Measuring Trauma and Health Status in Refugees
A Critical Review

Michael Hollifield, MD
Teddy D. Warner, PhD
Nityamo Lian, DOM, (NM)
Barry Krakow, MD
Janis H. Jenkins, PhD
James Kesler, MD
Jayne Stevenson, MD
Joseph Westermeyer, MD, PhD

A Kurdish Man

Context  Refuges experience multiple traumatic events and have significant associated health problems, but data about refugee trauma and health status are often conflicting and difficult to interpret.

Objectives  To assess the characteristics of the literature on refugee trauma and health, to identify and evaluate instruments used to measure refugee trauma and health status, and to recommend improvements.

Data Sources  MEDLINE, PsychInfo, Health and PsychoSocial Instruments, CINAHL, and Cochrane Systematic Reviews (searched through OVID from the inception of each database to October 2001), and the New Mexico Refugee Project database.

Study Selection  Key terms and combination operators were applied to identify English-language publications evaluating measurement of refugee trauma and/or health status.

Data Extraction  Information extracted for each article included author; year of publication; primary focus; type (empirical, review, or descriptive); and type/name and properties of instrument(s) included. Articles were excluded from further analyses if they were review or descriptive, were not primarily about refugee health status or trauma, or were only about infectious diseases. Instruments were then evaluated according to 5 criteria (purpose, construct definition, design, developmental process, reliability and validity) as described in the published literature.

Data Synthesis  Of 394 publications identified, 183 were included for further analyses of their characteristics; 91 (49.7%) included quantitative data but did not evaluate measurement properties of instruments used in refugee research, 78 (42.6%) reported on statistical relationships between measures (presuming validity), and 14 (7.7%) were only about statistical properties of instruments. In these 183 publications, 125 different instruments were used; of these, 12 were developed in refugee research. None of these instruments fully met all 5 evaluation criteria, 3 met 4 criteria, and 5 met only 1 of the criteria. Another 8 standard instruments were designed and developed in non-refugee populations but adapted for use in refugee research; of these, 2 met all 5 criteria and 6 met 4 criteria.

Conclusions  The majority of articles about refugee trauma or health are either descriptive or include quantitative data from instruments that have limited or untested validity and reliability in refugees. Primary limitations to accurate measurement in refugee research are the lack of theoretical bases to instruments and inattention to using and reporting sound measurement principles.

JAMA. 2002;288:611-621  www.jama.com

©2002 American Medical Association. All rights reserved.
have shown that unexplained somatic symptoms are associated with low acculturation, high treatment seeking, psychiatric disorders, and self-identified medical problems, but these studies have not shown objective evidence of medical disorders. In addition to psychiatric morbidity, refugees have a high prevalence of dental, nutritional, infectious, and pediatric illness, and they have greater self-rated impairment than the general population on the Medical Outcomes Survey SF (short form)-36 or SF-20, although neither survey has been proven valid or reliable in refugees. Refugees clearly experience multiple stressful events that are associated with adverse health outcomes. Furthermore, they may have increased morbidity, decreased life expectancy, and a vulnerability to medical illness and poor health habits, as do other traumatized populations. However, data about refugee trauma and health status are often conflicting and difficult to interpret because various methods and instruments are used for data collection, analyses, and reporting. Other methodological difficulties—such as translation and cultural differences, and inadequate resources to truly assess symptoms—complicate accurate measurement.

Under the auspices of the New Mexico Refugee Project (NMRP), created to evaluate and improve the measurement of torture, trauma, and health status in refugees, a systematic literature review and 2 levels of assessment were conducted: (1) to assess the characteristics of the refugee trauma and health literature, and (2) to systematically evaluate instruments either developed in or adapted for refugee trauma and health research using 5 criteria.

**METHODS**

Two primary sources were searched: (1) online databases using OVID from the inception date of each database through the first week of October 2001; MEDLINE from 1966, PsychInfo from 1967, Health and Psychosocial Instruments from 1985, CINAHL from 1982, and Cochrane Systematic Reviews from 1991, and (2) the NMRP database of 11487 citations, developed using Endnote (ISI ResearchSoft, Carlsbad, Calif), a commercially available software package for literature archiving and searching.

The key terms used in the initial searches are shown in the Box. Combination operators were applied to the primary source citation results to identify English-language publications evaluating (1) refugee health status, (2) refugee trauma, and (3) measurement of trauma or health status in refugees. This method initially yielded 216 and 248 citations from the online and NMRP databases, respectively. Removal of duplicates within and between these databases left 181 and 135 publications, respectively (Figure). Two authors (M.H., N.L.) then reviewed all papers in the NMRP hard-copy files, which yielded an additional 78 publications, bringing the total citations found by search criteria to 394.

Data were extracted from the 394 articles (primarily by M.H.), with consensus obtained (between M.H. and N.L.) for 22 articles, using a form to record author; year of publication; whether the article was primarily about refugees, health status, and/or trauma; if it was empirical, review, or descriptive; if it evaluated statistical properties of instrument(s) (ie, reliability, validity, item analyses, factor analyses), and if so, which ones; and if it described the development or adaptation of instruments. "Adaptation of instrument" was defined as an instrument developed in nonrefugee research that was used in refugee research to adequately test at least one statistical property. The extracted data were entered into an Excel 2000 database (Microsoft Corp, Redmond, Wash) for analyses. Publications were excluded if they were reviews or were primarily theoretical or descriptive, were not primarily about refugee health or trauma, or were only about infectious diseases.

Instruments that were either developed or adapted and tested in refugee research were then evaluated according to 5 criteria described by Weathers et al: purpose, construct definition, design, developmental process, and reliability and validity. To meet these criteria, an article had to clearly state the purpose of the instrument, what construct was being defined and measured, how the instrument was designed and why, by what methods and with what rationale it was developed, and report at least one measure of validity and reliability. Others previously have reviewed general instruments for assessing trauma and health status, our evaluation focused only on instruments either developed or adapted for use in refugee research.
RESULTS

Review of Publications for Content

Of the 394 publications identified by search criteria for further review, 187 were excluded for not meeting inclusion criteria (153 were not primarily about refugees and 34 were nonempirical or were not about trauma or health status measurement) and 24 publications had insufficient data for review (Figure). Thus, 183 articles were further analyzed. Of these, 178 (97%) were about health status (61% mental health, 16% physical health, and 20% both), and 82 (45%) about trauma. Ninety-one (49.7%) articles included quantitative data but did not evaluate measurement properties. Seventy-eight (42.6%) articles reported on the relationship of one measure to another, written with the assumption that validity was determined, even if the measures used had not shown validity. Fourteen articles (7.7%) were only about statistical properties of a measure (ie, reliability, validity, item analyses, factor analyses).

Review of Instruments

In the 183 relevant articles, 125 different instruments (ie, measurement tools such as questionnaires, surveys, and interview schedules) were described as being used to measure refugee trauma and/or health status. Only 12 instruments were developed specifically in a refugee sample and had sufficient detail in at least 1 article to allow for evaluation of the measure by the 5 criteria. Forty-one (22%) of the articles used 1 of these 12 instruments. TABLE 1 shows characteristics of these 12 instruments. Four of these 12 instruments measure health status (3 of which focused strictly on mental health), 3 measure trauma, 4 measure both trauma and health status, and 1 measures quality of care provided to refugees. None of these instruments fully met all 5 evaluation criteria. Three instruments completely met 4 criteria: the Harvard Trauma Questionnaire (HTQ),

| 178 Articles Eligible (125 Measurement Instruments) |
| 183 Articles Eligible (125 Measurement Instruments) |
| 105 Instruments Excluded From This Review (Not Developed in or Adapted for Refugee Research) |
| 211 Articles Excluded From This Review (187 Articles Did Not Meet Inclusion Criteria; 24 Articles Had Insufficient Data for Review) |

5 met only 1 of the criteria completely. None of these 12 instruments are published in the literature evaluated for this review, although components have been described and the full instruments may be available from the instrument developers or authors.

An additional 8 instruments developed in nonrefugee research that have been adapted for use in refugees were identified and evaluated (TABLE 2). Two of these instruments met all 5 criteria: the Hopkins Symptom Checklist-25 (HSCL-25) and Beck Depression Inventory. The other 6 instruments met 4 of the 5 criteria, but the purpose, construct definition, design, and development criteria were essentially met in Western, nonrefugee populations, and 6 of these instruments did not meet refugee-specific validity and/or reliability criteria.

Instruments Developed in Refugee Research

Developed and Described Trauma Measures. According to this search of the literature, 4 instruments that measure trauma have been developed in refugee research and are well described in the literature. The HTQ, developed by Mollica et al, is a self-report questionnaire with 4 parts. The purpose of part 1 is to measure 17 war-related traumatic experiences. The scale was conceptualized by expert, consensus methods from clinical experience, and designed to allow respondents to check as many of 4 responses for each experience that apply to them (“did not happen,” “experienced,” “witnessed,” or “heard about”). The HTQ manual describes its development, although it is not clear how items were chosen and designed. A convenience sample of 91 Southeast Asian patients attending a psychiatric outpatient clinic were administered the HTQ, 30 of whom were readministered the test a week later. Excellent statistical properties were demonstrated: interrater reliability for all events (κ=0.93); scale test-retest reliability (r=0.89); and internal scale consistency (Cronbach α=.90). In a separate study of 30 Asian refugees, full-scale 1-week test-retest reliability was r=0.62, but test-retest reliability for individual items ranged from poor (r=0.23) to excellent (r=0.90), meaning that some items were likely to be answered differently on the 2 tests while others were more stable. The HTQ, developed by a team with extensive refugee experience, was used in 22 of the 183 studies in this review. The trauma scale is reliable in clinical samples, although some items may not be reliable. It was rationally rather than empirically developed from clinical rather than community samples, and descrip-
tion about the construct and item development is scant. Thus, the 17-event list may be incomplete or biased, limiting generalizability. For example, experiences of women, as illustrated in the work by Alloity and Bonnerjea, are not well represented. Furthermore, the design of multiple possible responses may confuse the respondent and limit reliability and therefore validity. Finally, validity and reliability of the torture item in particular have not been reported.

The 32-item Resettlement Stressor Scale (RSS), developed by Clarke et al, from their experience with Cambodian adolescents, was designed to measure stress due to resettlement. In one study with 38 adolescents, the RSS score discriminated between those who had psychiatric illness and those who did not using diagnostic interviews, and accounted for 11.7% of PTSD score variance but did not account for the depression score variance. The War Trauma Scale (WTS), also developed by Clarke et al from their clinical experience, consists of 42 items in both an interview and self-report format, measuring traumatic experiences inflicted by the Pol Pot regime. The WTS full-scale score had an adequate internal consistency (α = .74), acceptable interrater reliability (κ = 0.88), and accounted for 15.4% of PTSD score variance and 6.7% of depression score variance. Both the RSS and the WTS demonstrated modest predictive validity of psychiatric disorder, and the WTS demonstrated acceptable reliability. They were developed by experienced investigators using rational rather than empirical methods. However, it is unclear from the literature how the items for each scale were constructed, develop-

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Measurement Focus</th>
<th>Purpose Described</th>
<th>Construct Definition Described</th>
<th>Design Described</th>
<th>Development Described</th>
<th>Testing in Refugees Described</th>
<th>Instrument Available in the Published Literature†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harvard Trauma Questionnaire (parts 1 and 4), Mollica et al, 1992</td>
<td>Trauma and health status‡</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Part</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Vietnamese Depression Scale, Kinzie et al, 1992; Kinzie and Manson, 1987</td>
<td>Health status‡</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Unnamed, Bolton, 2001</td>
<td>Health status‡</td>
<td>Yes</td>
<td>Yes</td>
<td>Part</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Resettlement Stressor Scale, Clarke et al, 1993</td>
<td>Trauma</td>
<td>Yes</td>
<td>Yes</td>
<td>Part</td>
<td>Part</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>War Trauma Scale, Clarke et al, 1993</td>
<td>Trauma</td>
<td>Yes</td>
<td>Yes</td>
<td>Part</td>
<td>Part</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Unnamed, Beiser and Fleming, 1986</td>
<td>Health status‡</td>
<td>Yes</td>
<td>Yes</td>
<td>Part</td>
<td>Part</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Post Migration Living Difficulties Scale, Silove et al, 1998</td>
<td>Trauma</td>
<td>Yes</td>
<td>Part</td>
<td>Part</td>
<td>Part</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Unnamed, Ekblad et al, 1999</td>
<td>Health status and quality of life</td>
<td>Yes</td>
<td>Part</td>
<td>Part</td>
<td>Part</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Unnamed, Weine et al, 2001</td>
<td>Quality of care provided</td>
<td>Yes</td>
<td>Yes</td>
<td>Part</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Survivor of Torture Assessment Record, Van Velsen et al, 1996</td>
<td>Trauma and health status</td>
<td>Yes</td>
<td>Part</td>
<td>Part</td>
<td>Part</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Unnamed, McCloskey et al, 1995</td>
<td>Trauma and health status‡</td>
<td>Yes</td>
<td>Part</td>
<td>Part</td>
<td>Part</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Unnamed, Cunningham and Cunningham, 1997</td>
<td>Trauma and health status</td>
<td>Yes</td>
<td>Part</td>
<td>Part</td>
<td>Part</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

*Part* indicates that criterion is only partially met. †Published literature evaluated for this review. ‡Focused primarily on mental health assessment.
oped, or designed, and they were reported in only 1 article.

The Post Migration Living Difficulties Scale (PMLD), developed by Silove et al.,54 is used to assess current life stressors of asylum seekers. Each of the 23 items of this administered survey is rated on a 5-point scale from “no problem” to “a very serious problem,” with a composite score determined.54 Its construct, development, and design are only partially described. Principal component analyses yielded 5 factors accounting for 69.8% of the variance of the 23 items: refugee determination process; health, welfare, and asylum problems; family concerns; general adaptation stressors; and social and cultural isolation.54 These 5 factors were evaluated among asylum seekers, refugees, and immigrants. Asylum seekers scored higher than immigrants on all 5 factors, and higher than refugees on factors 1, 2, and 3. Refugees scored higher than immigrants on factors 2 and 3.54 Thus, the PMLD is valid in discriminating between these 3 groups, but no other validity or reliability data are published. The PMLD is an important concept measuring life experiences other than war, but its usefulness is limited because of the lack of description about its design, development, reliability and validity, and scoring.

No refugee-specific instruments that assess prewar/conflict or nonwar/conflict trauma-related experiences in refugees were found. A number of general measures of lifetime trauma have been developed but have not been adapted or used with refugees.56

**Developed and Described Health Status Measures.** From our review, 2 instruments measuring health status that have been developed in refugee research are well described in the literature. Part 4 of the self-report HTQ, developed from clinical experience by Mollica et al,7 lists 30 symptom items, 16 generated from the Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition (DSM-III-R) criteria for PTSD, and 14 which are “presumably, culture-specific symptoms associated with PTSD.” Possible responses are “not at all,” “a little,” “quite a bit,” or “extremely.” In the same convenience sample of 91 patients described earlier, internal consistency was excellent (α=.96), the symptom prevalence ranged from 44% to 92%, and the 1-week item test-retest reliability ranged from poor to excellent (r=0.32-0.85; median, .59).7 An average item score of greater than 2.5 was predictive of a PTSD diagnosis by clinical interview (78% sensitive, 65% specific). The purpose, construct definition, and design of part 4 of the HTQ is clear. Modest reliability and fair validity in diagnosing PTSD was demonstrated in clinical populations. However, in a community study the sensitivity and specificity of the “greater than 2.5” cutoff score in diagnosing PTSD was 16% and 100%, respectively, and the most efficient score for diagnosis was 1.17 (sensitivity/specificity = 98%/100%).57

The HTQ includes a limited range of possible symptoms, some are not reliable, and their ability to predict impairment has not been shown. Finally, as the authors discuss, generalizability of the HTQ and the construct validity of PTSD in general and in refugees need further study.

The VDS, a self-report questionnaire developed by Kinzie et al58 to screen Vietnamese refugees for depression, was developed using a well described rational, consensus approach from extensive clinical experience. Culturally appropriate terms were added to existing Western symptoms of depression, and designed with items on a 3-point Likert scale. After pilot testing, the final 15-item scale measures 3 symptom types: physical symptoms associated with depression in the West, Western psychological symptoms of depression, and symptoms unrelated to Western concepts. The VDS is valid in discriminating between refugee patients with depression and those with

---

**Table 2. Evaluation of Instruments Adapted for Use in Refugee Research**

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Measurement Focus</th>
<th>Purpose Described</th>
<th>Construct Definition Described</th>
<th>Design Described</th>
<th>Development Described</th>
<th>Testing in Refugees Described</th>
<th>Instrument Available in the Published Literature*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hopkins Symptom Checklist-25</td>
<td>Health status †</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Beck Depression Inventory</td>
<td>Health status †</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Impact of Events Scale</td>
<td>Health status †</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Symptoms Checklist-90</td>
<td>Health status</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Health Opinion Survey</td>
<td>Health status</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No†</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Cornell Medical Index</td>
<td>Health status</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No‡</td>
<td>No</td>
<td>No‡</td>
</tr>
<tr>
<td>Posttraumatic Symptom Scale-10</td>
<td>Trauma and health status †</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Norbeck Social Support Questionnaire</td>
<td>Health status</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Not necessarily in the literature for this review.
†Focused primarily on mental health assessment.
‡Factor analysis verified constructs.
§No formal validity or reliability testing, but the instrument demonstrated stable and nonnormative symptoms.

©2002 American Medical Association. All rights reserved.

(Reprinted) JAMA, August 7, 2002—Vol 288, No. 5 615
anxiety or schizophrenia, and a cutoff score of 13 out of a possible 34 points demonstrated 91% sensitivity and 96% specificity for diagnosing DSM-III-defined major depression in a community sample. Reliability has not been reported.

**Less-Developed Health Status Measures.** Two additional refugee health status measures have been reported, but to our knowledge have not been formally named, used in other research, or published. Bolton used 3 ethnographic qualitative methods to investigate Rwandan’s perceptions of problems following the 1994 genocidal conflict and the local validity of Western concepts, and to adapt existing measures for local use. Two of the 18 identified problems about mental health were further developed. “Guhahamuka” (mental trauma), a concept that emerged after 1994, has 36 symptom items while “agahinda” (deep sadness or grief), an older concept, consists of 16 items. Guhahamuka and agahinda include all symptoms required for DSM-IV major depression and PTSD diagnoses, which was interpreted as supporting content validity of both syndromes in Rwandans. Agahinda was 95% sensitive and 38% specific for depression measured by the published cutoff point on the HSCL-25, and its test-retest reliability was modest but acceptable ($r=0.67$). This instrument is new and its design is not well described.

Beiser and Fleming used principal component analysis to identify 4 mental health factors (panic, depression, somatization, and well-being) in Southeast Asian refugees and Euro-Canadians from interviews using 6 existing scales as sources. The final 52-item administered instrument demonstrated acceptable internal consistency ($\alpha$ for the 4 scales ranged from .72-.91) and validity by discriminating between the 23 psychiatrically ill and the 30 well respondents. Refugees and Canadians scored similarly on all 4 scales. The design and development of this instrument is not well described and reliability data were not reported.

**Underdeveloped Potential Instruments.** Five authors report work without fully discussing the construction of the actual instrument (eg, type or number of items, scoring). We include these works because the concepts and methods hold promise for refugee research. Ekblad et al used a 7-question qualitative interview to define and compare quality of life (QOL) between 14 Iranian refugees and 8 Swedes at a primary health care clinic in Sweden. The authors found that 3 of their thematic domains parallel 3 of 6 domains of the World Health Organization quality of life instrument (social relationships, level of independence, environment), demonstrating a form of validity. Iranian refugees endorsed more social concepts of quality of life than Swedes, demonstrating discriminant validity. While this measure is incomplete, it is important since QOL is understudied in refugees and is an important component of overall health and welfare. Further development of QOL measures in refugees is needed.

Weine et al report on an interview to investigate important concepts of provider (primary care professionals, social service workers, or refugee mental health professionals) knowledge, as well as attitudes toward, and service provision patterns for, Bosnian refugees with PTSD. The instrument was developed by the authors using rational, consensus methods. In their study of 30 randomly selected providers, primary care professionals had less knowledge about and provided less service to refugees with PTSD than did mental health or social service workers, demonstrating a form of discriminant validity. No further design, development, or metric properties were reported.

Van Velsen et al report the development of the Survivor of Torture Assessment Record (STAR), a semistructured clinical interview that incorporates many instruments—such as the HSCL-25—and other investigator-chosen items to determine 3 scaled scores: trauma (scored 0-7), loss of health (scored 0-9), and social losses (scored 0-6). Validity was shown by the correlation between the trauma and loss-of-health scales ($r=0.59$), and by the fact that all 3 scales distinguished between depressed and nondepressed refugees in London. Development was not described further and to our knowledge reliability has not been reported. McCloskey et al used similar methods to integrate DSM-III-R criteria for PTSD, the Child Behavior Checklist, 11 items about political violence, and 3 items about family conflict into an evaluation of trauma and health status for Mexican and Central American women and their children. The authors report quantitative data but no statistical testing of these measures. While there are significant limitations to Van Velsen’s and McCloskey’s work, each demonstrate the integration of quantitative measures into a qualitative clinical interview, which can be used to enhance the validity of measures.

Cunningham and Cunningham developed 3 checklists to gather data about symptoms, trauma, and resettlement problems from case records at a treatment program for multinational refugees in Australia. The checklists were developed from literature about symptoms and trauma in refugees, and from the authors’ experience with resettlement problems. Principal component analyses yielded 6 trauma factors and 6 symptom factors, with 2 trauma items accounting for 43% of the PTSD score variance. This research demonstrates the concept of using factor analyses to define potentially relevant trauma and health constructs for refugees, but the checklists are limited by not being empirically developed or tested for their statistical properties.

**Nonrefugee Instruments Adapted for and Tested in Refugees**

We found 8 instruments developed in nonrefugee research that had at least 1 statistical property tested in refugee research (Table 2). Two instruments, the HSCL-25 and the Beck Depression Inventory, met all 5 evaluation criteria. The HSCL-25, a self-administered questionnaire originally designed to measure change in 15 anxiety and 10 depression symptoms in psychotherapy, has been validated in the general US population and used...
in many refugee studies. The content and design on a 4-point severity scale is acceptable to Indochinese populations, and reviews in the cultural psychiatry literature consider the measure valid.\textsuperscript{71,72} An average-item score greater than 1.75 indicates “clinically significant distress.”\textsuperscript{77} Mollica et al\textsuperscript{79} tested the HSCL-25 in 3 Indochinese groups, showing excellent test-retest reliability ($r=0.89$ for total scale; $r=0.82$ for each scale), good validity in predicting diagnosed depression (88% sensitivity, 73% specificity) or the presence of any major DSM-III-R-defined Axis I disorder with either scale or the total score (93% sensitive, 76% specific). The greater than 1.75 average-item score used as a diagnostic proxy for anxiety and depression is consistent with community data in general US populations.\textsuperscript{67,68} The HSCL-25 has good reliability and validity in clinical refugee samples, but is limited to symptoms of anxiety and depression, may not be a valid indicator of the full range of symptoms in refugees, and its ability to predict impairment has not been well studied in refugees.

The Impact of Events Scale (IES)\textsuperscript{73} has been used in 8 of the 183 studies in this review. The 15-item measure has 7 intrusion and 8 avoidance items on 3-point descriptive scales measuring intrusive thoughts and body sensations and avoidance behaviors after trauma. It is valid and reliable in general populations,\textsuperscript{73} and its development is well described. The 2 scales had satisfactory internal consistency ($\alpha=.82$ and .74, respectively) and accounted for 41% of the variance in IES scores in a study of 1787 Croatian and Bosnian children,\textsuperscript{75} confirming the 2-scale construct, although individual items fit differently than in the original 20-item version of the scales. Principal component analysis suggests a third scale, named “numbing,” which requires further validation.\textsuperscript{73,75} Higher intrusion and total scores in children with more trauma events demonstrated validity, although trauma events did not predict avoidance scores. The IES scores also distinguish between 3 groups of adult refugees who have experienced torture, nontorture trauma, and migrants who have not experienced war trauma.\textsuperscript{10} Neither scale nor item test-retest reliability was reported.

The Symptom Checklist-90 (SCL-90),\textsuperscript{77,78} developed to measure change in psychological symptoms with treatment, consists of 10 scales and has been used in 4 of the 183 studies in this review. The SCL-90 depression scale was valid in differentiating depressed from nondepressed Hmong refugees who were either patients in a psychiatric clinic or who were from a community sample, and the depression scale correlated well with the Zung Depression Scale ($r=0.67$), demonstrating concurrent validity.\textsuperscript{79,80} Further, the somatization scale correlated well ($r=0.40-0.52$) with the somatic concern item of the Brief Psychiatric Rating Scale and the somatic anxiety subscale of the Hamilton Anxiety Scales.\textsuperscript{81} The depression scale of a translated Vietnamese version of the SCL-90 correlated well with the VDS ($r=0.81$).\textsuperscript{2} However, we found no reliability testing of the SCL-90 in refugee research.

Other adapted instruments were tested in single studies identified in this review. The anxiety and depression scales from the Health Opinion Survey (HOS),\textsuperscript{82} an instrument derived from the general Cornell Medical Index to measure psychophysiological symptoms, were administered at time of interview to a community sample of 2180 Southeast Asian refugees from 3 countries. A factor analysis demonstrated that anxiety and depression were common and had the same meaning for all 3 groups.\textsuperscript{83} Further analyses reported in a subsequent article demonstrated that a single factor resembling the concept of neuroasthenia accounted for 40% of the distress scores on the HOS.\textsuperscript{84} A community sample of Vietnamese refugees demonstrated high and persistent levels of physical and psychological symptoms on the Cornell Medical Index (CMI), a general health-status questionnaire, compared with normative data from the United States and Britain.\textsuperscript{85,87} Neither the HOS nor the CMI have had validity (vs a standard) or reliability tested in refugee samples. The Posttraumatic Symptom Scale-10, a 10-question survey measuring symptoms of PTSD, was found to have excellent internal consistency ($\alpha=.92$) and test-retest reliability ($r=0.89$) in Bosnian refugees, but has not been tested for validity in refugees.\textsuperscript{88} The Beck Depression Inventory\textsuperscript{89} demonstrated excellent internal consistency ($\alpha=.93$) and excellent test-retest reliability ($r=0.92$), and distinguished depressed vs nondepressed Hmong refugees against a clinician interview (94% sensitivity, 78% specificity), demonstrating validity.\textsuperscript{79} The Norbeck Social Support Questionnaire (NSSQ), which measures dimensions of support that demonstrate excellent test-retest reliability and moderate concurrent validity in Western studies,\textsuperscript{90} was adapted to study the relationship of 3 kinds of support (social network size, emotional support, esteem support) to health in Namibian refugees. The authors found that support and coping style moderated the relationship between chronic stress (years in exile) and health status (anxiety, physical symptoms, physical signs, and hospitalization in the previous year), demonstrating a form of predictive validity.\textsuperscript{91} The NSSQ showed good internal consistency ($\alpha=.83$), but no further adaptations or reliability testing have been conducted among refugees.

**COMMENT**

Half (n=91) of the 183 articles about measurement of refugee trauma and health evaluated for this review reported quantitative data but did not report evaluation of association between or statistical properties of the instruments used. Forty-three percent (n=78) reported associations between measures, assuming that validity had already been determined. In these 183 articles, 125 different instruments were used. However, only 12 instruments have been developed and tested specifically in refugee research; 3 of these met 4 of 5 evaluation criteria, but none fully met all 5 criteria recommended for a developed instru-
ment, and none have been fully published in the literature evaluated in this review. Only 41 (22%) of the 183 articles used these 12 instruments. Another 8 well-described instruments adapted for use in refugee research were identified and evaluated; 2 of these instruments met all 5 criteria, 6 met 4 criteria (but these were all designed and developed in nonrefugee populations), and these 6 have not been tested for either validity or reliability among refugees. Of the 183 articles, 19% (35) used the 8 adapted instruments.

The primary limitations to accurate measurement of trauma and health status in refugees are the lack of theory-based construct definitions to guide the development and design of instruments specifically in refugee populations and inattention to use and reporting of sound measurement principles. These shortcomings may account for conflicting data between studies—different phenomena given the same name are being used to evaluate and compare populations. Improving measurement may help to clarify events that are traumatic and predictive of poor health, enable clinicians to better diagnose and care for patients, assist public health officials to develop better prevention models, allow scientists to conduct more useful research, and provide more accurate documentation of human rights abuses.

Improving Theory and Construct Definitions
What instruments often lack is good, theory-based construct definitions that guide the design and development of measures, as shown in Table 1. For example, legal definitions that distinguish “refugee” from “asylee” from “internally displaced person” are not necessarily predictors of trauma experiences or health status.92 It might be that, for research purposes, a “refugee” or “displaced person” is best defined as a person who has fled his/her social living context because of threat to the safety or integrity of themselves or family members due to any cause (eg, war, civil conflict, disaster, oppression, or persecution that is explicitly or implicitly sanctioned by the state). Operationalizing this definition into a measure would focus on persons who have been displaced because of a threat to their safety. However, the optimal construct definition of “refugee” is an empirical question that requires further study.

Likewise, there is need for further study of what constitutes refugee trauma. No empirically developed instruments assess the complete range of trauma experiences in refugees. It is difficult to define all relevant events and types of events that influence health status, including the understudied effects of non–conflict-related events.17,54 Since trauma may precede and postdate experiences related to war and conflict, genocide, disaster, or oppression and because subjective experiences are highly variable.3,4,21,54,93,94 Thus, further community-based empirical research to better define the range and type of events that are associated with adverse health status is needed. Refugee researchers might consider adopting methodologies from life events research to better define how and what events are weighted as traumatic and predictive of poor health.107,108 The concept of “polytrauma” might be developed and used in research, since refugees experience multiple events in multiple contexts over time. This concept, applicable to other populations, is especially important for refugee research to remind investigators of the multiple events to consider as moderators of health status.

Measurement constructs of health status are better developed than are those for trauma. Extant studies evaluate medical and psychological symptoms, disorders, diseases, and impairment. However, no community-based empirically developed instruments assess the full range of symptoms in refugees, and valid illness constructs associated with impairment are underdeveloped and require further study. For example, psychiatric symptom counts are less meaningful indicators of adverse health than if their relationship to impairment is established. Only a few community prevalence studies of psychiatric disorders using diagnostic instruments have been reported,3,17,19 and these did not assess the relationship of disorders to impairment. Symptoms, disorders, and even objective evidence of disease do not necessarily imply impairment, as demonstrated in people with renal disease, panic disorder, and heart disease.116,117 Illness, on the other hand, is defined by loss of functioning, and impairs a person in a highly contextual manner and is not necessarily defined in current medical nosology.118 Research has focused on infectious diseases, PTSD, anxiety, and depression, with some research focusing on physical injury, nutrition, and preventive health. Consideration must be given to other symptom complexes that are more strongly associated with impairment.7,20,55,97,115-121 For example, it is not clear that “PTSD” is the most appropriate construct for traumatized refugees with symptoms currently defined as PTSD. Culture and language complicate diagnosis,212 and polytrauma is pathogenic for disorders that are different compared with those found in nonrefugee populations.17,18,123 The work of Bolton60 is important in this regard as it demonstrates how community-based, empirical qualitative data validate refugees’ illness experiences. However, this work has yet to demonstrate how local illness constructs are related to impairment. In addition to improving measurement about the full range of symptoms and valid illness constructs, measures of self-rated impairment for refugees are needed, since negative self-perceptions of health may predict future physical illness, mortality, and quality of life, independent of objective health status.124-127

Improvements in Measurement Principles
Design and Development. Instruments developed in community refugee populations using empirical approaches combining qualitative and quantitative methods may create measures that are more valid in representing the experiences of refugees than

References 7, 10, 17, 18, 20, 95-106.
methods where data are only obtained rationally via expert and consensus approaches. Qualitative techniques, such as in-depth interviews and focus groups, help identify the range, depth, and meaning of possible responses in a population, and allow for development of culturally informed quantitative measures. These new instruments must then be validated using iterative statistical and field testing methods. Further, culturally informed quantitative instruments must be designed to be linguistically and visually acceptable and understandable to various refugee groups.

Testing Statistical Properties: Validity and Reliability. There are many kinds of validity and reliability that must be demonstrated for a measure to be accurate across groups. For example, the HTQ reports the validity of the “greater than 2.5” average item score in predicting PTSD, but this was in outpatient psychiatric patients, and this score was not corroborated in a community based sample. Nevertheless, this cutoff score has been used in other community samples of refugees, assuming its validity. Data from broad community sources are important for testing validity because demand characteristics may bias data from specific populations, such as clinical samples or research volunteers. For example, refugees who seek asylum may be motivated, even without awareness, to endorse high levels of trauma and symptoms to obtain refugee status or to please the interviewers. Although the HTQ and VDS were developed to be culturally appropriate, they may have limited validity because they were developed and tested in outpatient psychiatry clinics. Validity and generalizability of a measure can only be estimated by including people who express the full range of events or health status; extrapolation of clinical data to refugees in general would be a “clinician’s illusion.”

There are few studies that have established good reliability and validity in some aspects of instruments in refugee research. For example, a cutoff score of 13 on the VDS demonstrated excellent validity to DSM-III-R major depression diagnosis in a community sample, and the HSCL-25 anxiety and depression scales demonstrate excellent test-retest reliability in Southeast Asian refugees. However, proper instrument development should demonstrate internal consistency (ie, item intercorrelation), stability (ie, consistent scores over time, such as test-retest reliability), and validity of the construct (ie, correlation with a standard). Instrument validity for psychiatric disorders is difficult to establish, since psychiatric diagnostic interviews as standards may be insufficiently valid. Thus, determining what standards to use for validity testing can be a difficult methodological problem. Perhaps measures of impairment that are valid predictors of future health outcomes may be the best standards against which developed mental health measures should be tested.

Limitations
There are limitations to this review. First, electronic searches of the literature are not error free, and citations to some studies and instruments may not be included in the literature. A recent meta-analysis of psychological consequences of forced displacement in Yugoslavia that searched 2 databases that we did not use (PsycLit and PILOTS) found 95 publications about refugee mental health, 12 of which were appropriate for their analyses and 10 of which our search did not initially find. The only other review of refugee mental health that we located was published in 2000 and included only 12 studies. In addition, we review was limited to English-language publications. Second, it is difficult to extract accurate data from all publications. Some articles are difficult to obtain, some do not disclose all materials or methods used, and results are often unclear and difficult to interpret. Third, categorizing articles by type is partly subjective, and excluding articles about infectious diseases narrowed our review of instruments. Despite these limitations, our review indicates the need for improvements in the development, use, and reporting of instruments used to measure trauma and health status in refugee research.

Author Affiliations: Departments of Family and Community Medicine (Dr Hollifield) and Psychiatry (Drs Hollifield, Warner, Lian, and Stevenson), University of New Mexico Health Sciences Center, Albuquerque; Sleep and Human Health Institute, Albuquerque, NM (Dr Krakow); Departments of Anthropology and Psychiatry, Case Western Reserve University, Cleveland, Ohio (Dr Jenkins); Northern Colorado Family Practice Center, Greeley (Dr Kesler); University of Minnesota and the Minneapolis Veterans Affairs Center, Minneapolis, Minn (Dr Westermeyer). Author Contributions: Study concept and design: Hollifield, Warner, Krakow, Jenkins, Westermeyer. Acquisition of data: Hollifield, Lian, Kesler, Stevenson. Analysis and interpretation of data: Hollifield, Warner, Lian, Jenkins. Drafting of the manuscript: Hollifield, Warner, Lian, Kesler, Krakow, Jenkins, Kesler, Westermeyer. Critical revision of the manuscript for important intellectual content: Hollifield, Warner, Lian, Krakow, Jenkins, Kesler, Westermeyer. Statistical expertise: Hollifield, Warner, Jenkins. Obtained funding: Hollifield, Warner, Kesler, Krakow. Administrative, technical, or material support: Lian, Kesler, Westermeyer, Stevenson. Study supervision: Hollifield, Warner, Krakow, Jenkins, Westermeyer. Qualitative methods: Jenkins. Funding/Support: This work was supported by National Institute of Mental Health grant R01 MH 69574.

Acknowledgment: We thank James Ruiz and Valerie Eckert, MPH, for helping to assemble material for this work.

REFERENCES
9. Sileo D, Sinnerbrink I, Field A, Manicavasagar V, Steel Z. Anxiety, depression and PTSD in asylum-seekers: associations with pre-migration trauma and

©2002 American Medical Association. All rights reserved.
TRAUMA AND HEALTH STATUS IN REFUGEES


10. Thompson M, McGorry P. Psychological se-
   quelae of torture and trauma in Chilean and Salva-
   1995;29:81-86.

11. Tan ES, Krippinski J, Chiu E. Stresses Experienced
   Prior to Arrival in Australia. Sydney, Australia: Per-
   gammon Press; 1986.

12. Gong-Guy E. California Southeast Asian Mental
   Oakland, Calif: Asian Community Mental Health Ser-
   vices: California State Dept of Mental Health; 1987.

13. Mollica RF, Wysynk G, Lavelle J, Truong T, Tor
   S, Yang T. Assessing symptom change in Southeast
   Asian refugee survivors of mass violence and torture.

14. Van Velsen C, Gorst-Unsworth C, Turner S. Sur-
   vivors of torture and organized violence: demogra-
   phy and diagnosis. J Trauma Stress. 1996;9:181-
   193.

15. Weise S, Laub D. Narrative constructions of his-
   tories of trauma in testimony with Bosnian survivors

16. Cervantes RC, Salgado de Snyder VN, Padilla AM.
   Posttraumatic stress in immigrants from Central
   America and Southern Mexico. Hosp Community

17. Basoglu M, Paker M, Ozmuen E, Tasdemir O, Sa-
   hin D. Factors related to long-term traumatic stress
   responses in survivors of torture in Turkey. JAMA.

   cal effects of torture: a comparison of tortured with
   normal, Turkish, and activistic Turks in Turkey. Am J

   Cambodian refugees in New Zealand. Int J Soc Psy-

20. Allodi FA, Cowgill G. Ethical and psychiatric as-

21. Castillo R, Walitschn H, Ramirez Y, Escobar R. So-
   matization in primary care, with a focus on immi-
   grants and refugees. Am J Psychiatr. 1994;51:1232-
   1236.

22. Boyd D. Treating PTSD: Cognitive-Behavioral Strat-

23. Green BL, Lindy JD, Grace MC. Posttraumatic stress
   disorder: toward DSM-IV. J Neurol Ment Disord.

   Signs and symptoms of post-traumatic stress disor-

25. Van der Kolk BA, Pelcovitz D, Roth S, Mandel FS,
   McFarlane A, Herman JL. Dissociation, somatization,
   and affect dysregulation: the complexity of adapta-

26. Cheung P. Somatisation as a presentation in de-
   pressional and post-traumatic stress disorder among
   27:422-428.

27. Westermeyer J, Bouafyuly M, Neider J, Callies A.
   Somatization among refugees: an epidemiologic study.

28. Georgiades K. Development and validation of the Vietnam-
   ese-language depression rating scale. Am J Psychia-
   try. 1982;139:1276-1281.

29. Boswell SA, Manion SM, Viehd TL, Tolan NT, Anh
   B, Pho TN. Development and validation of a Vietnam-
   ese-language depression rating scale. Am J Psychia-
   try. 1982;139:1276-1281.

30. Bolton P. Local perceptions of the mental health

31. Bolton P. Cross-cultural validity and reliability test-
   ing of a standard psychiatric assessment instrument
   without a gold standard. J Nerv Ment Dis. 2001;189:
   238-242.

32. Beiser M, Fleming JA. Measuring psychiatric dis-
   order among Southeast Asian refugees. Psychol Med.

33. Ekblad S, Abazan A, Erikson N-G. Migration stress-
   related challenges associated with perceived quality
   of life: a qualitative analysis of Iranian refugees and Swed-
   ish populations. Transcultural Psychiatry. 1999;36:329-
   345.

34. Sartorius N, Janca A. World Health Organiza-
   tion Quality of Life Assessment Instrument. In:
   Thornicroft G, Tansella M, eds. Mental Health Out-
   comes Measures. New York, NY: Springer; 1996:
   153-177.

35. Weine SM, Kue C, Dzudza E, Razzano L, Povk-
   oric M. PTSD among Bosnian refugees: a survey of pro-
   viders’ knowledge, attitudes and service patterns. Com-

36. McCloskey LA, Fernandez-Esquerr ME, South-
   wick K, Locke C. The psychological effects of political
   and domestic violence on Central American and Mexi-
   can immigrant mothers and children. J Community Psy-

37. Achenbach T, Edelbrock C. Manual for the Child
   Behavior Checklist and Revised Child Behavior Pro-
   file. Burlington, VT: University Associates in Psychia-
   try; 1983.

38. Deogaonkar LR, Lipman RS, Rickels K, Uhlenhuth
   EH, Covi L. The Hopkins Symptom Checklist (HSCL):
   a measure of primary symptom dimensions. In: Basel
   P, ed. Modern Problems in Pharmacopsychiatry. Basel,
   Switzerland: Karger; 1974.

39. Winokur A, Winokur D, Rickels K, Cox D. Sym-
   toms of emotional distress in a family planning ser-
   vice: stability over a four-week period. Br J Psychia-
   try. 1984;144:395-399.

40. Mollica RF, Wysynk G, de-Mannefort D, Khunor
   F, Lavelle J. Indochinese versions of the Hopkins Symp-
   tom Checklist-25: a screening instrument for the psy-
   chiatric care of refugees. Am J Psychiatry. 1987;144:
   497-500.

41. Parloff M, Kelman H, Frank J. Comfort, effective-
   ness, and self-awareness as criteria of improvement
   in psychotherapy. Am J Psychiatry. 1954;111:343-
   351.

42. Butcher JN. Psychological evaluation. In: Wes-
   ternyer J, Williams C, Nguyen A, eds. Mental Health
   Services for Refugees. Washington, DC: Govern-

43. Kinzie JD, Manson SM. The use of self-rating scales
   in cross-cultural psychiatry. Hosp Community Psy-
   chiatry. 1987;38:190-196.

44. Horowitz MJ, Wilner N, Alavez N. Impact of Events
   Scale: a measure of subjective stress. Psycho-

©2002 American Medical Association. All rights reserved.
TRAUMA AND HEALTH STATUS IN REFUGEES

All rights reserved.

©2002 American Medical Association. All rights reserved.