

Validity issues in transcultural epidemiology[†]

MARK VAN OMMEREN

In a world with numerous refugees and increased concern for their well-being, governmental and non-governmental organisations are asking researchers for accurate estimates describing the extent of psychopathology in displaced populations. Although exact numbers are sought, the researcher soon learns that answers are filled with uncertainty. Turner and colleagues in this issue show that results from different assessment methods among Kosovan Albanian refugees in the UK do not agree with each other (Turner *et al*, 2003, this issue). An Albanian-speaking clinician administering diagnostic measures identified relatively low prevalence rates of post-traumatic stress disorder (PTSD) and depression compared with rates obtained from self-report measures in the same subsample. Studies of help-seeking Cambodian refugees in specialised clinics in the USA have indicated PTSD prevalence rates ranging between 22% and 92% (Abueg & Chun, 1996). Also, my colleagues and I have been confronted with quite different prevalence rates in two studies of a sample of Bhutanese refugees in Nepal (Shrestha *et al*, 1998; Van Ommeren *et al*, 2001).

Inconsistent findings in any research effort may result from random processes and non-equivalent measures, procedures, or samples, but may also be explained by problems of low validity. Problems of validity are not new to epidemiology (Dohrenwend, 1990), but are more likely to occur in transcultural epidemiology – defined here as research in which the views, concepts or measures of the investigator extend beyond the scope of one cultural unit to another (Prince, 1997).

Although crossing cultural units may be experienced as exotic or romantic, it is best to stay with good old conventional terminology to examine the effects of culture on the validity of transcultural studies.

Dimensions of validity of field research have been conceptualised by Cook & Campbell (1979) and clarified by Gliner & Morgan (2000). Table 1 presents definitions of classic types and subtypes of evidence of validity. Surprisingly, systematic and correct analysis of validity is uncommon in transcultural epidemiology. Rather, in the debate about the validity of transcultural studies, expressed opinions tend to be at polar ends – ranging from dismissing findings as socially constructed

medicalisation of social distress to presuming that epidemiological constructs, methods and findings are not affected by context.

The aim of this editorial is to generate awareness about the various ways in which context affects research validity. Such awareness may facilitate the identification and implementation of realistic and effective methods to reduce uncertainty in findings of transcultural studies.

MEASUREMENT VALIDITY AND RELIABILITY

Measurement validity and reliability (Table 1) are established in relation to the measure's intended purpose. Evidence of measurement validity and reliability cannot be assumed to generalise across populations. This lack of generalisability may be especially problematic when the original measure is translated into another language, as is common in transcultural

Table 1 Types of evidence for the validity of a research study

Type and subtype	Definition
Research validity	Extent of validity of the whole study
Measurement validity	Extent to which measure assesses what it purports to measure for a particular setting, population and purpose
Measurement reliability (e.g. test–retest, interrater and internal consistency)	Consistency of scores from the measure for a particular setting, population and purpose
Construct validity (e.g. discriminant, convergent and factorial evidence)	Extent to which a measure assesses the theoretical construct it is intended to measure
Diagnostic validity	Extent to which a category meets a consensus definition of psychiatric disorder and is distinguishable from other disorders
Content validity	Extent to which the measure's content represents the concept(s) to be measured
Criterion-related validity (i.e. predictive or concurrent evidence)	Strength of relationship with a measurable external criterion
Statistical validity	Proper use and interpretation of statistical methods and power
Internal validity (i.e. group equivalence, control of independent variables)	Extent to which a significant relationship is a causal relationship and not explicable by a third variable
External validity	Extent of generalisability to the target populations, to other populations, and across time and place
Population validity	Extent to which a sample represents the target population
Ecological validity	Extent of generalisability of findings across time and place to real life

Adapted with permission from Gliner & Morgan (2000). Main changes from Gliner & Morgan (2000) are: (a) nesting reliability under measurement validity, because conceptually reliability contributes to measurement validity, (b) separation of statistical validity from measurement issues, and (c) the addition of diagnostic validity.

[†]See pp. 444–448, this issue.

studies. Creating a culturally acceptable, comprehensible, relevant and semantically equivalent translation is difficult (Van Ommeren *et al*, 1999), making it essential to study the internal consistency and test–retest reliability of translated measures that might have changed during imperfect translations.

Construct and diagnostic validity

Construct validity is the degree to which a measure assesses the theoretical construct it has been designed for. If one assumes that diagnoses are atheoretical – as the later versions of the DSM strive to do – then trying to establish construct validity for measures of diagnoses is somewhat illogical. Avoiding this language issue, we discuss ‘diagnostic validity’, which is the extent to which a cluster of symptoms is markedly distressing or sufficiently impairing to warrant the label ‘psychiatric disorder’, and also is distinguishable from other disorders in terms of symptoms, course, clinical features, laboratory findings and findings from family studies (cf. Robins & Guze, 1970). Systems of diagnosis such as the DSM and ICD cannot be presumed to have high diagnostic validity across cultures, because there is evidence that sociocultural factors in varying degrees influence the clustering of symptoms and the extent to which symptoms are experienced as distressing (Mezzich *et al*, 1996).

Should the transcultural epidemiologist provide evidence of diagnostic validity in each research context? Researching evidence of diagnostic validity is a lengthy process. The current Western systems of disorders, DSM–IV (American Psychiatric Association, 1994) and ICD–10 (World Health Organization, 1992), have been created by numerous leading mental health researchers, who have had available more than a century of Western psychiatric and psychological literature, extensive data-sets for re-analysis, and, in the case of DSM–IV, funding for in-depth field trials. Even then, evidence of diagnostic validity is still sparse for many disorders. Accordingly, it may not always be realistic for transcultural epidemiologists to research diagnostic validity for the disorders they assess in various contexts. Nevertheless, this area of study benefits from continuous efforts to validate diagnostic categories (including the so-called ‘culture-bound’ disorders) in different contexts. The aforementioned definition of diagnostic validity suggests

that diagnostic validation is achieved through laboratory and family studies as well as through epidemiological and ethnographic studies of distress, disability, symptoms, course and clinical features.

Content and criterion-related validity

Literal translation can reduce a measure’s content validity, which is the extent to which a measure’s content represents the concept to be assessed. For example, the widely used Short Form–12 (Ware *et al*, 1996) contains the terms ‘bowling’ and ‘playing golf’ to assess physical functioning – terms that are unknown to many respondents in low-income countries. To use the Short Form–12 in such countries, locally meaningful equivalent terms must be substituted to maintain content validity.

Epidemiologists tend to focus their efforts on establishing criterion-related validity, which is the strength of relation between the measure and a measurable external criterion. The ideal external criterion is considered to be diagnosis by independent clinicians who are trained in using a semi-structured diagnostic instrument that has evidence of measurement validity and reliability (especially interrater reliability) for the local context. This poses a problem for transcultural epidemiology, because research is frequently conducted in contexts with very few mental health professionals, who may not have been trained in the use of standard semi-structured diagnostic instruments, which themselves seldom have any psychometric evidence for the local context.

Even though the aforementioned assessment standard of criterion-related validity is unlikely to occur in transcultural epidemiology, the researcher should try to gather data to test this validity. This effort is one of the strengths of the study by Turner *et al* in this issue.

INTERNAL AND EXTERNAL VALIDITY

Attempts to identify causes for differences in epidemiological findings between two sociocultural settings often have low internal validity. Internal validity refers to the degree to which a significant relationship is a causal relationship and is not explicable by a third variable. Societies can differ in so many ways that it is difficult to prove that

one variable is one of the causes of differences in epidemiological findings. Rather than finding causes for different prevalence rates across settings, it might be more realistic to compare patterns of findings across settings – see, for example, Patel *et al* (1999) and de Jong *et al* (2001).

Users of epidemiological data (such as policy-makers) need to know to what extent findings have external validity, i.e. generalisability to the target population, to other populations, and across time and place. Generalisability to the target population depends on the ability to randomly draw a representative sample from the entire population of relevant persons. The ability to do so requires the availability of reliable registers with contact information for the entire target population. However, the availability and quality of population registers vary and are likely to be poor in countries with fewer resources. Generalisability to the target population also depends on the study’s participation rate, i.e. the percentage of sampled people who are willing to participate in the study. Fortunately, participation rates appear to be much higher in research outside the industrialised world.

The extent to which findings from one cultural unit can be generalised to other populations is still open to debate. Can we generalise findings from one continent to another, or from one ethnic group to another within the same country? We still know little of the generalisability of epidemiological findings across populations. Multi-site studies are the answer. Moreover, in rapidly changing societies longitudinal studies may assess the extent to which findings generalise over time.

CONCLUSIONS

Systematically considering and addressing validity issues will reduce uncertainty in findings from transcultural epidemiological studies. The challenges inherent in addressing these issues are no reason for discouragement. Validity is a continuous construct. Perfectly valid studies tend to be unlikely in any science. A study certainly does not have to be highly valid in every regard to be valuable or useful. Yet, a sustained focus on validity issues – as has been demonstrated in the USA (Narrow *et al*, 2002) – will guide researchers to more-exact and useful epidemiological estimates.

DECLARATION OF INTEREST

None.

ACKNOWLEDGEMENTS

This paper has benefited from comments by Rob Baltussen, Etzel Cardeña, Daniel Chisholm, Laurence Kirmayer, Joop de Jong, George Morgan, Michael Spittel and Jos Van Ommeren.

REFERENCES

- Abueg, F. R. & Chun, K. M. (1996)** Traumatization stress among Asians and Asian Americans. In *Ethnocultural Aspects of Posttraumatic Stress Disorder: Issues, Research, and Clinical Applications* (eds A. J. Marsella, M. J. Friedman, E. T. Gerrity, et al). Washington, DC: American Psychological Association.
- American Psychiatric Association (1994)** *Diagnostic and Statistical Manual of Mental Disorders (4th edn)* (DSM-IV). Washington, DC: APA.
- Cook, T. D. & Campbell, D. T. (1979)** *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.
- de Jong, J. T. V. M., Komproe, I. H., Van Ommeren, M., et al (2001)** Lifetime events and posttraumatic stress disorder in 4 postconflict settings. *JAMA*, **286**, 555–562.
- Dohrenwend, B. P. (1990)** 'The problem of validity in field studies of psychological disorders' revisited. *Psychological Medicine*, **20**, 195–208.
- Gliner, J. A. & Morgan, G. A. (2000)** *Research Methods in Applied Settings: An Integrated Approach to Design and Analysis*. Mahwah, NJ: Erlbaum.
- Mezzich, J. E., Kleinman, A., Fabrega, H., et al (eds) (1996)** *Culture and Psychiatric Diagnosis*. Washington: American Psychiatric Press.
- Narrow, W. E., Rae, D. S., Robins, L. N., et al (2002)** Revised prevalence estimates of mental disorders in the United States. *Archives of General Psychiatry*, **59**, 115–123.
- Patel, V., Araya, R., Lima, M. S., et al (1999)** Women, poverty and common mental disorders in four restructuring societies. *Social Science and Medicine*, **49**, 1461–1471.
- Prince, R. H. (1997)** What's in a name? *Transcultural Psychiatry*, **34**, 151–154.
- Robins, E. & Guze, S. B. (1970)** Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *American Journal of Psychiatry*, **126**, 107–111.
- Shrestha, N. M., Sharma, B., Van Ommeren, M., et al (1998)** Impact of torture on refugees displaced within the developing world: symptomatology among Bhutanese refugees in Nepal. *JAMA* **280**, 443–448.
- Turner, S. W., Bowie, C., Dunn, G., et al (2003)** Mental health of Kosovan Albanian refugees in the UK. *British Journal of Psychiatry*, **182**, 444–448.
- Van Ommeren, M., Sharma, B., Thapa, S., et al (1999)** Preparing instruments for transcultural research: use of the Translation Monitoring Form with Nepali-speaking Bhutanese refugees. *Transcultural Psychiatry*, **36**, 285–301.
- , **de Jong, J. T. V. M., Sharma, B., et al (2001)** Psychiatric disorders among tortured Bhutanese refugees in Nepal. *Archives of General Psychiatry*, **58**, 475–482.
- Ware, J. E., Kosinski, M., & Keller, S. D. (1996)** A 12 item short form health survey: construction of scales and preliminary tests of reliability and validity. *Medical Care*, **34**, 220–233.
- World Health Organization (1992)** *Tenth Revision of the International Classification of Diseases and Related Health Problems (ICD-10)*. Geneva: WHO.

MARK VAN OMMEREN, PhD, Department of Mental Health and Substance Dependence, World Health Organization, Avenue Appia, 1211 Geneva 27, Switzerland. E-mail: vanommerenm@who.int

(First received 4 April 2002, accepted 12 April 2002)
