

Cultural adaptation of mental health measures: improving the quality of clinical practice and research

KAMALDEEP BHUI, SALAAD MOHAMUD, NASIR WARFA, THOMAS J. CRAIG and STEPHEN A. STANSFELD

The need for accurate information about the mental health problems of multicultural communities requires valid measures of mental health for use in a number of languages and cultural contexts. Measures of psychopathological symptoms leading to a diagnosis have been especially criticised for their universal application, without attention to their limitations across cultures. Yet, measures are crucial to assess recovery and the performance of services, and to take account of carer and user views. We summarise the main challenges in the cultural adaptation of such measures in our work with adults and adolescents of South Asian, African and Caribbean origin.

CONCEPTUAL ISSUES

Instruments are usually selected on the basis of the mental disorder of interest and psychometric properties. However, a further essential criterion is the cultural relevance and validity. It is not unusual to find translations of questionnaires in use for which information on the quality and the exact procedure of translation is uncertain and often unpublished. The preliminary selection of instruments must address the ease of translating the conceptual targets. This is a function of the conceptual proximity between the source and target languages. Although translations from French to English may be easier because of their conceptual proximity – compared with (say) English to Somali – there is still scope for idioms being unique to one language. For example, ‘feeling blue’ or ‘butterflies in the stomach’ may require alternative conceptually equivalent terms. Even where the same idiom exists, its significance and prevalence may not be the same in different cultural groups. An example of this is the expression of panic disorder among Khmer refugees: this includes dizziness, triggering catastrophic cognitions of

‘wind overload’ as an idiom of distress (Hinton *et al*, 2001). This requires that instruments that assess panic disorder must include idioms such as ‘wind overload’, which are largely specific to the target population and quite inappropriate for other cultural groups. Even where the core experience and expressions of distress are similar in two or more cultures, judgements about what constitutes a symptom of illness, and the threshold for diagnosis of psychopathology at different symptom levels, are not always identical across cultures. This may be tackled by employing interview-based measures administered by local clinicians who are, perhaps, best placed to come to judgements about psychopathological illness. Such approaches are expensive and time-consuming, and do not necessarily accommodate indigenous constructs of mental illness into professional classifications of mental disorder. Some instruments requiring ‘yes’ or ‘no’ responses, for example the Mini International Neuropsychiatric Interview (MINI; Sheehan *et al*, 1997), are easier to use because the interviewer makes less complex judgements, but this approach relies on careful translation of the stem questions. The scale of this task should not be underestimated. In our recent translation of the MINI into Somali, in addition to the forward- and back-translation work, six translators took 50 hours of negotiation and discussion to ensure the most appropriate conceptual terms were identified and that the translated questions tapped these core concepts.

Flaherty *et al* (1988) classified questionnaires according to their track record in cross-cultural settings. Three typologies emerged:

- (a) instruments with a proven cross-cultural equivalence;
- (b) instruments with established validity and reliability in the original culture but not tested in other cultures;

- (c) instruments with high face validity but untested in the country of origin.

Sartorius & Kuyken (1994) proposed a useful schema. The ethnocentric approach assumes total conceptual and psychometric equivalence, so instruments are freely used across cultures. The pragmatic approach is advocated if there is 60–75% overlap; this involves limited adaptation, on the assumption that psychometric properties are not significantly altered, and that conceptual limitations exist but are not marked. Edwards (1994) calls this the ‘derived etic’. A recent innovation to develop a Chinese version of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN; Cheng *et al*, 2001) demonstrates this approach. The SCAN was validated against psychiatric gold standards, and incorporated detailed consensus methods using an international panel of researchers. Judgements about psychopathological symptoms were made using videotaped material with carefully translated subtitles (Cheng *et al*, 2001). This study correctly identified symptoms; these are simpler to assess than diagnoses. However, where there is little conceptual overlap this derived etic approach cannot guarantee psychometric and conceptual equivalence. In this situation the ‘emic plus etic’ method is recommended: this advocates the use of the original, literally translated questionnaire alongside an indigenous instrument. Such an approach has been used with Shona speakers in Zimbabwe and among the Serer of Senegal (Beiser *et al*, 1976; Patel *et al*, 1997). Finally, if there is no conceptual overlap, then translation and validation are impossible. In our global society such a situation is unlikely to be found, except rarely among isolated peoples in remote parts of the world.

TRANSLATION ISSUES

Grammatical differences and variations in the semantic range offered by each language, and the related idioms and cultural symbols, must all be considered, as each can contribute to the latent meanings within any communication. A major challenge is to ensure literal and conceptual equivalence and comparability at multiple levels of meaning. So translators must have sufficient experience. It is crucial that they understand both languages (original and target), and know about the cultural

understanding of mental distress and disorder. Edwards (1994) and Flaherty *et al* (1988) emphasise conceptual, functional, scalar, content, technical and criterion equivalence. These can rarely all be achieved for all instruments. The only recommended convention is an iterative process of independent forward- and back-translation by bilingual translators with a reconciliation of the two original language versions (Jones *et al*, 2001). Edwards (1994) recommended choosing translators who have learnt the language of the original version as a second language, in preference to those who use the source language as their dominant language. Empirical evaluations of different approaches to translation are lacking. Researchers have used different numbers of forward- and back-translators, with no evidence to recommend an optimal number. For example, Mumford *et al* (1991) used six psychiatrists to perform the forward-translation into Urdu and six lay bilingual people to back-translate the Hospital Anxiety and Depression Scale. Greater numbers of forward-translators make the reconciliation between the different translations more difficult, but conceptual problems are more likely to be identified. Another approach is translation by an expert committee; alternatively the committee may choose the most appropriate version produced by translators and recommend modifications to maximise face and content validity.

VALIDATION OF MEASURES

Face and content validity are ensured during the translation process, but can be further strengthened by focus groups, pilot studies and consultation with community agencies. Ensuring criterion validity requires a gold standard. This may be a psychiatric gold standard with little or no cultural adaptation to recommend it. For example, there is no Somali gold standard to assess mental disorders, and the cultural, linguistic and symbolic influences in language are rich, complex and a composite of Arabic, East African, Islamic, Italian and English influences.

Etic assumptions about the universality of psychopathological disorders, instruments, thresholds, and ceiling and floor effects facilitate large international comparative studies of incidence, prevalence

KAMALDEEP BHUI, MD, SALAAD MOHAMUD, MSc, Barts & London Medical School, Queen Mary, University of London; NASIR WARFA, MSc, Barts & London Medical School, Queen Mary, University of London and GKT Medical School, King's College, London; THOMAS J. CRAIG, PhD, GKT Medical School, King's College, London; STEPHEN A. STANSFELD, PhD, Barts & London Medical School, Queen Mary, University of London

Correspondence: Dr K. Bhui, Department of Psychiatry, Barts & London, Queen Mary's School of Medicine, Medical Sciences Building, London E1 4NS, UK. Tel: 020 7882 7842; fax: 020 7882 7924; e-mail: k.s.bhui@qmul.ac.uk

(First received 27 November 2002, final revision 13 February 2003, accepted 14 March 2003)

and psychopathological profiles (Mason *et al*, 1997; Cheng *et al*, 2001). Such studies contribute to knowledge about the epidemiology of diagnostic entities, but reveal less about local cultural influences that limit, or confound, the measurement of disorders. We then learn little of indigenous, culturally constructed disorders that fall outside our classifications of psychopathological illness. Importing instruments from migrants' country of origin is a possible solution, but may reveal surprises. For example, in a recent study we hypothesised that the Amritsar Depression Inventory (ADI), an instrument developed and validated in the Punjab, would be better at detecting morbidity among Punjabi interviewees than the 12-item General Health Questionnaire (GHQ). Both instruments were administered to consecutive attenders in five general practices and validated against the Clinical Interview Schedule (Bhui *et al*, 2000). Somewhat to our surprise, the ADI was a poorer screening instrument among Punjabis, and proved no better than chance as a screening instrument among Punjabis who had migrated to the UK over 30 years ago. It was as good as the GHQ for English interviewees. One of the explanations we considered is that acculturation as a process modifies the validity of instruments, and so culture-specific instruments may be necessary only in newly migrant groups. Universal approaches, based on notions of human similarity, aver that such problems are minimal and overstated. Relativistic ideologies argue for incomparability across cultures owing to an emphasis on context, behaviour and indigenous concepts of distress rather than psychiatric constructs. In the future, more authentic interpretations of data must acknowledge the methodological trade-offs and the limitations.

CLINICAL PRACTICE

All of the conceptual issues identified above apply in clinical situations; however, there is rarely time to subject clinical information to the same analysis as outlined for research instruments. Interpreters can help where there are language difficulties, but in a clinical situation, triangulation with information from families about social function, unusual thinking and beliefs should be supplemented by a bilingual clinician's assessment. Interpreters can also be asked about empathy, grammar, unusual beliefs and any unexplained phenomena noticed (but not understood) by them. The normalising tendency of some interpreters should also be investigated before and after a consultation. What seems to confound clinical assessment through an interpreter is the lack of contextual, prosodic and cultural information. Considering this information is what constitutes complexity in the clinical assessment in clinical practice and research.

REFERENCES

- Beiser, M., Benfari, R. C., Collomb, H., *et al* (1976) Measuring psychoneurotic behavior in cross-cultural surveys. *Journal of Nervous and Mental Disease*, **163**, 10–23.
- Bhui, K., Bhugra, D. & Goldberg, D. (2000) Cross-cultural validity of the Amritsar Depression Inventory and the General Health Questionnaire amongst English and Punjabi primary care attenders. *Social Psychiatry and Psychiatric Epidemiology*, **35**, 248–254.
- Cheng, A. T. A., Tien, A. Y., Chang, C. J., *et al* (2001) Cross-cultural implementation of a Chinese version of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) in Taiwan. *British Journal of Psychiatry*, **178**, 567–572.
- Edwards, N. C. (1994) Translating written material for community health research and service delivery: guidelines to enhance the process. *Canadian Journal of Public Health*, (Jan–Feb), 67–70.

Flaherty, J. A., Gaviria, F. M., Pathak, D., et al (1988)

Developing instruments for cross-cultural psychiatric research. *Journal of Nervous and Mental Disease*, **176**, 257–263.

Hinton, D., Um, K. & Ba, P. (2001) *Kyal Goeu* ('wind overload') Part I: A cultural syndrome of orthostatic panic among Khmer refugees. *Transcultural Psychiatry*, **38**, 403–433.

Jones, P. S., Lee, J. W. & Phillips, L. R. (2001) An adaptation of Brislin's translation model for cross-cultural research. *Nursing Research*, **50**, 300–304.

Mason, P., Harrison, G., Croudace, T., et al (1997)

The predictive validity of a diagnosis of schizophrenia. A report from the International Study of Schizophrenia (IsoS) coordinated by the World Health Organization and the Department of Psychiatry, University of Nottingham. *British Journal of Psychiatry*, **170**, 321–327.

Mumford, D. B., Tareen, I. A., Bajwa, M. A., et al (1991) The translation and evaluation of an Urdu version of the Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, **83**, 81–85.

Patel, V., Simunyu, E., Gwanzura, F., et al (1997) The Shona Symptom Questionnaire: the development of an

indigenous measure of common mental disorders in Harare. *Acta Psychiatrica Scandinavica*, **95**, 469–475.

Sartorius, N. & Kuyken, W. (1994) Translation of health status instruments. In *Quality of Life Assessment: International Perspective* (eds J. Orley & W. Kuyken). Berlin: Springer Verlag.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., et al (1997) Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, **12**, 232–241.