/191) or 47% of those approached (130/276) NOTE: completers more likely to be older and Caucasian; final analysis (combining Groups 1 and 2) included only Caucasians and African-Americans (840/1079 or 78% of those who completed screen; 840/1474 or 57% of those approached)	I. Random sample from 4 medical centers in southeastern US II. Good (telephone, trained clinicians) III. Yes, within 2 months IV. 57% of total sample V. Yes **Level of Evidence: I**

Return to Contents

Author, Year Screen	Gold Standard	Screen Sample I. Age, gender, special population II. Response Rate	Interview Sample I. Age, gender, special population II. Response rate	QUADAS Item Ratings I. Representativeness II. Quality of Gold Standard III. Concurrent IV. % Interviewed V. Blind Interviews RCE Level of Evidence
Andrykowski 1998[47] **PCL**	SCID NP PTSD module	I. Mean age = 57 years 95% Caucasian, 4% African-American, 1% Asian-American 22% high school education or less NOTE: all had diagnosis of Stage 0 to IIIA breast cancer, without surgery, chemotherapy, or radiotherapy for 6-72 months, in remission II. 84/107 (79%) consented; 2 later deemed ineligible	I. Same as screen sample II. Same as screen sample NOTE: participants completed PCL-C and SCID NP PTSD during one telephone interview	I. Women in remission from breast cancer II. Fair (telephone, doctoral-level students) III. Yes IV. 100% of those consenting; 77% of those invited V. No **Level of Evidence: IV**
Dobie 2002[53] **PCL**	CAPS	I. Mean age = 48 years 100% female 75% white, 9% black, 15% other 40% married 35% high school education or less II. 16% of those randomly selected for telephone interview (282/1763); 11% of total pool (282/2545)	I. Same as screening II. Same as screening NOTE: participants were older and more often divorced than eligible non-participants	I. Female Veterans (1 site) II. Good (in-person, clinician) III. Yes IV. 11% of total sample V. Yes **Level of Evidence: III**
Freedy 2010[11] **PCL**	*See above*			
Lang 2005[49] **PCL**	CIDI 2.1	I. Primary care from VA or university-affiliated clinic II. Approximately 60% of patients approached in clinic consented; 275/401 completed PCL-C (69% [65% reported in text]) (returned by mail) NOTE: reported that a randomly selected half of those who completed consent form and short set of instruments in waiting room were selected for diagnostic interview	I. Mean age = 48 years 48% male 57% Caucasian 53% married 23% high school education or less II. 186/401 completed CIDI (46% [44% reported in text]) 154/401 completed PCL-C and CIDI (38% [36.5% reported in text])	I. Primary care clinics (VA or university-affiliated) II. Fair (telephone, licensed psychologist or research assistant) III. Not reported IV. 38% of enrolled V. Yes **Level of Evidence: II**

Return to Contents

Author, Year Screen	Gold Standard	Screen Sample I. Age, gender, special population II. Response Rate	Interview Sample I. Age, gender, special population II. Response rate	QUADAS Item Ratings I. Representativeness II. Quality of Gold Standard III. Concurrent IV. % Interviewed V. Blind Interviews RCE Level of Evidence
Lang 2003[54] **PCL**	CIDI	I. 100% female Veterans (1 site) II. 56% agreed to participate and returned questionnaires (221/394) NOTE: 25 of 419 survey packets were undeliverable	I. Mean age = 53 years 82% Caucasian, 12% African-American; 6% other/ unknown 39% married 80% with 9-15 years of education NOTE: interviewed women were older, more likely Caucasian, more likely divorced, separated, or widowed; less likely to be never married II. 87% of those screened willing to be interviewed (192/221); 46% of those approached (192/419) 26% randomly selected for interview (49/192)	I. Female Veterans (1 site) II. Fair (telephone, CIDI designed for lay interviewers) III. Yes, within 1 month IV. 26% (randomly selected, n=49) V. Yes **Level of Evidence: II**
Prins 2003[9] **PCL-S**	*See above*			
Walker 2002[56] **PCL (not specified)**	CAPS	I. Mean age = 41 years 100% female 79% Caucasian; 6% African-American; 8% Asian, 2% Hispanic, 1% Native-American 51% married 57% college graduates II. Adjusted return rate of 62% (1225/1912 eligible)	I. Not reported II. Overall 21% (261 of 1225 who returned questionnaire) or 14% (261/1912 eligible) – See NOTE NOTE: 305 returned questionnaires and had history of childhood sexual maltreatment, 152 of 204 reached (74%) agreed to interview (or 50% of those with history of maltreatment who returned questionnaires) From sample of 250 without childhood maltreatment, 116 of 155 reached (75%) agreed to interview (or 46% of sample) 7 had missing PCL data; final sample was n=261	I. Women only; random sample of HMO members II. Unclear (Not reported if face-to-face or telephone; qualifications of administrators not reported) III. Yes, within 2 months IV. 50% of those who reported childhood maltreatment; 46% of sample without maltreatment V. Not reported **Level of Evidence: III**
Yeager 2007[8] **PCL (not specified)**	*See above*			

Return to Contents

Author, Year Screen	Gold Standard	Screen Sample I. Age, gender, special population II. Response Rate	Interview Sample I. Age, gender, special population II. Response rate	QUADAS Item Ratings I. Representativeness II. Quality of Gold Standard III. Concurrent IV. % Interviewed V. Blind Interviews RCE Level of Evidence
Gaynes 2010[12] **M-3**	MINI	I. Not reported (eligible patients were age 18 or older, English speaking, mentally competent, and attending primary care academic family medicine clinic) II. 54% of those approached (n=723)	I. Mean age = 45 years 71% female 67% white, 28% black, 5% other 49% married 55% high school education or less 21% unemployed II. complete date for 89% (647/723 who consented	I. One family medicine clinic, sample similar to overall clinic II. Fair (In person or telephone, research assistant) III. Yes, within 30 days IV. 89% V. Yes **Level of Evidence: I**
Houston 2011[50] **PDT-4A**	SCID	I. Not reported (eligible patients were age 18 or older, non-psychotic, and seen in primary care clinic) II. Not reported, 343 of those who completed an initial questionnaire were "qualified" for the study after initial interview by investigating physician	I. Mean age = 47 years 69% female 86% white 48% married II. 78% (343/440)	I. One primary care clinic II. Fair (telephone, "trained rater") III. Not reported IV. Not reported V. Not reported **Level of Evidence: IV**
Means-Christensen 2006[51] **ADD**	CIDI	I. Not reported II. 61% of patients approached (7738/12724)	I. Mean age = 42 years 62% female 65% Caucasian, 16% African American, 10% Hispanic, 4% Asian, 5% other II. 867 of 1494 that screened positive (58%) agreed to interview; 569 (38%) completed interview 452 of random sample of 1107 that screened negative (41%) agreed to interview; 232 (21%) completed interview	I. More interviews among positive screens II. Fair (telephone, trained CIDI interviewers) III. Yes, median of 14 days IV. 38% of those that screened positive; 21% of random sample of those that screened negative V. No **Level of Evidence: IV**
Kroenke 2007[14] **GAD**	SCID	I. Not reported II. 92% (2740/2982) completed questionnaire (including GAD-7); of 2740, the first 2149 were used for development and validation of the GAD-7	I. Mean age=47years 69% female 81% non-Hispanic white, 7% black, 9% Hispanic, 3% other 65% married 34% high school education or less II. 77% (1654/2149) agreed to interview; 58% (965/1654) randomly selected for interview	I. Yes-15 primary care sites in 12 states (part of research network) II. Fair (telephone, 1 of 2 mental health professionals) III. Yes, approximately 1 week IV. 100% (this analysis based on those who completed GAD-7 and were interviewed) V. Yes **Level of Evidence: I**

QUADAS = QUality Assessment of Studies of Diagnostic Accuracy included in Systematic reviews tool (Whiting 2003)[6]
RCE = Rational Clinical Examination (Simel 2008)[7] (see Appendix C)
*Baseline data from n=188 (56% of those who completed an initial screen)

Return to Contents

APPENDIX F. SPECIFIC ASSOCIATION OF RCE LEVEL OF EVIDENCE RATINGS TO QUADAS ITEM RATINGS AS APPLIED IN THIS REVIEW

RCE Level of Evidence Rating	**QUADAS Item 1** Sample size of screening sample	**QUADAS Item 2** Representativeness of screening sample viz. target population/ selection method	**QUADAS Item 3** Sample size/ representativeness of Interview sample viz. screening sample	**QUADAS Item 4** Quality of gold standard and its administration	**QUADAS Item 5** Blinded/concurrent diagnostic evaluations
I	Large	Representative of target population/randomly selected or consecutive sample	All of screening sample or randomly selected representative sample	In person by trained diagnostician	Yes
II	Small	Representative of target population/randomly selected or consecutive sample	All of screening sample or randomly selected representative sample	In person by trained diagnostician	Yes
III*	Small or Large	Representative sample or convenience/non-representative sample	Random selection or non-representative interview sample	In person or by telephone by trained diagnostician	Yes
IV†	Small	Convenience/non-representative sample	Non-random interview sample	Telephone by trained research assistants	No
V	*Not included in Systematic Review*				

QUADAS = QUality Assessment of Studies of Diagnostic Accuracy included in Systematic reviews tool (Whiting 2003)[6]
RCE = Rational Clinical Examination (Simel 2008)[7] (see Appendix C)
*Level III studies have either a small sample size and lower ratings on QUADAS 2 or QUADAS 3, or a larger sample size and lower ratings on both QUADAS 2 and QUADAS 3
†Level IV studies may have a higher rating on one of the QUADAS 1-4 criteria but have lower ratings in the other 3 criteria

Return to Contents