

Measurement in Health and Disease

Suggested answers: Validity of the PI HAQ

1. *What is 'face validity' and how would Study 1 establish it?* Face validity means the instrument appears to subjects to measure what it is intended to measure. This is important in this study, as respondents must understand what they are being asked. They establish it by asking patients what terms they understand.
2. *What is 'content validity' and in what sense would Study 2 establish it?* Content validity has two meanings: that all aspects are included and that experts think that the scale covers the correct material. In Study 2, they assume that the long version has content validity and look at whether the short version is similar to it.
3. *In the results for Study 2, what can we deduce about the distributions of the long and short value scales?* They are negatively skew, because the observations reach the maximum but not the minimum and have means closer to the top of the range than the bottom, particularly so for the short form.
4. *In Study 3, why did they use Spearman's ranked product-moment correlation coefficient and Wilcoxon's signed rank test to assess short term reliability? What would the Wilcoxon test tell us?* They used these rank methods because the distributions are not Normal. The Wilcoxon test will tell us whether scores tend to change in a consistent direction from occasion to occasion.
5. *In the results for Study 3, the authors say that 'Patients who gave identical value scores at entry and exit had given a range of scores (that is, had not simply ticked the maximum score to every domain each time)'. Why is this important?* There is a danger in questionnaire layouts with a series of similar questions that respondents will tick the same boxes all the way down. The PI HAQ scale shown on page 993 is of this form. The finding suggests that respondents are thinking about each question rather than, for example, ticking "very important" all the time.
6. *What is the difference between 'criterion validity' and 'construct validity'?* 'Criterion validity' looks at what the measure is supposed to do and compares it with some other measure of the same thing or something closely related, e.g. an anxiety score compared with a clinical diagnosis of anxiety, or with some outcome that the measure is supposed to predict. 'Construct validity' looks at how the measure would be related to other variables if it really measured what we want it to measure.
7. *How does Study 4a assess 'construct validity'?* Study 4a looks at how the value scale is related to many variables which we think should not be related to value though they will be related to the level of disease.
8. *In Study 4a, why is it important that the value scale was independent of the level of disability, clinical status, psychological status, and personality, and that both values and change in values were independent of recent change in any variable? What aspect of validity does this address?* This is divergent validity, they are showing that the scale is measuring something different from either disease severity or psychological state.

9. *In the results for Study 4b, in what sense do the authors use the term 'discriminant validity'?* They are using it in the sense that their scale should distinguish between groups defined by disease severity, not as a synonym for divergent validity..
10. *How does Study 4b assess 'criterion validity'?* It uses a different way to assess the combined disability and the value which respondents attach to it, the time trade off method. This is much used by economists in the development of quality of life studies. They have used it here because they have no direct criterion.