Translated Mental Health Screening Tools in Refugee Populations
August 2016
Sara E. Julian

Addressing mental health issues in refugee populations can be extremely difficult, but with many effective treatment options available, it becomes especially important that these disorders be recognized and correctly assessed. The World Health Organization estimates the lifetime prevalence of major mental health conditions worldwide at about 12% (Kessler, *et al.*, 2011), while The Global Burden of Disease Study recently identified major depressive disorder as the second largest cause of loss in Disability Adjusted Life Years worldwide (Murray, et al., 2010). Immigrant populations may be especially vulnerable to psychiatric disorders, and refugees and asylees even more so (Jaranson, *et al,* 2007). United States PCPs may miss as many of 50% of patients suffering from major depression (Mitchell, *et al.*, 2009). However, a 2002 review by the USPSTF indicated that the use of screening tools such as the PHQ-2 and 9 can greatly increase the accurate diagnosis of depression, from 10% to 47%, on an initial encounter (Pigone, *et al.*, 2002).

Translation of an inventory for use in a refugee population can pose several problems beyond simple accuracy, and many translations have been found to be problematic. Challenges include cultural relevance, eliciting accurate answers in the face of stigma, and differences in symptom presentation and conception of disease. Finally, ideal cutoff values vary from one version to another, and clinicians are often advised to keep this in mind.

Translation must account for the cultural relevance of questions. For example, the item referring to concentration is often adapted to the more common pastimes of its target culture, such as spending time with friends, watching TV, reading the newspaper, etc. Local idioms must be accounted for, as in the Vietnamese version of the Edinburgh Postnatal Depression Scale (EPDS), where an item referring to "things getting on top of me," or feeling overwhelmed by one's responsibilities, was interpreted by Vietnamese women as referring to carrying or packing one's belongings as in a flood or natural disaster.

Cultural stigma against mental illness or the expression of distress can further complicate both the administration of these scales and their initial validation. In one of the validation studies detailed below, it is interesting to note that Vietnamese women were more likely to be scored as depressed according to their EPDS, but not in their clinical interview (Barnett, *et al.,* 1999). The authors speculate that this difference may reflect a greater stigma attached to the expression of distress aloud as opposed to in writing. Questions regarding self-harm, suicide and harming others are often the most socially sensitive.

Translations also need to account for cross-cultural differences in the manifestation of mental illness. For example, several studies exist that suggest Asian individuals with mood disorders are more likely to endorse somatic and cognitive symptoms, and less likely to endorse mood symptoms or suicidal ideation, than their Western counterparts (Kleinmann, 1977). Translations of the PHQ-9 depression scale for Sub-Saharan Africa will often add an item for irritability, based on its inclusion in the DSM-V criteria for depression. This usually increases the sensitivity of the scale (Weobong, *et al*, 2009; Adewuya, *et al*, 2006; Bhana, *et al,* 2015).

A 2016 meta-analysis by Shrestha, *et al* on the reliability and validity of translations of EPDS found only 15 studies, out of 1281 seeking to validate translations of the Edinburgh Postnatal Depression Scale, which met inclusion criteria based on study design and appropriateness of translation. The tools presented here have all gone through a similar process of translation, adaptation and validation similar to that described by Shrestha, *et al.* This process included translation by bilingual speakers, often with mental health training, review by either native or bilingual mental health professionals, discussion and approval of each item by a representative native group, and blind back-translation.

For the sake of brevity, this paper will focus on some of the better validated translations in languages most commonly encountered in the Primary Care Clinics at UVA Hospital in Charlottesville, VA. These languages include Spanish, Nepali, and Arabic. Ideal cutoff values for screening purposes have been noted.

The Hopkins Symptom Checklist and Harvard Trauma Questionnaire

The first screening tests of choice for many clinicians dealing with refugee populations are the Harvard Trauma Questionnaire (HTQ) and the Hopkins Symptom Checklist (HSCL). The HTQ was based on the HSCL, and was designed specifically for refugee populations. It was originally written, in collaboration with Southeast Asian psychiatric practitioners, in English, Laotian, Vietnamese and Cambodian. Both the HTQ and HSCL have been extensively validated in these languages, and versions have been produced in many other languages, including English, Arabic, Farsi, and Serbo-Croatian (Kleijn, *et al.*, 2001).

The HTQ has the advantage of recording what traumatic events a patient has gone through and how closely affected they were by these events. New versions are frequently in development, usually intended to better capture traumas common to specific conflicts. For example, a version was created in 2003 tailored to the events experienced by refugees from Iraq and Afghanistan during the conflicts in these countries.

While internal consistency in most versions has been shown to be excellent, sample size in these validation studies is often limited, and the tests have been shown to have only average sensitivity and specificity in comparison to assessment by psychiatric staff (e.g., Kleijn, et al., Molica, et al., 2003). It is designed to take about 50-60 min to complete, which may limit its use in a busy practice.

Screening Tools for Depression

A more practical option may be the use of more specific, shorter screening tools. For depression, the English version of the PHQ-9 is already widely used both as a screening tool and to track symptoms, and seems to have the most evidence to support its use.

The PHQ-9 consists of nine items, corresponding to the nine criteria for depression from the DSM-IV, and was originally developed from part of a larger PHQ questionnaire. It has been validated in multiple large clinical settings and is the most commonly used depression screening tool in primary care settings. It has been translated and validated into several different languages, including Portugese, Mandarin Chinese, Japanese, French, Greek, Korean and Hatian Creole (Sawaya, *et al.*, 2016); Malay, Chinese, Thai and Turkish (Azah, *et al,* 2005); and SeTswana (Hanlon, *et al.*, 2015), among others.

Spanish:
Many primary care clinicians may have had some familiarity with the Spanish version of the PHQ-9. Spanish versions have been validated in Chile, Spain, Honduras, Mexico, and among Spanish speakers in the United States in inpatient, outpatient, psychiatric, non-psychiatric and community populations (Familiar, *et al.,* 2014). One study by Familiar, *et al.* investigated the factor structure of the PHQ-9 in a large sample of female Mexican teachers. Their analysis suggested a monodimensional model of the data with strong factor loadings, at 0.71 to .90, supporting the idea that, like the English version, their Mexican Spanish version is picking up a single construct, further supporting its validity. Multiple official Spanish versions are available online (Spritzer, *et al*, 2009) and some can be found in the Appendix listed at the end of the paper below. Most recommend a cutoff score of >10.

Nepali:
In order to adapt the PHQ-9 to use in Nepal, Kohrt *et al* attempted to integrate the Nepali concepts of the heart-mind (responsible for affect and memory) and the brain-mind (responsible for cognition, rationality and social behavior). The Nepali concept of depression differs significantly from the Western concept. There is no one word in Nepali that corresponds to the term depression, while some terms that on the surface may seem to be accurate translations, such as *mannasik rog* or mental illness, and *mannasik samasya* mental problems, refer specifically to the brain-mind, and are highly stigmatizing. However, Kohrt *et al.* implemented an extensive translation and vetting process, including the removal of all English idioms, medical jargon, replacement with local idioms, and the use of representative focus groups to assess the acceptability of each item.

The study looked at scores on the Nepali PHQ-9 in comparison to results of the Composite International Diagnostic Interview in 125 primary care patients from multiple clinics in Chitwan, Nepal. Although their n was relatively small, they found relatively high reliability, with area under the curve of 0.94 and Chronbach's alpha of 0.84, and moderate internal consistency (.54 to .86).

Interestingly, they noted that endorsing a heart-mind problem was significantly associated with depression. This led them to suggest a specific algorithm for use in primary care settings. The algorithm calls for the use of a question about heart-mind problems, along with the often-added question 10 of the PHQ-9, which assesses functional impairment, in much the same way as the PHQ-2 is used in the US. Those who endorse a heart-mind problem may go on to be further screened using the PHQ-9. They also identified the terms "thinking too much" and the English word "tension" (increasingly commonly used in Nepal in recent years) as other effective and relatively non-stigmatizing idioms that may be used to assess for depression. The use of this algorithm could have reduced the number of patients tested by half and false positives by 18%, without significant effect on sensitivity, specificity or positive predictive value. The authors recommend a cutoff score of >1 for initial screening, and of >10 on the PHQ-9.

Arabic:
An Arabic version of the PHQ-9 has also been created and validated. Although several studies have looked at the internal consistency of this version, a 2016 study by Samaya *et al*. also describes measures of validity. This study compared PHQ-9 scores with those of the Psychiatric Diagnosis Screening Questionnaire and clinical diagnosis among 186 Arabic-speaking psychiatric out-patients in Lebanon. Their results showed good reliability (Cronbach's alpha = 0.88) and

good convergence with the PDSQ (AUC=0.7). When compared to clinical diagnosis, the PHQ-9 did show low specificity (46%). However, they note that low PHQ-9 sensitivity has often been observed in other studies of psychiatric populations.

English:

The English version of the PHQ-9 has also been validated in populations of English speakers outside of the US, UK and Australia. One especially noteworthy study by Sung, *et al*. (2013) looked at the psychometric properties of the PHQ-9 in a group of 400 Singaporeans of diverse ethnic backgrounds, including Chinese, Malay, Indian, and "Other." They compared this to diagnosis based on the 16-item Quick Inventory of Depressive Symptomatology – Self-Report, and a structured clinical interview. The PHQ-9 was shown to have good criterion validity, with AUC of 0.80 for both major and minor depression. It had good internal consistency, with a Cronbach alpha of 0.870, and despite the culturally diverse target group, showed a significant correlation with the QIDS-SR16 (r=0.733, p<0.001). While the most common cutoff score is >9, sensitivity and specificity do vary from one version to another.

Depression Screening in Special Populations

The use of the PHQ can be problematic among pregnant and chronically ill patients because of confounding of somatic symptoms. This has led many practitioners to use other screening measures such as the Hospital Anxiety and Depression Scale (HADS) for inpatients and the severely ill, and the Edinburgh Postnatal Depression Scale (EPDS) for pregnant women.

The data validating translation of the HADS is not as robust as that for the PHQ-9. It has been translated into Spanish, French, German, Italian, Arabic (Montazeri, *et al.* 2003), Norwegian (Helvik, *et al*, 2011), Chinese (Chen, *et al.*, 2013; Wong & Yu, 2013), Nepali (Risal, *et al.*, 2015) and other languages with good validity, with Cronbach alphas of 0.78 - 0.86. Of note, most translations are not in the public domain, and the purchase of some is restricted to psychiatrists, clinical psychologists and speech and language pathologists.

Like the PHQ-9, the EPDS has also been validated in a variety of translated versions, including Spanish, Arabic, Vietnamese, Turkish, Tagalog, French, German, Chinese, and Japanese (Garcia-Esteve, *et al.*, 2003; Barnett, *et al.*, 1999; Small, *et al.*, 2007). Two meta analyses exist of EPDS translations in low- and lower-middle income countries which may be useful to clinicians. The first, by Shrestha, *et al.*, looked at the validity of EPDS translations in 14 studies covering 12 countries. A summary of these studies and their designs can be found in the Appendix. The second, a meta

analysis of 25 studies of women in nine different Sub-Saharan African countries showed a media Cronbach alpha of 0.86 (0.71-0.87). For screening purposes, these studies showed the greatest sensitivity at a cut-off score of ≥9, with a pooled sensitivity of 0.94 (95% confidence interval , 0.68-0.99) and a pooled specificity of 0.77 (95% CI, 0.59-0.88).

The results of an Australian study (Small, *et al*, 2007), comparing EPDS scores among English-, Turkish-, Vietnamese-, and Tagalog-speaking women are also worth noting. For each group the scale showed good internal validity, with Cronbach alphas of greater than 0.80. However, researchers were most surprised by the degree of similarity in the women's responses, as both inter-item and item-total correlations were quite similar. Factor analyses in this and previous studies of the EPDS support a monodimensional model of postpartum depression. These two facts again support the idea that the EPDS is measuring a single construct.

It is worth noting that while screening tools can be immensely helpful, other major factors in a correct diagnosis are therapeutic relationship, the experience of the clinician, and number of encounters (Mitchell, *et al.,* 2009). Clinicians should keep in mind the limited sensitivity and specificity of many of these tests in making clinical judgments.

References:
1. Adewuya AO, Ola BA, Afolabi OO. Validity of the patient health questionnaire (PHQ-9) as a screening tool for depression amongst Nigerian university students. J Affect Disord. 2006;96(1-2):89-93.
2. Al-Qadhi W, Ur Rahman S, Ferwana MS, Abdulmajeed IA. Adult depression screening in saudi primary care: Prevalence, instrument and cost. *BMC Psychiatry*. 2014;14:190-244X-14-190. doi: 10.1186/1471-244X-14-190 [doi].
3. Barnett, B., Matthey, S. & Gyaneshwar, R. Arch Womens Ment Health (1999) 2: 67. doi:10.1007/s007370050038.
4. Bhana A, Rathod SD, Selohilwe O, Kathree T, Petersen I. The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa. BMC Psychiatry. 2015;15:118.
5. Chen H, Bautista D, Ch'ng YC, Li W, Chan E, Rush AJ. Screening for postnatal depression in Chinese-speaking women using the Hong Kong translated version of the Edinburgh Postnatal Depression Scale. Asia Pac Psychiatry. 2013;5(2):E64-72.
6. Ghubash R, Abou-saleh MT, Daradkeh TK. The validity of the Arabic Edinburgh Postnatal Depression Scale. Soc Psychiatry Psychiatr Epidemiol. 1997;32(8):474-6.
7. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the patient health questionnaire (PHQ): A diagnostic meta-analysis. *J Gen Intern Med*. 2007;22(11):1596-1602. doi: 10.1007/s11606-007-0333-y [doi].
8. Gross R, Glasser S, Elisha D, et al. Validation of the hebrew and arabic versions of the outcome questionnaire (OQ-45). *Isr J Psychiatry Relat Sci*. 2015;52(1):33-39.
9. Hanlon C, Medhin G, Selamu M, et al. Validity of brief screening questionnaires to detect depression in primary care in Ethiopia. J Affect Disord. 2015;186:32-9.

10. Jaranson JM, Ekblad S, Kroupin GV, Eisenman DP. Mental Health and Illness in Immigrants: Epidemiology and Risk Factors. In: Immigrant Medicine, Walker PF, Barnett ED. (Eds), Elsevier, Philadelphia 2007. p.627.

11. Kleijn WC, Hovens JE, Rodenburg JJ. Posttraumatic stress symptoms in refugees: assessments with the Harvard Trauma Questionnaire and the Hopkins symptom Checklist-25 in different languages. Psychol Rep. 2001;88(2):527-32.

12. Kleinman A. Depression, somatization and the 'new cross-cultural psychiatry'.*Soc. Sci. Med.* 1977; 11:3–10.

13. Kohrt BA, Jordans MJ, Tol WA, Luitel NP, Maharjan SM, Upadhaya N. Validation of cross-cultural child mental health and psychosocial research instruments: Adapting the depression self-rating scale and child PTSD symptom scale in nepal. *BMC Psychiatry*. 2011;11(1):127-244X-11-127. doi: 10.1186/1471-244X-11-127 [doi].

14. Kohrt BA, Luitel NP, Acharya P, Jordans MJ. Detection of depression in low resource settings: validation of the Patient Health Questionnaire (PHQ-9) and cultural concepts of distress in Nepal. BMC Psychiatry. 2016;16:58.

15. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606-13.

16. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet. 2009;374(9690):609-19.

17. Montazeri A, Vahdaninia M, Ebrahimi M, Jarvandi S. The Hospital Anxiety and Depression Scale (HADS): translation and validation study of the Iranian version. Health Qual Life Outcomes. 2003;1:14.

18. Pignone MP, Gaynes BN, Rushton JL, et al. Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2002;136(10):765-76.

19. Risal A, Manandhar K, Linde M, Koju R, Steiner TJ, Holen A. Reliability and Validity of a Nepali-language Version of the Hospital Anxiety and Depression Scale (HADS). Kathmandu Univ Med J 2015;50(2):115-24

20. Sawaya H, Atoui M, Hamadeh A, Zeinoun P, Nahas Z. Adaptation and initial validation of the patient health questionnaire - 9 (PHQ-9) and the generalized anxiety disorder - 7 questionnaire (GAD-7) in an arabic speaking lebanese psychiatric outpatient sample. *Psychiatry Res*. 2016;239:245-252. doi: 10.1016/j.psychres.2016.03.030 [doi].

21. Small R, Lumley J, Yelland J, Brown S. The performance of the Edinburgh Postnatal Depression Scale in English speaking and non-English speaking populations in Australia. Soc Psychiatry Psychiatr Epidemiol. 2007;42(1):70-8.

22. Simon GE, Goldberg DP,Von Korff M, Ustun TB.Understanding cross-national differences in depression prevalence.*Psychol. Med.* 2002; 32:585–594.

23. Sulaiman, A. H., Bautista, D., Liu, C.-Y., Udomratn, P., Bae, J. N., Fang, Y., Chua, H. C., Liu, S.-I., George, T., Chan, E., Tian-mei, S., Hong, J. P., Srisurapanont, M., Rush, A. J. and the Mood Disorders Research: Asian & Australian Network (2014), Differences in psychiatric symptoms among Asian patients with depression: A multi-country cross-sectional study. Psychiatry Clin Neurosci, 68: 245–254. doi:10.1111/pcn.12118.

24. Sung SC, Low CC, Fung DS, Chan YH. Screening for major and minor depression in a multiethnic sample of Asian primary care patients: a comparison of the nine-item Patient Health Questionnaire (PHQ-9) and the 16-item Quick Inventory of Depressive Symptomatology - Self-Report (QIDS-SR16). Asia Pac Psychiatry. 2013;5(4):249-58

25. Wafki F, Amine B, Ibn Yacoub Y, et al. Depression among the moroccan systemic sclerosis. *Clin Rheumatol*. 2012;31(1):73-77. doi: 10.1007/s10067-011-1785-3 [doi].

26. Weissman MM, Bland RC,Canino GJ *et al*. Cross-national epidemiology of major depression and bipolar disorder. *JAMA*1996; 276: 293–299.

27. Weobong B, Akpalu B, Doku V, Owusu-Agyei S, Hurt L, Kirkwood B, et al. The comparative validity of screening scales for postnatal common mental disorder in Kintampo. Ghana J Affect Disord. 2009;113(1):109–17.

28. Wong VT, Yu DK. Usefulness of the Hospital Anxiety and Depression Scale for screening for psychiatric morbidity in Chinese patients with Graves' ophthalmopathy. East Asian Arch Psychiatry. 2013;23(1):6-12.

Appendix:

1. PHQ in Different Languages. Multicultural Mental Health Resource Centre. http://www.multiculturalmentalhealth.ca/clinical-tools/assessment/screening-for-common-mental-disorders/phq-in-different-languages/. Accessed 8/17/2016

2. Patient Health Questionnaire (PHQ) Screeners. Pfiser. http://www.phqscreeners.com/select-screener/36. Accessed 8/17/2016

3. Edinburgh Postnatal Depression Scale (EPDS). Perinatal Services BC. http://www.perinatalservicesbc.ca/health-professionals/professional-resources/health-promo/edinburgh-postnatal-depression-scale-(epds). Accessed 8/20/2106.

Table 3, Shrestha, *et al.*, 2016

Culturally Sensitive Translation of the Edinburgh Postnatal Depression Scale in 12 low- and lower-middle-income countries

Country and Author	Forward translation	Back translation	Resolution of differences in translations by Committee approach	Pretesting	Amendments	Test of Equivalence	Translated version
1. Nepal							
Nepal et al., (1999) [24]	Yes (psychiatrist)	Yes (Another set)	NM	Yes	Not needed	NM	Nepalese
2. India							
Patel et al., (2002) [28]	Yes	Yes	NM	NM	NM	NM	Konkani
Fernandes et al. (2010) [29]	Yes (1 health professional)	Yes (Another set: health professional)	Yes (investigator, translators & another member of the study team)	Yes	NM	NM	Kannada
3. Nigeria							
Uwakwe (2003) [30]	Yes (3 nurses)	Yes (Another set: 2 medical students & layperson)	Yes (investigator&translators)	Yes	Not needed	NM	Igbo (Eastern)
Adewuya et al., (2006) [31]	Yes (Psychiatrist & linguist)	Yes (Another set: psychiatrist & linguist)	NM	NM	NM	NM	Yoruba (Western)
4. Pakistan							
Rahman et al., (2005) [39]	Yes	Yes	\Yes	NM	NM	NM	Urdu
5. Mongolia							
Pollock et al., (2006) [32]	Yes (2 medical translators)	Yes (Another set: 2 medical translators)	Yes (investigator, psychiatrist & doctor)	NM	Yes	NM	Mongolian
6. Bangladesh							
Gausia et al., (2007) [19]	Yes (1 Principal investigator)	Yes (Another set: translator & 2 native English speakers)	Yes (investigator,1 psychologist, 1 psychiatrist, 1 paediatrician, 3 physicians, 2 lay persons)	Yes (Probing)	Yes	Yes	Bengali
7. Sri Lanka							
Rowel et al., (2008) [37]	Yes	Yes	NM	Yes	NM	NM	Sinhalese
8. Ethiopia							
Hanlon et al., (2008) [33]	Yes (Physicians)	Yes (Another set: physicians)	Yes (investigator,2 senior psychiatrists)	Yes (Probing)	Yes	NM	Amharic
Tesfaye et al., (2010) [35]	*Re-validation of the earlier EPDS –Amharic version (Hanlon et al., [33])*		NM	Yes (Probing)	Yes	NM	

9. Ghana							
Weobong et al., (2009) [38]	Yes (Native & UK professionals)	Yes (Native & UK professionals)	Yes (study team)	(Qualitative study)	Not needed	NM	Twi
10. Zimbabwe							
Chibanda et al., (2010) [36]	Yes (research assistant)	Yes	Yes (study team)	NM	NM	NM	Shona
11. Vietnam							
‡Fisher et al., (2004) [23]	Yes	Yes	Yes (study team)	Yes (Probing)	Yes	NM	Vietnamesea
12. Malawi							
Stewart et al., (2013) [34]	Yes (1 health professional, psychiatrist (UK), 2 social science graduates)	Yes (Another set: 1 non-mental health professional)	Yes (antenatal clinic nurses)	Yes	Yes	NM	Chichewa (bVisual Prompt card

NM not mentioned; aFisher et al. [23] revised Vietnamese version developed by Small et al. (1999) [43]; bVisual Prompt card depicting a range of happy and sad faces was used along with the Chichewa-version-EPDS