

## Questionnaire development: 2. Validity and reliability

Linda Del Greco, EdD  
Wikke Walop, PhD  
Richard H. McCarthy, MD, CM

**I**n order to have confidence in the results of a study, one must be assured that the questionnaire consistently measures what it purports to measure when properly administered. In short, the questionnaire must be both valid and reliable. In this article we will define validity and reliability and provide some examples of how to think about these issues and how to take some first steps in verifying them.

The importance of validity and reliability cannot be emphasized too strongly. For example, the thermometer must indicate the correct temperature to be valid and must repeatedly give the same reading to be reliable. If the thermometer were reliable but not valid it would give consistently inaccurate readings; if it were valid but not reliable it would indicate different temperatures at each use, the correct temperature being occasionally indicated. In both of these situations the thermometer could not be relied on to contribute to sound clinical judgements.

### Validity

#### Content validity

In part 1 of this series we discussed what areas need to be addressed in the formulation of a questionnaire. Once the questionnaire is drafted one must determine whether the domain has been adequately covered (content validity).<sup>1-3</sup> For example, suppose it was decided that appetite consisted

*From the Department of Epidemiology and Biostatistics, McGill University, the Division of Clinical Epidemiology, Department of Medicine, Montreal General Hospital, and Cornell University Medical Center, Westchester Division, New York Hospital, White Plains, NY*

*Second of a five-part series. Part 1 appeared in the Mar. 15, 1987, issue of CMAJ.*

*Reprint requests to: Dr. Linda Del Greco, New York Hospital, 21 Bloomingdale Rd., White Plains, NY 10605, USA*

of three attributes. In order for the questionnaire to have content validity all three attributes must be questioned sufficiently. A tally of the number of questions addressing each attribute will immediately indicate any imbalance. If an imbalance exists the results may be biased, particularly when the questionnaire yields a single score, as in measurements of functional health status.

#### Face validity

Face validity is not really validity but refers to the appearance of the questionnaire: Does it look "professional" or carelessly and poorly constructed? Professional-looking questionnaires are more likely to elicit serious responses. Therefore, face validity is an important consideration for both the pretest and the final product.

#### Criterion validity

Criterion validity indicates the effectiveness of a questionnaire in measuring what it purports to measure. The responses on the questionnaire being developed are checked against an external criterion, or gold standard, which is a direct and independent measure of what the new questionnaire is designed to measure.<sup>1-3</sup> For example, the number of radical mastectomies reported by surgeons can be validated by reviewing hospital records. Discriminative questionnaires cannot usually be validated by such means because of the absence of an external criterion.

#### Construct validity

Construct validity refers to the extent to which the new questionnaire conforms to existing ideas or hypotheses concerning the concepts (constructs) that are being measured.<sup>1-3</sup> Construct validity presents the greatest challenge in questionnaire dev-

elopment. For example, one could theorize that appetite is logically related to weight retention or gain. Therefore, one could administer the questionnaire to people who have difficulty in gaining or losing weight. If the questionnaire exhibits construct validity there should be a marked difference in how these two groups respond.

Another example is age. One could hypothesize that appetite changes with age, healthy adolescents having larger appetites than senior citizens. If the hypothesis is correct, the questionnaire exhibits construct validity if it discriminates between these two groups.

Another method of establishing construct validity is to ask other questions that measure a variable related to appetite. For example, one could hypothesize that a good appetite is a sign of good health. There is construct validity if a strong correlation exists between the results of the new questionnaire and those of an established measure of health status: a healthy person will score well and an ill person poorly on both.

Creativity and logic are required to establish construct validity. The more ways one can test the construct validity of a new measure, the more confidence one can have in the performance of the measure.

## Reliability

Reliability, or reproducibility, indicates whether the questionnaire performs consistently. There are three ways of examining reliability. The first is to examine the questionnaire's test-retest reliability: the ability of the questionnaire to yield similar results when administered to the same person on two separate occasions. The more reliable the questionnaire the higher the correlation between the results. The interval between the administrations is important. If it is too short the results may be confounded because the subject responds from memory; if it is too long the attribute being examined may have changed, and the low correlation may indicate this change rather than poor reliability.

A second method is to examine interobserver reliability. The same subject is evaluated by two interviewers, using the same questionnaire. The results will correlate well if the questionnaire has good interobserver reliability.

The third method examines the consistency within the questionnaire: the degree to which a

subject answers similar questions in a similar manner. One method is to administer two equivalent forms of a questionnaire at the same time to a subject. This method is rarely used because of the difficulty in formulating or finding two equivalent questionnaires.

A more feasible method for testing the consistency of homogeneous (single-attribute) questionnaires is the split-half method: the even- and odd-numbered questions are separated and are considered to be two equivalent questionnaires. The internal consistency of a homogeneous questionnaire can also be examined after a single administration by applying an appropriate statistical procedure.

The split-half method cannot be used with heterogeneous questionnaires because division of the questionnaire will not yield "equivalent" forms. In this situation one may repeat questions throughout the questionnaire; only the original question is kept in the final form.

Finally, one can look at the logical patterns of answers. For example, one question might ask: "Would you say you are never tired, sometimes tired or always tired?" The next question could ask: "Do you sometimes feel tired in the afternoon? Yes, No or Not applicable?" Subjects answering "never tired" to the first question should answer "No" to the second.

As a result of testing validity and reliability, questions are rewritten, eliminated or added. This process is repeated until the questionnaire meets the standards set by the researcher. Special care must be taken when eliminating and adding questions to ensure that the content validity is not jeopardized.

To conduct a study in an area where people speak foreign languages may require a translation of the questionnaire. If a translation is not necessary the researcher can proceed directly to the development of the code book. Given the importance of language, translation will be dealt with in the next issue.

## References

1. Anastasi A: *Psychological Testing*, 4th ed, Macmillan, New York, 1976: 134-140
2. Last JM, Abramson JH, Greenland S et al (eds): *A Dictionary of Epidemiology*, Oxford U Pr, New York, 1983: 107
3. Kirshner B, Guyatt G: A methodological framework for assessing health indices. *J Chronic Dis* 1985; 38: 27-36

---

## Without health

*Without health life is no life; it is unlivable . . . Without health, life spells but languor and an image of death.*

— François Rabelais (1494?-1553)