



Diagnostic test accuracy of screening tools for post-traumatic stress disorder among refugees and asylum seekers: A systematic review and meta-analysis

Olivia Magwood^{a,b,*}, Kara Bellai-Dussault^c, Grace Fox^c, Chris McCutcheon^b, Owen Adams^d, Ammar Saad^{a,c}, Azaad Kassam^{e,f,g}

^a Bruyère Research Institute, 85 Primrose Ave, Ottawa, ON K1R 6M1, Canada

^b Interdisciplinary School of Health Sciences, University of Ottawa, 125 University, Ottawa, ON K1N 6N5, Canada

^c School of Epidemiology and Public Health, University of Ottawa, 600 Peter Morand Crescent, Ottawa, ON K1G 5Z3, Canada

^d Canadian Medical Association, 1410 Blair Towers Place, Suite 500, Ottawa, ON K1J 9B9, Canada

^e Department of Psychiatry, University of Ottawa, 75 Laurier Ave. E, Ottawa, ON K1N 6N5, Canada

^f Pinecrest-Queensway Community Health Centre, 1365 Richmond Rd #2, Ottawa, ON K2B 6R7, Canada

^g Ottawa Newcomer Health Centre, 291 Argyle, Ottawa, ON K2P 1B8, Canada

ARTICLE INFO

Keywords:

Refugee

Asylum seeker

PTSD

Diagnostic test accuracy

Mental health screening

ABSTRACT

Background: Refugees and asylum seekers often experience traumatic events resulting in a high prevalence of post-traumatic stress disorder (PTSD). Undiagnosed PTSD can have detrimental effects on resettlement outcomes. Immigration medical exams provide an opportunity to screen for mental health conditions in refugee and asylum seeker populations and provide links to timely mental health care.

Objective: To assess the diagnostic accuracy of screening tools for PTSD in refugee and asylum seeker populations.

Methods: We systematically searched Medline, Embase, PsycINFO, CENTRAL and CINAHL up to 29 September 2022. We included cohort-selection or cross-sectional study designs that assessed PTSD screening tools in refugee or asylum seeker populations of all ages. All reference standards were eligible for inclusion, with a clinical interview considered the gold standard. We selected studies and extracted diagnostic test accuracy data in duplicate. Risk of bias and applicability concerns were addressed using QUADAS-2. We meta-analyzed findings using a bivariate random-effects model. We partnered with a patient representative and a clinical psychiatrist to inform review development and conduct.

Results: Our review includes 28 studies (4,373 participants) capturing 16 different screening tools. Nine of the 16 tools were developed specifically for refugee populations. Most studies assessed PTSD in adult populations, but three included studies focused on detecting PTSD in children. Nine studies looked at the Harvard Trauma Questionnaire (HTQ) with diagnostic cut-off points ranging from 1.17 to 2.5. Meta-analyses revealed a summary point sensitivity of 86.6% (95%CI 0.791; 0.917) and specificity of 78.9% (95%CI 0.639; 0.888) for these studies. After evaluation, we found it appropriate to pool other screening tools (Posttraumatic Stress Disorder Checklist, the Impact of Event Scale, and the Posttraumatic Diagnostic Scale) with the HTQ. The area under the curve for this model was 79.4%, with a pooled sensitivity of 86.2% (95%CI 0.759; 0.925) and a specificity of 72.2% (95% CI 0.616; 0.808).

Conclusions: Our review identified several screening tools that perform well among refugees and asylum seekers, but no single tool was identified as being superior. The Refugee Health Screener holds promise as a practical instrument for use in immigration medical examinations because it supports the identification of PTSD, depression, and anxiety across diverse populations. Future research should consider tool characteristics beyond sensitivity and specificity to facilitate implementation in immigration medical exams.

Registration: Open Science Framework: [10.17605/OSF.IO/PHNJV](https://doi.org/10.17605/OSF.IO/PHNJV)

* Corresponding author at: Bruyère Research Institute, 85 Primrose Ave, Ottawa, ON K1R 6M1, Canada.

E-mail address: omagwood@bruyere.org (O. Magwood).

1. Introduction

Many refugees and asylum seekers experience trauma related to war, persecution, loss of family members, and poor access to resources (Javanbakht et al., 2019), which can have long lasting impacts on mental health (Bogic et al., 2015; Steel et al., 2002). Post-traumatic stress disorder (PTSD) is a psychological condition triggered by the direct or indirect experience of a traumatic event (American Psychiatric, 2013), and has been observed among approximately 30% of all refugees and asylum seekers (Blackmore et al., 2020) and up to 53% among refugee and asylum seeking children (Kien et al., 2019). The high prevalence of PTSD in this population has led to the development of numerous screening tools to identify the condition early in the resettlement process (Davidson et al., 2010; Gadeberg et al., 2017).

When refugees and asylum seekers migrate to a new country, they undergo an immigration medical examination (IME) to determine their health status. Examinations are typically conducted by a registered medical practitioner and include a standard physical exam, questionnaire regarding current or past medical conditions, and laboratory or clinical tests (Wickramage and Mosca, 2014). The results of the IME are used to inform a clinical course of action where individuals are referred to health care professionals. The purpose of the IME is to support the wellbeing of migrating populations, but it is usually limited to physical health. The IME offers a unique opportunity to screen refugees and asylum seekers for mental health conditions and ensure timely referral to a specialist. This is important because undiagnosed mental health conditions can have a significant impact on employment, intergenerational trauma, and overall resettlement success (Blackmore et al., 2020). Screening for mental health concerns not only helps identify people in need of support, but also can help resettlement states avoid over-treating individuals who may be experiencing natural reactions to traumatic situations, thereby efficiently allocating mental health resources to those most in need. Immigration officials (e.g. UK, Australia, New Zealand, Canada) have highlighted this as an area of concern and are interested in incorporating mental health screening tools into IMEs to support the resettlement of refugees and asylum seekers (Hough et al., 2019; Immigration New Zealand (INZ) 2021).

Several systematic reviews have been conducted on PTSD in refugee and asylum seeker populations, but the focus has been on prevalence (Blackmore et al., 2020; Henkelmann et al., 2020; Peconga and Hogh Thogersen, 2020), access to care (Due et al., 2020), and community-based or psychological interventions (Crumlish and O'Rourke, 2010; Thompson et al., 2018; Uphoff et al., 2020; Soltan et al., 2022). Three systematic reviews have looked at mental health screening tools for refugees and asylum seekers, but none have tried to compare the diagnostic accuracy of tools for specific conditions (Davidson et al., 2010; Gadeberg et al., 2017; Hollifield et al., 2002). Gadeberg et al. focused on children and adolescents but found only three studies measuring sensitivity and specificity (Gadeberg et al., 2017). To date, there is a limited evidence base of reviews assessing the validity of PTSD screening tools in refugee populations and none conducted to guide the implementation of mental health screening tools within IMEs. The objective of this review is to inform immigration policy by assessing the diagnostic test accuracy of PTSD screening tools in refugee and asylum seeker populations.

2. Materials and methods

2.1. Patient engagement

We partnered with a patient representative with lived experience of resettlement and expertise in the field of mental health. Our patient representative was consulted in the writing of the protocol and was critical in the development of the inclusion criteria, search parameters and data extraction items, and was included in group discussions related to data synthesis and knowledge translation. They reviewed this

manuscript in its entirety and ensured its relevance and appropriateness to the population of interest (see GRIPP-2 Short Form in Appendix 1) (Staniszewska et al., 2017).

2.2. Protocol and registration

We conducted this systematic review according to an a priori protocol (Magwood et al., 2021). We report this systematic review according to the Preferred Reporting Items for Systematic Reviews and Meta Analyses reporting checklist Diagnostic Test Accuracy extension (PRISMA-DTA) (Appendix 2) (McInnes et al., 2018).

2.3. Eligibility criteria

We included studies which met the following criteria:

2.3.1. Types of studies

We included primary studies assessing the diagnostic test accuracy of screening tools for PTSD in refugee and asylum seeker populations. Prospective and retrospective cohort-selection or cross-sectional study designs were eligible for inclusion (Mathes and Pieper, 2019). We excluded diagnostic case-control designs, which are characterized by investigators recruiting disease-positive (i.e., case) and disease-negative (i.e., control) participants. Case-control design is prone to bias, potentially leading to inflated estimation of the diagnostic performance (Whiting et al., 2013; Park, 2019). We included studies which reported sensitivity and/or specificity. If measures of sensitivity or specificity were not reported, we included studies which provided the data necessary to calculate them (e.g., extraction of a 2 × 2 contingency table). For feasibility, studies in languages other than English, French, Spanish, and Arabic were not eligible. No date limitation was applied.

2.3.2. Participants

Studies were eligible if the screening tools were used in refugee and/or asylum seeker populations of any age. Screening done in any geographical location and in any clinical setting was eligible. We included articles that included refugees and/or asylum seekers among other population groups as long as subgroup data exclusive to refugee and/or asylum seeker populations were available.

2.3.3. Index tests

We included studies which assessed and reported the diagnostic accuracy of a screening tool. The term "screening tool" was inclusive of any questionnaire, checklist, or interview guide designed to identify psychiatric symptoms. Screening tools may be self-administered or administered by a clinical or non-clinical professional. The index tests in the studies we retained either had a prior history of development and validation among groups at risk of PTSD or were being compiled from established tools and tested in refugee groups. A screening tool was considered invalid if questions were not established before screening. Screening tools administered using any method of delivery (e.g., written or verbal, self-assessment or clinician-administered, etc.) and at any point in time were eligible. The reference standard for diagnosis of PTSD is the clinical interview, but we also included studies that compared screening tools to other diagnostic proxies.

2.3.4. Target conditions

We included screening tools administered with the intent to identify psychiatric or clinical symptoms of PTSD, as defined according to the criteria of either the Diagnostic and Statistical Manual of Mental Disorders (DSM) or the International Classification of Diseases (ICD) (American Psychiatric, 2013; World Health Organization, 1992). We included studies which used the definitions from any published version of the DSM or ICD criteria. Studies which included screening tools for PTSD in addition to other psychiatric conditions were included as long as the diagnostic test accuracy data were available for the PTSD

component independent of other conditions.

2.4. Information sources

We developed a search strategy in consultation with a health sciences librarian to search the following bibliographic databases: Medline (Ovid), Embase (Ovid), PsycINFO (Ovid), CENTRAL (Ovid) and CINAHL (Ebsco). We searched from database inception up to 29 September 2022. The search consisted of keywords and MeSH headings. We elected not to search the gray literature but have consulted experts in the field and authors of included studies, whenever possible, to identify other relevant literature. Additionally, we screened the reference lists of relevant systematic reviews and considered their included studies against our eligibility criteria.

2.5. Search

Our search strategy was translated for all databases and was peer reviewed by a health scientist librarian. Complete search strategies are available in Appendix 3.

2.6. Study selection

We used the review management tool Covidence to manage identified studies and facilitate title/abstract and full-text screening (Veritas Health Innovation 2019). All reviewers contributed to screening and piloted the inclusion criteria on the first 250 citations. After achieving a 90% consensus, the remaining titles and abstracts were screened independently and in duplicate. Conflicts were resolved through discussion. Titles and abstracts deemed potentially eligible were assessed independently through full-text review by two reviewers. Discrepancies were resolved through discussion with the entire review team.

2.7. Data collection

We developed a standardized data extraction sheet. Three reviewers independently piloted the data extraction sheet on a sample of three studies, and revisions were made as necessary. All reviewers contributed to data extraction. Relevant data items were extracted from all studies in duplicate. Consensus was achieved by a third reviewer. We made one attempt to contact authors for studies in which reporting was unclear ($n = 2$) (Sondergaard et al., 2003; Nehring et al., 2021).

2.8. Definitions for data extraction

We extracted the following variables from all included studies: Publication and year, country, setting, study sample size, participant characteristics, reference standard, index test (screening tool), language (s), number and types of items/domains, response format/scale design, threshold for positivity (cut-off) and interpretation, target populations (child/adolescent/adult), developed for refugee populations (y/n), adapted for refugee populations (y/n), validated for refugee populations (y/n), description of cultural/linguistic elements, professional background/training of assessor, presence of interpreter (y/n), mode of administration, sensitivity/specificity, study conclusions and implications for future research.

2.9. Risk of bias and applicability

We critically appraised the methodological rigor of included studies using QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2) (Whiting et al., 2011). The QUADAS tool consists of 4 key domains that discuss patient selection, index test, reference standard, flow of patients through the study, and timing of the administration of the index tests and reference standard. We tailored QUADAS-2 signaling questions for our review: We defined a low risk of bias reference standard as a

semi-structured clinical interview according to DSM or ICD criteria. Studies which used diagnostics proxies or other approaches were considered high risk for bias. We also defined an appropriate interval between screening with the index test and confirmation with the reference standard as “up to three months.” Two reviewers applied QUADAS-2 independently. Consensus was achieved through discussion. We created summary figures using the Risk-of-bias VISualization (ROBVIS) tool (McGuinness and Higgins, 2021).

2.10. Diagnostic accuracy measures

Our primary measures of interest were sensitivity and specificity, where the unit of assessment is the individual patient.

2.11. Synthesis of results

For each included study, two-by-two contingency tables containing the raw data (true-positive, false-negative, false-positive, and true-negative findings) were tabulated on the basis of the provided raw data or by calculation from the sensitivity and specificity described in each study. We then calculated diagnostic accuracy measures of sensitivity and specificity for each tool and all available cut-offs (thresholds for positivity). Individual study results for sensitivity and specificity were plotted on a forest plot to visually assess and explore study variability. With regard to the explanation for the variability seen between studies, we identified 2 possible sources of variability: (1) the index test (screening tool), which varied both in format and administration across studies, and (2) threshold effect, which is the criteria used to define a positive test result.

To make determinations about suitability for meta-analysis, we assessed the clinical heterogeneity of all tools. Studies were first grouped by screening tool, and we conducted meta-analyses if more than four studies used a common tool. Next, we mapped all tools against the DSM-IV criteria of re-experience, avoidance and increased arousal dimensions of PTSD (See Appendix 4). We identified the Harvard Trauma Questionnaire, Posttraumatic Diagnostic Scale, Posttraumatic Stress Disorder Checklist for DSM-5 and Impact of Event Scale as being sufficiently similar to pool their findings in a meta-analysis.

2.12. Meta-analysis

We used a bivariate random-effects model to pool data from studies which used the same index tests to account for correlation between sensitivity and specificity. Several studies included diagnostic testing accuracy results for multiple thresholds. We first meta-analyzed findings using the optimal cut-off reported for each study population, following guidance from the Cochrane DTA Handbook (Macaskill et al., 2010). We then generated a summary receiver operating (sROC) curve for these studies. Studies using the same index test with a common cut-off were pooled to generate summary points for sensitivity and specificity. We conducted an additional analysis by pooling multiple index tests that were deemed comparable. Typical measures of heterogeneity, such as the I^2 statistic, were not used to assess statistical heterogeneity, as recommended by the Cochrane DTA Handbook (Macaskill et al., 2010). We used the version 4.0.3 of R to conduct all analyses, following the recommended analyses from Shim and colleagues (Shim et al., 2019). A copy of the code used for this analysis can be found in Appendix 5.

3. Results

3.1. Study selection

Our search retrieved 8928 citations. After removal of duplicates, two reviewers screened 4631 titles and abstracts independently against our inclusion criteria. One hundred and seven citations were identified for full text review, of which 25 met all inclusion criteria. We conducted a

backward citation analysis by searching the reference lists of these 25 studies. An additional 21 studies were assessed at full text, of which 3 were included. Overall, we included 28 studies that met all eligibility criteria. Study flow and reasons for exclusion for all studies assessed at full text are reported in Fig. 1 and Appendix 6, respectively.

3.2. Study characteristics

Our review includes 28 studies (see Table 1) from which 16 different screening tools were identified. Thirty-three validation tests were conducted, with some studies evaluating multiple tools or different versions of the same tool. Nine studies looked at the Harvard Trauma Questionnaire (HTQ) (Sondergaard et al., 2003; de Fouchier et al., 2012; Jakobsen et al., 2011; Jakobsen et al., 2017; Lhewa et al., 2007; Mollica et al., 1992; Renner et al., 2006; Smith Fawzi et al., 1997; Blackmore et al., 2022), making it the most frequently researched tool. This is followed by the Impact of Event Scale (IES) (Sondergaard et al., 2003; Renner et al., 2006; Morina et al., 2017; Sack et al., 1998), which was validated in four different studies, and the Refugee Health Screener (RHS) (Bjarta et al., 2018; Hollifield et al., 2016; Kaltenbach et al., 2017), validated in three studies. The PROTECT Questionnaire (PQ) (Mewes et al., 2016; Wulfes et al., 2019), the PTSD Checklist (PCL-5) (Heeke et al., 2020; Ibrahim et al., 2018), and the Posttraumatic Diagnostic Scale (PDS) (Wulfes et al., 2019; Turner et al., 2003) were each validated by two studies. The remaining screening tools were only validated in single studies. Nine of the 16 tools were developed specifically for refugee populations. The most common tools used similar response formats using a Likert-scale. For example, the RHS-15 contains 14 items that address issues of symptoms related to anxiety, depression, pain, and ability to cope. In addition, there is a “Distress Thermometer”. The symptoms items are to “indicate the degree to which the symptom has been bothersome to you over the past month” and are scored from 0 (not at all) to 4 (extremely). The coping item responses range from 0 (able to handle or cope with anything) to 4 (unable to handle or cope with anything). The Distress Thermometer (DT) is presented like a thermometer ranging from 0 (no distress, things are good) to 10 (extreme distress, I feel as bad as I ever have). The 13 symptom and 1 coping item responses are summed to obtain a 14-item total score

(Hollifield 2016). Similarly, the HTQ and IES scales assess the previous week’s experiences using a 4-point Likert scale ranging from 1 (not at all) to 4 (extremely), which are used to establish an overall score.

Most studies used a diagnostic interview as the reference standard ($n = 21$) (Sondergaard et al., 2003; Nehring et al., 2021; de Fouchier et al., 2012; Jakobsen et al., 2011; Jakobsen et al., 2017; Lhewa et al., 2007; Mollica et al., 1992; Smith Fawzi et al., 1997; Blackmore et al., 2022; Morina et al., 2017; Sack et al., 1998; Kaltenbach et al., 2017; Ibrahim et al., 2018; Turner et al., 2003; Ahmad et al., 2000; Brink et al., 2016; Dao et al., 2012; Eytan et al., 2007; Hocking et al., 2018), but the diagnostic instrument varied. For example, some studies used the Composite International Diagnostic Interview (CIDI) or the Structured Clinical Interview (SCID), while some did not report details of the diagnostic interview. Five studies used other screening tools, or diagnostic proxies, as the reference standard (Bjarta et al., 2018; Hollifield et al., 2016; Heeke et al., 2020; Barbieri et al., 2019; Lillee et al., 2015). In order to address concerns of PTSD diagnostic criteria failing to apply across cultures, two studies developed their own culturally specific reference standards (Renner et al., 2006; Hall et al., 2014). For example, Hall et al. developed a community diagnosis procedure, where social workers identified children in need of treatment, and these referrals were validated through interviews with the children and their caregivers (Hall et al., 2014). Renner et al. used focus groups to evaluate transcripts of interviews and considered psychological and somatic complaints to determine if participants were traumatized (Renner et al., 2006).

Four different modes of delivery were employed to administer the screening tools. Most instruments were self-report questionnaires that were either self-administered ($n = 12$) (Sondergaard et al., 2003; de Fouchier et al., 2012; Jakobsen et al., 2011; Jakobsen et al., 2017; Mollica et al., 1992; Smith Fawzi et al., 1997; Morina et al., 2017; Bjarta et al., 2018; Kaltenbach et al., 2017; Heeke et al., 2020; Brink et al., 2016), with staff assistance ($n = 10$) (Nehring et al., 2021; Lhewa et al., 2007; Renner et al., 2006; Blackmore et al., 2022; Sack et al., 1998; Mewes et al., 2016; Turner et al., 2003), or through an interview ($n = 5$) (Hollifield et al., 2016; Wulfes et al., 2019; Ibrahim et al., 2018; Barbieri et al., 2019). Eight of the tools were designed to be administered as interviews (Sondergaard et al., 2003; Renner et al., 2006; Ahmad et al., 2000; Dao et al., 2012; Eytan et al., 2007; Hocking et al., 2018; Lillee

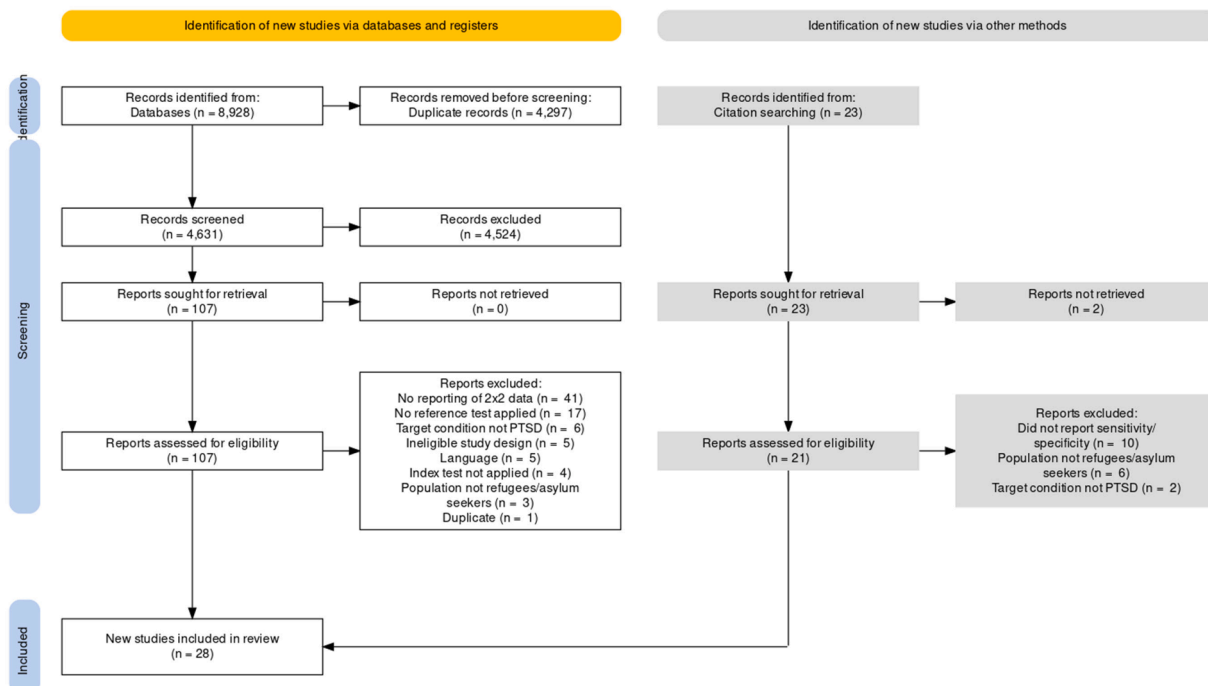


Fig. 1. PRISMA Flow Diagram.

Table 1
Characteristics of included studies.

Study	Study setting (country)	Screening Tool	Method of Delivery	Screening Personnel	Language of Tool	Reference Standard	Target Population	Target Condition	Sample size, gender, mean age	Country of origin
Ahmad et al. (2000) (Ahmad et al., 2000)	Sweden	Posttraumatic Stress Symptoms in Children (PTSS-C)	Semi-structured interview	Interviewers experienced working with children and trained in PTSS-C	English	Diagnostic Interview for Children and Adolescents (DICA), DSM-IV adjusted	Children	PTSD	n=66; gender not reported; mean age=12	Kurdistan
Barbieri et al. (2019) (Barbieri et al., 2019)	Italy	ICD-11 PTSD criteria	Self-report via interview	Cultural mediator; physician; psychologist	Arabic, English, French	PTSD Checklist for DSM-V (PCL-05)	Adults	PTSD, CPTSD	n=120; female=17, male=103; mean age=25.1	19 African countries
Bjærtå et al. (2018) (Bjarta et al., 2018)	Sweden	Refugee Health Screener 13 (RHS-13)	Self-report (paper or online questionnaire)	Bilingual staff provided technical assistance	English, Swedish, Arabic, Dari, Farsi, Somali, Tigrinya	PC-PTSD-4, DSM-IV	Adults	PTSD, anxiety, depression	n=520; female=136, male=367; mean age=not reported	Afghanistan, Syria, Iraq, Iran, Eritrea, Somalia
Blackmore et al. (2022) (Blackmore et al., 2022)	Australia	Harvard Trauma Questionnaire (HTQ)	Self-report questionnaire (assisted)	Interpreters	English	Structured clinical interview, DSM-V (research version)	Adults	PTSD	n=52; female=52; male=0; mean age=27.6	Afghanistan
Brink et al. (2016) (Brink et al., 2016)	USA	Karen Mental Health Screener	Self-report questionnaire	Physician with mental health training and interpreter	English, Karen	Structured Clinical Interview (SCID-4-CV), DSM-IV	Adults	PTSD, depression	n=177; female=126, male=54 (as per original sample of 180); mean age=38.1	Burma
Dao et al. (2012) (Dao et al., 2012)	USA	Posttraumatic Stress Disorder Interview for Vietnamese Refugees (PTSD-IVR)	Interview	Clinician fluent in Vietnamese	Vietnamese	LEAD procedure for diagnostic interview, DSM-IV-TR	Adults	PTSD	n=127; female=103, male=24; mean age=46.2	Vietnam
de Fouchier et al. (2012) (de Fouchier et al., 2012)	France	Harvard Trauma Questionnaire (HTQ)	Self-report questionnaire	N/A	French	Structured Clinical Interview (SCID), DSM-IV-TR	Adults	PTSD	n=52; female=23, male=29; mean age=36.6	DRC, Ivory Coast, Cameroon, Guinea, Rwanda, Mali
Eytan et al. (2007) (Eytan et al., 2007)	Switzerland	Mini International Neuropsychiatric Interview (MINI), MDE and PTSD sections	Interview	Nurses and interpreters	French, English	Diagnostic interview with cultural adaptation, DSM-IV	Adults	PTSD, major depression	n=101; female=26, male=75; mean age=30	33 countries, mostly African and Central and Eastern European
Hall et al. (2014) (Hall et al., 2014)	Ethiopia	Child Posttraumatic Stress Disorder Symptom Scale Interview format (CPSS-I)	Interview (parent and children pairs)	Native Somali speakers with 3-day training in structured interview techniques	Somali	Author-developed "community diagnosis" procedure	Children	PTSD	n=159 (child/caregiver pairs); child: female =92, male= 67; mean age= 11; caregiver: female=140; male=19; mean age=39.2	Somalia
Heeke et al. (2020) (Heeke et al., 2020)	Germany	Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5)	Self-report questionnaire	Psychologists and interpreters	Farsi, Arabic, Turkish, Russian	6-item subset of ICD-11	Adults	PTSD, anxiety, depression	n=167; female=42, male=125; mean age=34.2	Syria, Afghanistan, Iraq, Turkey
Hocking et al. (2018) (Hocking et al., 2018)	Australia	STAR-MH	Interview	Non-mental health workers and interpreters	English	Mini International Neuropsychiatric Interview (MINI), DSM-IV/ICD-10	Adults	PTSD, major depressive disorder	n=185; female=56, male=129; mean age=33	36 countries (countries not specified)
Hollifield et al. (2016) (Hollifield et al., 2016)	USA	Refugee Health Screener 15 (RHS-15)	Self-report via interview	Nurse	Nepali, Karen, Burmese, Arabic	Posttraumatic Symptom Scale Self Report (PSS-SR), DSM-IV	Adolescents and adults	PTSD, anxiety, depression	n=179; female=86, male=93; mean age=32.4	Burma, Bhutan, Iraq

(continued on next page)

Table 1 (continued)

Study	Study setting (country)	Screening Tool	Method of Delivery	Screening Personnel	Language of Tool	Reference Standard	Target Population	Target Condition	Sample size, gender, mean age	Country of origin
Ibrahim et al. (2018) (Ibrahim et al., 2018)	Kurdistan	Post-Traumatic Stress Disorder Checklist (PCL-5)	Self-report via interview	Interpreters trained to administer the tool	Kurdish, Arabic	Culturally informed clinical interview, DSM-V	Adults	PTSD, depression	n=98; female=49, male=49; mean age=32.9	Iraq, Syria
Jakobsen et al. (2011) (Jakobsen et al., 2011)	Norway	Harvard Trauma Questionnaire (HTQ), 16-tem	Self-report questionnaire	Interpreter	Arabic, Dari, Farsi, Bosnian, Somali	Composite International Diagnostic Interview (CIDI)	Adults	PTSD, anxiety, depression	n=64; female=30, male=34; mean age=33	Africa, Europe (Balkans), Asia (Somalia only country specified)
Jakobsen et al. (2017) (Jakobsen et al., 2017)	Norway	Harvard Trauma Questionnaire (HTQ), 16-item	Self-report questionnaire (on laptop)	Interpreter	Translated versions used but language not specified	Composite International Diagnostic Interview (CIDI)	Adolescents	PTSD, anxiety, depression	n=160; female=0, male=160; mean age=16.2	Majority from Afghanistan and Somalia (other countries not specified)
Kaltenbach et al. (2017) (Kaltenbach et al., 2017)	Germany	Refugee Health Screener 13 and 15 (RHS-13; RHS-15)	Self-report questionnaire	Clinical psychologists and interpreters	Albanian, Kurdish, Serbian	Semi-structured clinical interview using Posttraumatic Stress Disorder Checklist-5 (PCL-5)	Adults	PTSD, anxiety, depression	n=56; gender and mean age not reported for analysis sample; [For total sample of n=86: female=31, male=55; mean age=28.8]	Syria, Afghanistan, Albania, Kosovo, Serbia, Iraq, Macedonia, Georgia
Lhewa et al. (2007) (Lhewa et al., 2007)	USA	Harvard Trauma Questionnaire (HTQ)	Self-report questionnaire (assisted)	Tibetan research assistant	English, Tibetan	Diagnostic interview, DSM-IV	Adults	PTSD, anxiety, depression	n=57; female=12, male=45; mean age=34	Tibet
Lillee et al. (2015) (Lillee et al., 2015)	Australia	Kessler Psychological Distress Scale (K10)	Interview	Physicians and interpreters	English	PTSD treatment screener, administered by physician	Adults	PTSD	n=300; female=154, male=146; mean age=34	18 African and Asian countries
Mewes et al. (2016) (Mewes et al., 2016)	Germany	PROTECT Questionnaire (PQ)	Self-report questionnaire (on laptop, assisted)	Interpreters	Farsi, Arabic, Kurdish, English	Structured Clinical Interview (SCID), DSM-IV/DSM-V	Adults	PTSD, depression	n=91; female=28, male=63; mean age not reported for analysis sample [For total sample of 141, mean age=31.9]	Iran, Afghanistan, Syria, African countries (not specified)
Mollica et al. (1992) (Mollica et al., 1992)	USA	Harvard Trauma Questionnaire (HTQ)	Self-report questionnaire	N/A	Khmer, Lao, Vietnamese	Semi-structured clinical interview, DSM-III-R	Adults	PTSD, depression	n=91; female=57, male=34; mean age=43.5	Cambodia, Lao, Vietnam
Morina et al. (2013) (Morina et al., 2017)	Germany, Italy, UK	Impact of Event Scale-Revised (IES-R)	Self-report questionnaire	N/A	Translated versions used but language not specified	Mini International Neuropsychiatric Interview (MINI), DSM-IV/ICD-10	Adults	PTSD	n=796; female=406, male=390; mean age not reported for analysis sample [For total sample of 854, mean age=41.6]	Bosnia-Herzegovina, Croatia, Kosovo, Macedonia, Serbia
Nehring et al., 2021 (Nehring et al., 2021)	Germany	Child Behavior Checklist (CBCL-PTSD)	Self-report questionnaire	Bilingual physicians and interpreters	German	Children ≥ 6: Structured interview (Kinder-DIPS), DSM-IV; Children < 6: Semi-structured interview (PTSDSSI) of main caregiver	Children	PTSD	n=61; female=25, male=36; mean age=8.9	Syria
Renner et al. (2006) (Renner et al., 2006)	Austria	Harvard Trauma Questionnaire (HTQ) HTQ: 16 Item	Self-report questionnaire (assisted) Self-report questionnaire (assisted)	Researchers and Interpreters	German, English, Russian, Farsi	Interviews recorded and transcribed, then focus group of research team members diagnose participants as	Adults	PTSD	n=150; female=40, male=110; mean age=32.4 (Chechnyan), 32.5 (Afghani), 27.5 (West African)	Chechnya, Afghanistan, Cameroon, Gambia, Ghana, Guinea, Liberia, Sierra Leone

(continued on next page)

Table 1 (continued)

Study	Study setting (country)	Screening Tool	Method of Delivery	Screening Personnel	Language of Tool	Reference Standard	Target Population	Target Condition	Sample size, gender, mean age	Country of origin
		HTQ: Cultural	Self-report questionnaire (assisted)			“traumatized” or “not traumatized”				
		Impact of Event Scale-Revised (IES-R)	Self-report questionnaire (assisted)							
		Clinician Administered PTSD Scale (CAPS-1)	Interview							
Sack et al. (2008) (Sack et al., 1998)	USA	Impact of Event Scale (IES)	Self-report questionnaire (assisted)	Masters-level clinician and interpreter	English	Diagnostic interview using PTSD section of the Diagnostic Instrument for Children and Adolescents (DICA), DSM-III-R	Adolescents and Adults	PTSD	n=180; female=86, male=94; mean age=20.1	Cambodia
Smith Fawzi et al. (1997) (Smith Fawzi et al., 1997)	USA	Harvard Trauma Questionnaire (HTQ)	Self-report questionnaire	N/A	Vietnamese	Structured Clinical Interview (SCID), DSM-III-R	Adults	PTSD	N=51; gender and mean age not reported for analysis sample [For total sample of n=62; gender not reported; mean age=51]	Vietnam
Söndergaard et al. (2003) (Söndergaard et al., 2003)	Sweden	Health Leaflet (HL) Harvard Trauma Questionnaire (HTQ) Impact of Event Scale (IES-R)	Interview Self-report questionnaire Self-report questionnaire	Social worker	Not reported	Structured Clinical Interview (SCID) and Clinician-Administered PTSD Scale (CAPS)	Adults	PTSD	n=75; gender not reported; mean age not reported	Iraq
Turner et al. (2003) (Turner et al., 2003)	UK	Post-traumatic Diagnostic Scale (PDS)	Self-report questionnaire (assisted)	N/A	Albanian	Clinician-Administered PTSD Scale (CAPS), DSM-IV	Adults	PTSD, depression	n=120; female=64, male=56; mean age=37.1	Albania
Wulfes et al. (2019) (Wulfes et al., 2019)	Germany	PROTECT Questionnaire (PQ) Posttraumatic Diagnostic Scale (PDS-8)	Self-report questionnaire via interview Self-report questionnaire via interview	Psychotherapists; masters students; psychotherapy trainees	German, Arabic, Persian, Kurdish, Turkish	Structured Clinical Interview (SCID), DSM-IV	Adults	PTSD, Major Depressive Episode (MDE)	n=118; female=42, male=76; mean age=32.9	Syria, Iraq, Afghanistan, Iran, Sudan

et al., 2015; Hall et al., 2014). A variety of specialist personnel were involved in administering the screening tools. Fourteen studies reported the use of interpreters (Nehring et al., 2021; Jakobsen et al., 2011; Jakobsen et al., 2017; Renner et al., 2006; Blackmore et al., 2022; Sack et al., 1998; Kaltenbach et al., 2017; Mewes et al., 2016; Heeke et al., 2020; Ibrahim et al., 2018; Brink et al., 2016; Eytan et al., 2007; Hocking et al., 2018; Lillee et al., 2015), two used cultural mediators (Lhewa et al., 2007; Barbieri et al., 2019), and four involved bilingual research staff or volunteers (Bjarta et al., 2018; Dao et al., 2012; Hall et al., 2014). Medical professionals administered the tools in eleven studies (Nehring et al., 2021; Blackmore et al., 2022; Sack et al., 1998; Hollifield et al., 2016; Kaltenbach et al., 2017; Wulfes et al., 2019; Heeke et al., 2020; Brink et al., 2016; Dao et al., 2012; Eytan et al., 2007; Hocking et al., 2018; Barbieri et al., 2019; Lillee et al., 2015). The time to administer the screening tools ranged from 5 min to 90 min, but only six studies reported this information (Kaltenbach et al., 2017; Ahmad et al., 2000; Hocking et al., 2018; Barbieri et al., 2019; Lillee et al., 2015).

The screening tools were validated in refugee populations from 48 different countries. Four studies reported the region of origin without specifying the country. The most frequently studied refugee populations originated from Iraq ($n = 8$), Afghanistan ($n = 8$), Syria ($n = 7$), and Somalia ($n = 6$). The mean sample size was 174, ranging from 51 to 796. Most of the screening tools were designed for adults ($n = 13$). Only three studies included children and used tools designed specifically to screen for PTSD in children (Nehring et al., 2021; Ahmad et al., 2000; Hall et al., 2014). One study looked exclusively at adolescents but used a tool designed for adults (Jakobsen et al., 2017).

The screening tools were translated into 22 languages, with Arabic ($n = 7$), Farsi ($n = 5$), Kurdish ($n = 4$), Somali ($n = 3$) and Vietnamese ($n = 3$) the most frequently used. Six studies used the original English versions of the screening tools with interpreters (Sack et al., 1998; Ahmad et al., 2000; Eytan et al., 2007; Hocking et al., 2018; Barbieri et al., 2019; Lillee et al., 2015; Hall et al., 2014). In addition to translating the tools, some research teams adapted the instruments to the cultures being studied. For example, 14 questions of the Harvard Trauma Questionnaire are intended to be specific to the culture and refugee experiences of the population being screened (Mollica et al., 1992). Seven studies reported making changes to the screening tools to make them more culturally and contextually relevant (de Fouchier et al., 2012; Mollica et al., 1992; Renner et al., 2006; Smith Fawzi et al., 1997; Dao et al., 2012; Eytan et al., 2007; Hall et al., 2014).

3.3. Risk of bias and applicability

Fig. 2 presents the overall methodological quality and applicability assessments using QUADAS-2 for all 28 studies included in the review. Fig. 3 summarizes the results of the quality assessment on a study-level.

We found risk of bias of the selection of patients to be high in 5 (18%) included studies and unclear in 8 (29%) studies. The primary reason for high risk of bias was due to recruitment of a non-consecutive or non-

random sample ($n = 5$). Risk of bias due to the administration and interpretation of the index tests was unclear in 16 (57%) included studies. For all 16 studies, it was unclear whether the results of the index test were interpreted without knowledge of the reference standard results. Risk of bias associated with administration and interpretation of the reference standard was high in three (11%) studies and unclear in 11 (39%) studies. The primary reasons for high risk of bias were use of an inappropriate reference standard ($n = 2$) and interpreting the results of the reference standard unblinded to the index test results ($n = 1$). We found risk of bias about the flow and timing of index test and reference standard administration to be high in 3 (11%) studies and unclear in 3 (11%) studies. The primary reason for high risk of bias in this domain was attributed to administration of the reference standard to a subsample of the study population ($n = 3$).

In terms of applicability, one (4%) study was of high concern because the selected patients were prisoners of war, a characteristic that could make this population distinct from other refugees and asylum seekers. Two (7%) studies had unclear concerns regarding population selection due to strict inclusion criteria. Applicability of the index tests was of unclear concern for one (4%) included study and applicability of the reference standard was of unclear concern for 3 (11%) included studies.

3.4. Results of individual studies

We present the sensitivity and specificity for all studies and all reported cut-offs in Figs. 4 and 5, respectively. We stratified these findings by index test. To note, two studies did not provide sufficient data to extract contingency table data (Sondergaard et al., 2003; Bjarta et al., 2018).

3.5. Synthesis of results

Multiple studies used the Harvard Trauma Questionnaire as their screening index test, which made it possible to meta-analyze the results of these studies. A meta-analysis of sensitivity and specificity is presented in Figs. 6 and 7, respectively, for studies using the Harvard Trauma Questionnaire at the optimal cut-off reported by authors, ranging from 1.17 to 2.5. The area under the curve for this bivariate model was 84.0%, with a summary point sensitivity of 86.6% (95%CI 0.791; 0.917) and specificity of 78.9% (95%CI 0.630; 0.888) (See Fig. 8).

Six studies which used the Harvard Trauma Questionnaire analyzed the tools performance using Mollica and colleagues (1992) original cut-off of 2.5 (Mollica et al., 1992), which allowed us to meta-analyze these results. The summary point sensitivity for this model was 58.4% (95%CI 0.316; 0.810) with a specificity of 91.1% (95%CI 0.742; 0.973). The results of this meta-analysis are presented in Figs. 9 and 10.

After careful evaluation, some instruments were deemed similar enough to the Harvard Trauma Questionnaire to be meta-analyzed: the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5), the Impact of Event Scale-Revised (IES-R) and the Posttraumatic Diagnostic Scale

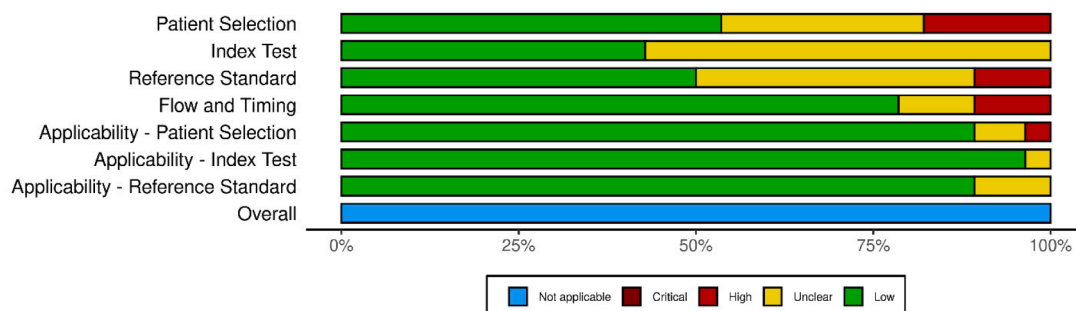


Fig. 2. Risk of bias and applicability concerns of included studies.

		Risk of bias and applicability (QUADAS-2)							
		D1	D2	D3	D4	D5	D6	D7	Overall
Study	Ahmad	+	-	+	+	+	+	+	○
	Barbieri	+	-	-	+	+	+	+	○
	Bjarta	-	-	✗	+	+	+	-	○
	Blackmore	-	+	+	+	-	+	+	○
	Brink	+	-	+	+	+	+	+	○
	Dao	-	+	+	+	+	+	+	○
	De Fouchier	-	-	-	+	+	+	+	○
	Eytan	+	+	+	+	+	+	+	○
	Hall	✗	-	-	+	-	-	-	○
	Heeke	-	-	-	+	+	+	+	○
	Hocking	+	-	+	+	+	+	+	○
	Hollifield	+	+	-	+	+	+	+	○
	Ibrahim	+	-	+	+	+	+	+	○
	Jakobsen 2011	+	-	+	-	+	+	+	○
	Jakobsen 2017	+	-	+	+	+	+	+	○
	Kaltenbach	+	+	+	+	+	+	+	○
	Lhewa	+	+	+	+	+	+	+	○
	Lillee	+	-	✗	+	+	+	+	○
	Mewes	✗	-	-	-	+	+	+	○
	Mollica	+	+	+	+	+	+	+	○
	Morina	+	+	-	✗	+	+	+	○
	Nehring	+	+	+	+	+	+	+	○
	Renner	-	+	-	+	+	+	-	○
	Sack	✗	-	-	-	+	+	+	○
Smith Fawzi	-	-	+	+	✗	+	+	○	
Sondergaard	✗	+	✗	✗	+	+	+	○	
Turner	-	+	-	✗	+	+	+	○	
Wulfes	✗	-	-	+	+	+	+	○	

D1: Patient Selection
D2: Index Test
D3: Reference Standard
D4: Flow and Timing
D5: Applicability - Patient Selection
D6: Applicability - Index Test
D7: Applicability - Reference Standard

Judgement
✗ High
- Unclear
+ Low
○ Not applicable

Fig. 3. Risk of bias and applicability concerns summary of included studies.

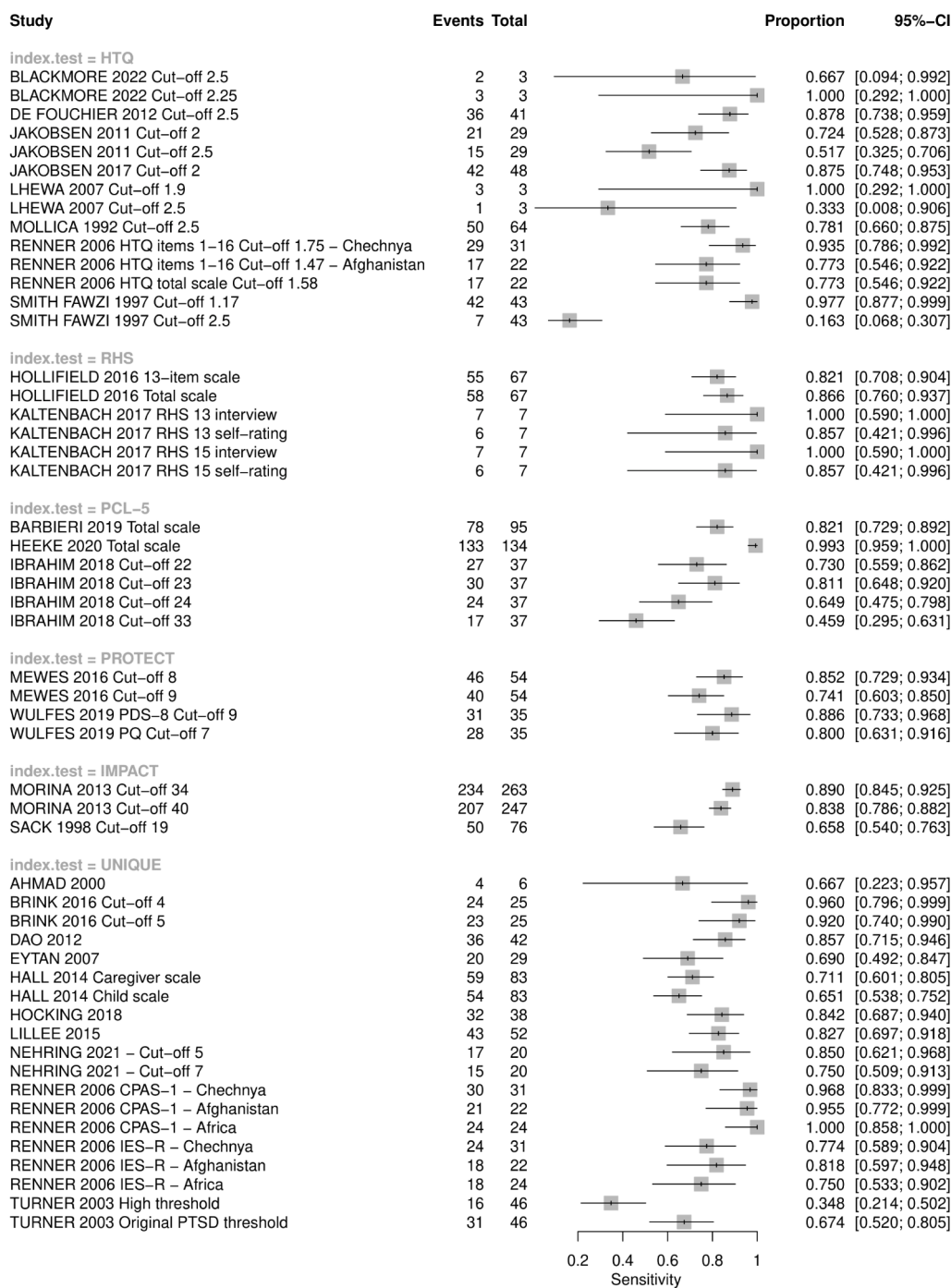


Fig. 4. Forest plot of sensitivity of index tests to detect PTSD in included studies.

(PDS). The summary ROC analyses for these 11 studies are presented in Fig. 11 (Jakobsen et al., 2011; Jakobsen et al., 2017; Lhewa et al., 2007; Renner et al., 2006; Morina et al., 2017; Sack et al., 1998; Heeke et al., 2020; Ibrahim et al., 2018; Turner et al., 2003). The area under the curve for this model was 79.4%, with a pooled sensitivity of 86.2% (95%CI 0.759; 0.925) and a specificity of 72.2% (95%CI 0.616; 0.808).

4. Discussion

The findings of our systematic review demonstrated the validity of PTSD screening tools among refugees and asylum seeker populations. However, the performance of these tools varies depending on the tool used, selected threshold for positivity, and population characteristics.

Several tools, including the Harvard Trauma Questionnaire, Refugee Health Screener, and Posttraumatic Stress Disorder Checklist appear promising given their significant testing and higher performance measures. Our meta-analyses of the Harvard Trauma Questionnaire demonstrate that the optimal cut-off for this screening tool varies based on the cultural background of the participants. Mollica and colleagues, who first developed this questionnaire, identified a cut-off of 2.5 as optimal among Indochinese refugees but cautioned that the tool would need to be validated for each individual refugee population (Mollica et al., 1992). This suggests that expression and understanding of PTSD symptoms is influenced by culture.

Many of the studies included in this review highlighted the cross-cultural challenges of using a standardized screening tool among

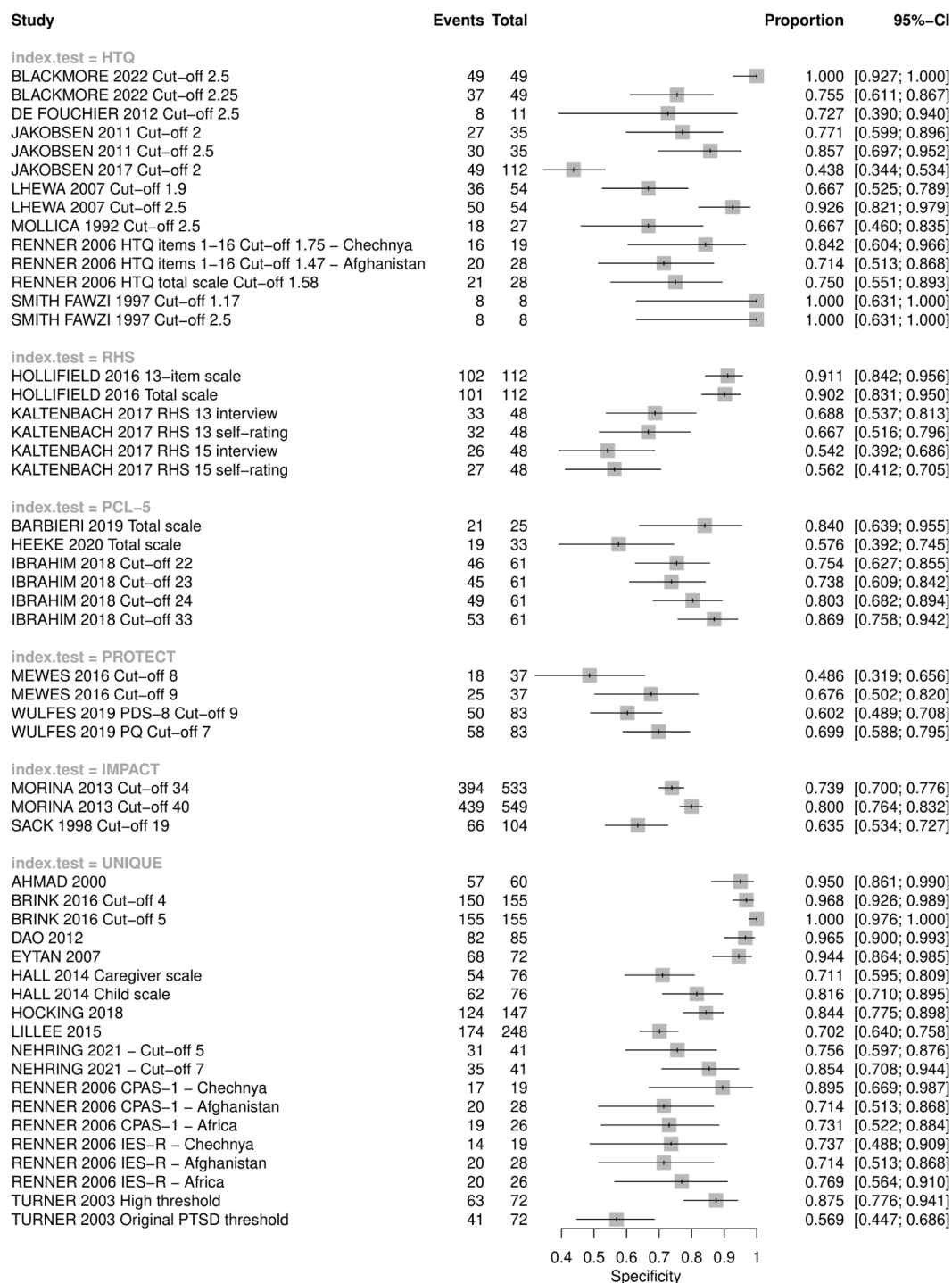


Fig. 5. Forest plot of specificity of index tests to detect PTSD in included studies.

diverse refugee populations (Jakobsen et al., 2011; Mollica et al., 1992; Smith Fawzi et al., 1997; Sack et al., 1998; Hollifield et al., 2016; Mewes et al., 2016; Wulfes et al., 2019; Heeke et al., 2020; Ibrahim et al., 2018; Ahmad et al., 2000; Hocking et al., 2018; Barbieri et al., 2019). The diagnostic criteria for PTSD were originally developed for Western populations, and some argue that they have poor applicability across cultures (Summerfield, 1999). In their summary of the evidence of the cross-cultural validity of PTSD, Hinton and Lewis-Fernandez concluded that there is a core set of PTSD symptoms that apply universally (e.g. reexperiencing the event), but many symptoms and expressions of trauma vary across cultures (e.g. avoidance) (Hinton and

Lewis-Fernandez, 2011). Some studies attributed the poor performance of specific screening items to cultural differences (de Fouchier et al., 2012; Lhewa et al., 2007; Hocking et al., 2018). Other studies identified symptoms that were unique to specific cohorts of refugees, which may also apply to non-refugees from cultural groups with similar traumatic experiences (de Fouchier et al., 2012; Renner et al., 2006; Brink et al., 2016; Hocking et al., 2018; Hall et al., 2014). This suggests that, for some groups, identification of a specific traumatic syndrome exclusively experienced by a single cohort of refugees at the time of resettlement may be sufficient to refer a refugee for follow-up care without screening for universal PTSD symptoms.

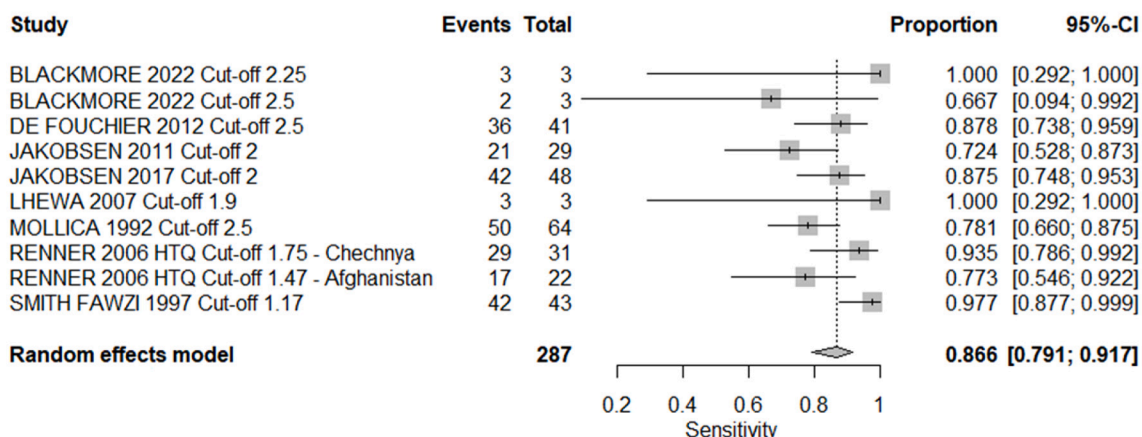


Fig. 6. Forest plot for sensitivity of studies using the Harvard Trauma Questionnaire.

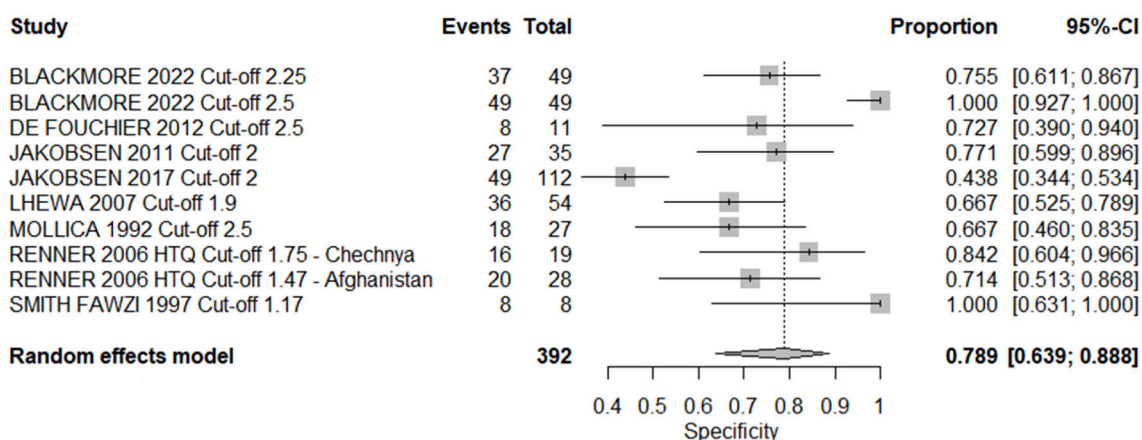


Fig. 7. Forest plot for specificity of studies using the Harvard Trauma Questionnaire.

SROC curve (bivariate model) for Diagnostic Test Accuracy

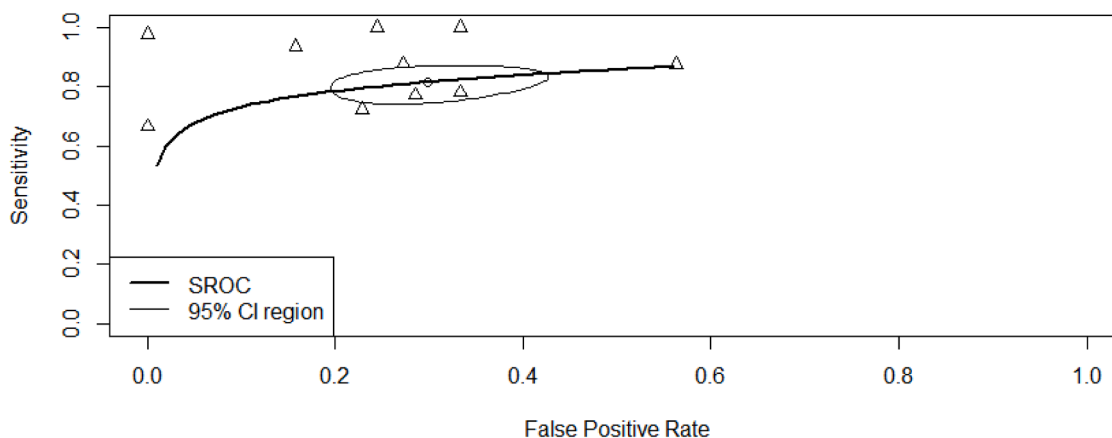


Fig. 8. Summary ROC curve for studies using the Harvard Trauma Questionnaire.

Several authors highlighted the need for cultural validation of screening tools for specific refugee populations (Jakobsen et al., 2011; Mollica et al., 1992; Renner et al., 2006; Smith Fawzi et al., 1997; Sack et al., 1998; Kaltenbach et al., 2017; Mewes et al., 2016; Wulfes et al., 2019; Ibrahim et al., 2018; Ahmad et al., 2000; Hocking et al., 2018; Barbieri et al., 2019). Hocking et al. argued that the STAR-MH may be

more cross-culturally valid due to their unique development and validation process. Their signaling questions were developed from qualitative reports of expressions of trauma found in forced migration populations, and they validated these questions with communities. However, further diagnostic test accuracy studies are needed to confirm this performance among diverse refugee groups. Cultural interpreters,

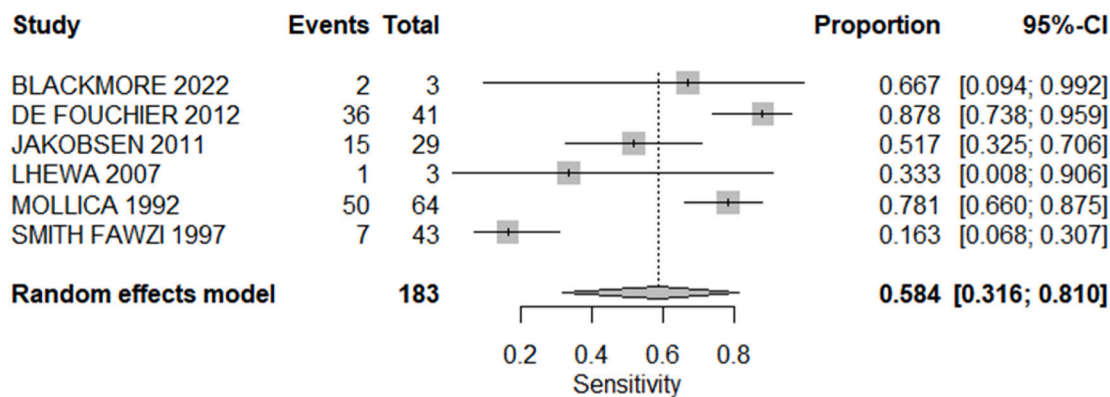


Fig. 9. Forest plot for sensitivity of studies using the Harvard Trauma Questionnaire with a cut-off of 2.5.

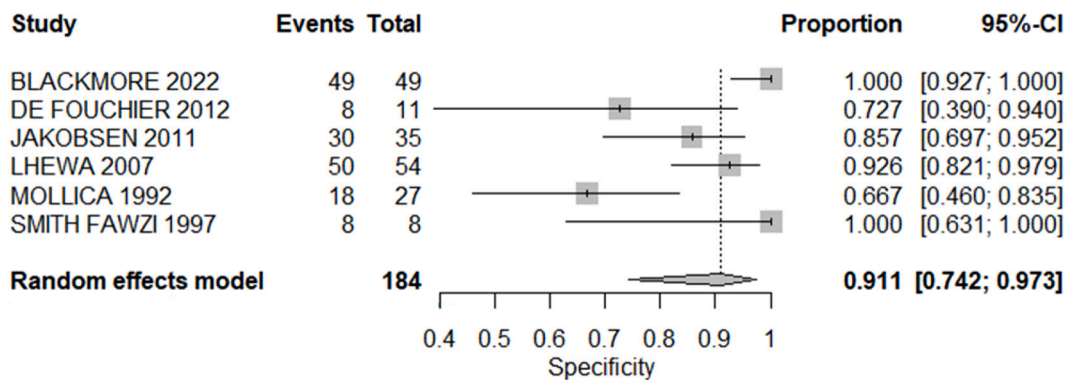


Fig. 10. Forest plot for specificity of studies using the Harvard Trauma Questionnaire with a cut-off of 2.5.

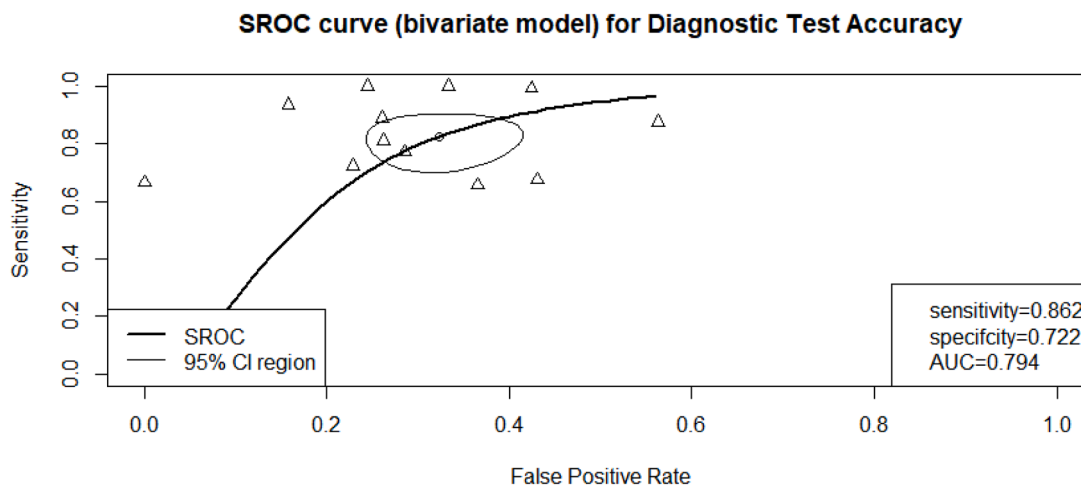


Fig. 11. Summary ROC curve for studies using the HTQ, PCL-5 or PDS.

who speak the language of the refugee population and understand the cultural manifestations of trauma, can help overcome literacy and other language barriers and may enhance the cultural validity of PTSD screening (Lhewa et al., 2007; Kaltenschach et al., 2017; Dao et al., 2012). They may also play a role in managing the stigma associated with mental illness in some cultures, a factor which may influence the performance of screening tools (Jakobsen et al., 2011; Lhewa et al., 2007; Wulfes et al., 2019; Brink et al., 2016; Lillee et al., 2015; Hall et al., 2014).

4.1. Implementation considerations

Many of the tools included in this review were not originally developed as screening tools, but rather to support clinical diagnosis of PTSD. However, stakeholders have expressed an interest in adopting one or more tools during immigration medical exams which could potentially be administered by an immigration official or panel physician (Hough et al., 2019). It is notable that many validation studies used specialized personnel to administer the screening tools, even when using self-administered questionnaires. Although several tools have been translated into the languages of refugee populations, many of these

validation studies ($n = 15$) employed interpreters or bilingual staff. In nine studies, medically-trained personnel — nurses, physicians, or psychiatrists— participated in screening. However, with so little detail provided about the administration of the screening tools, it is difficult to interpret how important the presence of these personnel is to the performance of the tools. For knowledge users looking to implement these tools, it would be useful to know if non-medical personnel can conduct screening and the level of training they should receive. Further, many refugees and asylum seekers originate or temporarily reside in low resource settings, and these are often the settings in which IMEs are conducted. Future research should consider the feasibility and appropriateness of self-administered tools to offset the screening burden placed on medical professionals. Previous work has suggested that self-assessments are possible and may be facilitated by digital technologies (e.g., laptops, tablets), but concerns regarding health literacy remain (Magwood et al., 2022).

One of the most important pieces of under-reported data is the time it takes to administer the screening tools. For decision makers looking to implement a new screening tool, administration time will have a significant impact on resources, especially considering the need for specialized personnel, such as translators or cultural interpreters. If one instrument is slightly less accurate than another but it takes substantially less time to use, the trade-off could be worthwhile. For the refugees that are being screened, there is a possible risk of re-traumatization if the screening process takes too long or if questions are invasive. De Fouchier et al. reported finding it necessary to abridge the Harvard Trauma Questionnaire because participants found it “too long and overwhelming emotionally” (de Fouchier et al., 2012). Similarly, Turner et al. suggest that the setting and timing of the administration of the screening tool can influence performance (Turner et al., 2003). There are stressors particular to life in refugee camps (e.g., poor living conditions, chronic uncertainty) and post-migration resettlement (e.g., isolation) that can exacerbate the symptoms of mental illness, and these setting-related influences should be considered when implementing a screening program.

4.2. Strengths and limitations

We conducted a diagnostic test accuracy review in accordance with a registered protocol (Magwood et al., 2021). Our review addresses a knowledge gap identified by national-level immigration officials to inform future immigration policies and programs. This is complemented by a comprehensive knowledge translation plan targeting immigration policymakers. We engaged a patient representative throughout our review to ensure the relevance and appropriateness of our review to the target population.

Several limitations of this work should be noted. First, we focused our review on PTSD among refugees and asylum seekers, but we recognize that depression, anxiety and somatic symptoms are also highly prevalent among these populations (Blackmore et al., 2020). In fact, increased trauma exposure in refugee populations is associated with comorbid mental health conditions outcomes (Im et al., 2022). Rates of co-morbidity are particularly high among refugees and contribute to negative physical, mental health, and resettlement outcomes (Im et al., 2022). High rates of mental disorder comorbidity present challenges for traditional diagnostic and treatment approaches and emphasize the importance of considering multiple common mental health symptoms rather than focusing exclusively on individual diagnostic categories. Moreover, a focus on psychiatric diagnosis alone displaces attention away from critical determinates of wellness such as resilience, suffering, meaning-making, acculturation – complexities of the displacement experience not well-captured by screening tools. If screening takes place as part of resettlement, immigration officials may desire to screen for several conditions at once. Some tools, such as the Refugee Health Screener, offer this possibility, but we did not assess the diagnostic test accuracy of their depression or anxiety components. Second, we

restricted our inclusion criteria to articles published in English, French, Spanish and Arabic. Our search did capture some studies published in German which were excluded from our review. Given that Germany is a major country of resettlement, it is possible that we excluded relevant studies that may have contributed to the evidence base. Additionally, we assessed validity by analyzing sensitivity and specificity, and recognize that these values are based on screening tool cut-offs. Previous studies have assessed validity similarly but also conducted an additional cluster analysis, which could be considered for future research (Gadeberg et al., 2017). Finally, our review may have been strengthened by further exploring important characteristics of the screening tools that would further inform implementation, such as timing of screening, mode of administration, or professional background of the administrator.

Inconsistent reporting of administration details and of individual study results precluded us from conducting subgroup analyses which could have further explored factors that influenced the performance of these screening tools. We had initially planned to conduct several subgroup analyses, but the small number of homogenous studies precluded us from conducting these additional analyses. Obtaining data to conduct our analyses was at times challenging, as many studies did not clearly report necessary values, which often had to be calculated based on the performance metrics and the prevalence reported in each study. Similarly, we could not conduct our planned subgroup analyses by age or gender because the performance metrics of the studies, including sensitivity and specificity, are rarely broken down by these characteristics. In some studies, no information on gender was provided at all. Reporting guidelines specific to mental health screening tools would help researchers conducting future validation studies to report on the information needed for appraisal and synthesis. Ongoing initiatives to develop an equity extension of the widely used Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline may help in this regard (Rizvi et al., 2022).

5. Conclusions

The findings of this review indicate that no single tool distinguished itself as the ‘best’ tool that can screen for PTSD in all refugee and asylum seeker populations. However, there are many screening tools available that perform well and are validated for several populations. These tools are ready for implementation among select groups of refugees for whom validated thresholds for positivity have been identified. For example, the Harvard Trauma Questionnaire. Although it has only been validated in three studies and does not appear to be as accurate as some other tools, the Refugee Health Screener holds promise as a practical instrument for use during immigration medical examinations. Its advantage over other tools is that it screens for PTSD, anxiety, and depression, the three most prevalent mental health conditions among refugees and asylum seekers. It is designed to be applied to all populations, but more studies are needed to evaluate its validity across cultures. This systematic review has only examined the sensitivity and specificity of PTSD screening tools. Future systematic reviews could look at other performance metrics, or the accuracy of these tools in detecting other mental health conditions, such as anxiety and depression. Sensitivity and specificity are only one aspect of an effective screening tool. More research is needed into the cost-effectiveness of these tools, their implementation requirements, and potential harms, such as re-traumatization. Importantly, an effective screening tool is only valuable if there are culturally appropriate mental health services available to those who require linkage to care.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Lindsay Sikora, MIS (University of Ottawa) for her review of the search strategies.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jmh.2022.100144.

References

- American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association, p. 443. [https://www.appi.org/Products/DSM-Library/Desk-Reference-to-the-Diagnostic-Criteria-From-\(1\)](https://www.appi.org/Products/DSM-Library/Desk-Reference-to-the-Diagnostic-Criteria-From-(1)).
- Ahmad, A., Sundelin-Wahlsten, V., Sofi, M.A., Qahar, J.A., von Knorring, A.L., 2000. Reliability and validity of a child-specific cross-cultural instrument for assessing posttraumatic stress disorder. *Eur. Child Adolesc. Psychiatry* 9 (4), 285–294. Dec.
- Barbieri, A., Visco-Comandini, F., Alunni Fegatelli, D., Schepisi, C., Russo, V., Calò, F., et al., 2019. Complex trauma, PTSD and complex PTSD in African refugees. *Eur. J. Psychotraumatol.* 10 (1), 1700621. Dec 31.
- Bjarta, A., Leiler, A., Ekdahl, J., Wasteson, E., 2018. Assessing severity of psychological distress among refugees with the refugee health screener, 13-item version. *J. Nerv. Ment. Dis.* 206 (11), 834–839. Nov.
- Blackmore, R., Boyle, J.A., Fazel, M., Ranasinha, S., Gray, K.M., Fitzgerald, G., et al., 2020. The prevalence of mental illness in refugees and asylum seekers: a systematic review and meta-analysis. *PLoS Med.* 17 (9), e1003337. Sep 21.
- Blackmore, R., Gray, K.M., Melvin, G.A., Newman, L., Boyle, J.A., Gibson-Helm, M., 2022. Identifying post-traumatic stress disorder in women of refugee background at a public antenatal clinic. *Arch. Womens Ment. Health* 25 (1), 191–198. Feb 1.
- Bogic, M., Njoku, A., Priebe, S., 2015. Long-term mental health of war-refugees: a systematic literature review. *BMC Int. Health Hum. Rights* 15 (1), 29. Oct 28.
- Brink, D.R., Shannon, P.J., Vinson, G.A., 2016. Validation of a brief mental health screener for Karen refugees in primary care. *FAMPRJ* 33 (1), 107–111. Feb.
- Crumlish, N., O'Rourke, K., 2010. A systematic review of treatments for post-traumatic stress disorder among refugees and asylum-seekers. *J. Nerv. Ment. Dis.* 198 (4), 237–251. Apr.
- Dao, T.K., Poritz, J.M., Moody, R.P., Szeto, K., 2012. Development, reliability, and validity of the Posttraumatic Stress Disorder Interview for Vietnamese refugees: a diagnostic instrument for Vietnamese refugees. *J. Trauma Stress* 25 (4), 440–445. Aug.
- Davidson, G.R., Murray, K.E., Schweitzer, R.D., 2010. Review of refugee mental health assessment: best practices and recommendations. *J. Pac. Rim Psychol.* 4 (1), 72–85. May.
- de Fouchier, C., Blanchet, A., Hopkins, W., Bui, E., Ait-Aoudia, M., Jehel, L., 2012. Validation of a French adaptation of the Harvard Trauma Questionnaire among torture survivors from sub-Saharan African countries. *Eur. J. Psychotraumatol.* 3 (1), 19225.
- Due, C., Green, E., Ziersch, A., 2020. Psychological trauma and access to primary healthcare for people from refugee and asylum-seeker backgrounds: a mixed methods systematic review. *Int. J. Ment. Health Syst.* 14 (1), 71. Sep 11.
- Eytan, A., Durieux-Paillard, S., Whitaker-Clinch, B., Loutan, L., Bovier, P.A., 2007. Transcultural validity of a structured diagnostic interview to screen for major depression and posttraumatic stress disorder among refugees. *J. Nerv. Ment. Dis.* 195 (9), 723–728. Sep.
- Gadeberg, A.K., Montgomery, E., Frederiksen, H.W., Norredam, M., 2017. Assessing trauma and mental health in refugee children and youth: a systematic review of validated screening and measurement tools. *Eur. J. Public Health* 27 (3), 439–446. Jun 1.
- Hall, B.J., Puffer, E., Murray, L.K., Ismael, A., Bass, J.K., Sim, A., et al., 2014. The importance of establishing reliability and validity of assessment instruments for mental health problems: an example from Somali children and adolescents living in three refugee camps in Ethiopia. *Psychol. Inj. Law* 7 (2), 153–164. Jun 1.
- Heeke, C., O'Donald, A., Stammel, N., Botche, M., 2020. Same same but different? DSM-5 versus ICD-11 PTSD among traumatized refugees in Germany. *J. Psychosom. Res.* 134, 110129. Jul.
- Henkelmann, J.R., de Best, S., Deckers, C., Jensen, K., Shahab, M., Elzinga, B., et al., 2020. Anxiety, depression and post-traumatic stress disorder in refugees resettling in high-income countries: systematic review and meta-analysis. *BJPsych Open* 6 (4), e68. Jul 2.
- Hinton, D.E., Lewis-Fernandez, R., 2011. The cross-cultural validity of posttraumatic stress disorder: implications for DSM-5. *Depress. Anxiety* 28 (9), 783–801. Sep.
- Hocking, D.C., Mancuso, S.G., Sundram, S., 2018. Development and validation of a mental health screening tool for asylum-seekers and refugees: the STAR-MH. *BMC Psychiatry* 18 (1), 69. Dec.
- Hollifield, M., Toolson, E.C., Verbillis-Kolp, S., Farmer, B., Yamazaki, J., Woldehaimanot, T., et al., 2016. Effective screening for emotional distress in refugees: the refugee health screener. *J. Nerv. Ment. Dis.* 204 (4), 247–253. Apr.
- Hollifield, M., Warner, T.D., Lian, N., Krakow, B., Jenkins, J.H., Kesler, J., et al., 2020. Measuring trauma and health status in refugees: a critical review. *JAMA* 288 (5), 611–621. Aug 7.
- Hough, C., O'Neill, E., Dyer, F., Beaney, K., & Crawshaw, A. (2019). The Global Mental Health Assessment Tool (GMHAT) pilot evaluation: Final report: Home Office Research report 108. *Public Health England*. <https://www.gov.uk/government/publications/the-global-mental-health-assessment-tool-gmhat-pilot-evaluation-final-report>.
- Ibrahim, H., Ertl, V., Catani, C., Ismail, A.A., Neuner, F., 2018. The validity of Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5) as screening instrument with Kurdish and Arab displaced populations living in the Kurdistan region of Iraq. *BMC Psychiatry* 18 (1), 259. Aug 16.
- Immigration New Zealand (INZ). Refugee settlement health assessments | Immigration New Zealand [Internet]. 2021 [cited 2021 Dec 16]. Available from: <https://www.immigration.govt.nz/assist-migrants-and-students/other-industry-partners/panel-ph-ysician-network/refugee-settlement-health-assessments>.
- Im, H., Swan, L.E., Warsame, A.H., Isse, M.M., 2022. Risk and protective factors for comorbidity of PTSD, depression, and anxiety among Somali refugees in Kenya. *Int. J. Soc. Psychiatry* 68 (1), 134–146. Feb 1.
- Jakobsen, M., Thoresen, S., Johansen, L., 2011. The validity of screening for post-traumatic stress disorder and other mental health problems among asylum seekers from different countries. *J. Refug. Stud.* 24 (1), 171–186.
- Jakobsen, M., Meyer DeMott, M.A., Heir, T., 2017. Validity of screening for psychiatric disorders in unaccompanied minor asylum seekers: use of computer-based assessment. *Transcult. Psychiatry* 54 (5–6), 611–625. Dec.
- Javanbakht, A., Amiradri, A., Abu Suhaiban, H., Alsaud, M.I., Alobaidi, Z., Rawi, Z., et al., 2019. Prevalence of possible mental disorders in syrian refugees resettling in the United States screened at primary care. *J. Immigr. Minority Health* 21 (3), 664–667. Jun.
- Kaltenbach, E., Hårdtner, E., Hermenau, K., Schauer, M., Elbert, T., 2017. Efficient identification of mental health problems in refugees in Germany: the Refugee Health Screener. *Eur. J. Psychotraumatol.* 8 (sup2), 1389205. Nov 24.
- Kien, C., Sommer, I., Faustmann, A., Gibson, L., Schneider, M., Krzczal, E., et al., 2019. Prevalence of mental disorders in young refugees and asylum seekers in European Countries: a systematic review. *Eur. Child Adolesc. Psychiatry* 28 (10), 1295–1310. Oct.
- Lhewa, D., Banu, S., Rosenfeld, B., Keller, A., 2007. Validation of a Tibetan translation of the Hopkins Symptom Checklist 25 and the Harvard Trauma Questionnaire. *Assessment* 14 (3), 223–230. Sep.
- Lillee, A., Thambiran, A., Laugharne, J., 2015. Evaluating the mental health of recently arrived refugee adults in Western Australia. *J. Public Ment. Health* 14 (2), 56–68. Jun 15.
- Macaskill P., Gatsonis C., Deeks J.J., Harbord R.M., Takwoingi Y. Chapter 10: analysing and presenting results. In: deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration, 2010. Available from: <http://srda.cochrane.org/>.
- Magwood O., Bellai Duseault K., Fox G., McCutcheon C., Adams O., Saad A., et al. Diagnostic test accuracy of screening tools for post-traumatic stress disorder among refugees and asylum seekers: a protocol for a systematic review of diagnostic test accuracy studies. [Internet]. Ottawa, Canada: *cochrane Equity Methods*; 2021. Available from: https://methods.cochrane.org/sites/methods.cochrane.org/equity/files/uploads/for_cochrane_-_dta_sr_protocol_-_refugee_mental_health_screening_.pdf.
- Magwood, O., Kassam, A., Mavedatnia, D., Mendonca, O., Saad, A., Hasan, H., et al., 2022. Mental health screening approaches for resettling refugees and asylum seekers: a scoping review. *Int. J. Environ. Res. Public Health* 19 (6), 3549. Jan.
- Mathes, T., Pieper, D., 2019. An algorithm for the classification of study designs to assess diagnostic, prognostic and predictive test accuracy in systematic reviews. *Syst. Rev.* 8 (1), 226. Sep 3.
- McGuinness, L.A., Higgins, J.P.T., 2021. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. *Res. Synth. Methods* 12 (1), 55–61.
- McInnes, M.D.F., Moher, D., Thoms, B.D., McGrath, T.A., Bossuyt, P.M., the P.D.T.A.G., et al., 2018. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 319 (4), 388–396. Jan 23.
- Mewes, R., Friele, B., Bloemen, E., 2016. Validation of the Protect Questionnaire: a tool to detect mental health problems in Asylum seekers by non-health professionals. *Torture* 28 (2), 56–71.
- Morina, N., Ewers, S.M., Passardi, S., Schnyder, U., Knaevelsrud, C., Müller, J., et al., 2017. Mental health assessments in refugees and asylum seekers: evaluation of a tablet-assisted screening software. *Confl. Health* 11 (1), 18. Dec.
- Mollica, R.F., Caspi-Yavin, Y., Bollini, P., Truong, T., Tor, S., Lavelle, J., 1992. The Harvard Trauma Questionnaire. Validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. *J. Nerv. Ment. Dis.* 180 (2), 111–116. Feb.
- Nehring, I., Sattel, H., Al-Hallak, M., Sack, M., Henningsen, P., Mall, V., et al., 2021. The child behavior checklist as a screening instrument for PTSD in refugee children. *Children* 8 (6), 521. Jun.
- Park, S.H., 2019. Diagnostic case-control versus diagnostic cohort studies for clinical validation of artificial intelligence algorithm performance. *Radiology* 290 (1), 272–273. Jan.

- Peconga, E.K., Høgh Thøgersen, M., 2020. Post-traumatic stress disorder, depression, and anxiety in adult Syrian refugees: what do we know? *Scand. J. Public Health* 48 (7), 677–687. Nov.
- Renner, W., Salem, I., Ottomeyer, K., 2006. Cross-cultural validation of measures of traumatic symptoms in groups of asylum seekers from Chechnya, Afghanistan, and West Africa. *Soc. Behav. Personal.* 34 (9), 1101–1114.
- Rizvi, A., Lawson, D.O., Young, T., Dewidar, O., Nicholls, S., Akl, E.A., et al., 2022. Guidance relevant to the reporting of health equity in observational research: a scoping review protocol. *BMJ Open* 12 (5), e056875. May 1.
- Sack, W.H., Seeley, J.R., Him, C., Clarke, G.N., 1998. Psychometric properties of the Impact of Events Scale in traumatized Cambodian refugee youth. *Personal. Individ. Differ.* 25 (1), 57–67.
- Shim, S.R., Kim, S.J., Lee, J., 2019. Diagnostic test accuracy: application and practice using R software. *Epidemiol. Health* 41, e2019007.
- Smith Fawzi, M.C., Murphy, E., Pham, T., Lin, L., Poole, C., Mollica, R.F., 1997. The validity of screening for post-traumatic stress disorder and major depression among Vietnamese former political prisoners. *Acta Psychiatr. Scand.* 95 (2), 87–93. Feb.
- Soltan, F., Cristofalo, D., Marshall, D., Purgato, M., Taddese, H., Vanderbloemen, L., et al., 2022. Community-based interventions for improving mental health in refugee children and adolescents in high-income countries. *Cochrane Database Syst. Rev.* 5, CD013657. May 9.
- Sondergaard, H.P., Ekblad, S., Theorell, T., 2003. Screening for post-traumatic stress disorder among refugees in Stockholm. *Nord. J. Psychiatry* 57 (3), 185–189.
- Staniszewska, S., Brett, J., Simera, I., Seers, K., Mockford, C., Goodlad, S., et al., 2017. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. *Res. Involv. Engagem.* 3, 13.
- Steel, Z., Silove, D., Phan, T., Bauman, A., 2002. Long-term effect of psychological trauma on the mental health of Vietnamese refugees resettled in Australia: a population-based study. *Lancet* 360 (9339), 1056–1062. Oct 5.
- Summerfield, D., 1999. A critique of seven assumptions behind psychological trauma programmes in war-affected areas. *Soc. Sci. Med.* 48 (10), 1449–1462. May.
- Thompson, C.T., Vidgen, A., Roberts, N.P., 2018. Psychological interventions for post-traumatic stress disorder in refugees and asylum seekers: a systematic review and meta-analysis. *Clin. Psychol. Rev.* 63, 66–79. Jul.
- Turner, S.W., Bowie, C., Dunn, G., Shapo, L., Yule, W., 2003. Mental health of Kosovan Albanian refugees in the UK. *Br. J. Psychiatry* 182, 444–448. May.
- Uphoff, E., Robertson, L., Cabieses, B., Villalón, F.J., Purgato, M., Churchill, R., et al., 2020. An overview of systematic reviews on mental health promotion, prevention, and treatment of common mental disorders for refugees, asylum seekers, and internally displaced persons. *Cochrane Database Syst. Rev.* 9, CD013458. Sep 4.
- Veritas Health Innovation (updated 2019). *Covidence systematic review software*. Melbourne, Australia. Available at www.covidence.org.
- Whiting, P.F., Rutjes, A.W., Westwood, M.E., Mallett, S., 2013. Quadas- steering group. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J. Clin. Epidemiol.* 66 (10), 1093–1104. Oct.
- Whiting, P.F., Rutjes, A.W., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., et al., 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* 155 (8), 529–536. Oct 18.
- Wickramage, K., Mosca, D., 2014. Can migration health assessments become a mechanism for global public health good? *Int. J. Environ. Res. Public Health* 11 (10), 9954–9963.
- World Health Organization, 1992. *Classifications of Mental and Behavioural Disorder. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. World Health Organization, Geneva. <https://www.who.int/publications/i/item/9241544228ICD-10>.
- Wulfes, N., del Pozo, M.A., Buhr-Riehm, B., Heinrichs, N., Kröger, C., 2019. Screening for posttraumatic stress disorder in refugees: comparison of the diagnostic efficiency of two self-rating measures of posttraumatic stress disorder: screening for PTSD in refugees. *J. Trauma Stress* 32 (1), 148–155. Feb.