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.

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

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

†See pp.444–448, this issue.

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.
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 translational 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 translational 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.

**CONCLUSIONS**

Systematically considering and addressing validity issues will reduce uncertainty in findings from translational 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.

[377]
DECLARATION OF INTEREST
None.

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

REFERENCES


