June 1994

Validity and reliability

Muhammad Masood Kadir
Aga Khan Unviversity, masood.kadir@aku.edu

Asma Fozia Qureshi
Aga Khan Unviversity

Validity and Reliability

Muhammad Masood Kadir, Asma Fozia Qureshi (Department of Community Health Sciences, The Aga Khan University, Karachi.)

What is validity?

validity is used in epidemiology to assess the degree to which the information collected accurately answers the research question; i.e., the extent to which the results are accurate and the extent to which the conclusions derived can be generalized.

The term validity in epidemiological research is used in three different ways[1]:

A. Internal Validity (Study Validity)
B. External Validity (Generalizability)
C. Measurement validity (Variable).

A. Internal Validity is the degree to which the observed findings lead to correct inferences about phenomena taking place in the study sample. A study is not valid if it cannot provide accurate information, or cannot enable well-founded inferences to be drawn from the population studied. For example, a study shows a higher risk of lung cancer among coffee thinkers. This increased risk attributed to coffee drinking is incorrect as coffee drinkers are more likely to smoke and therefore, show a higher risk of lung cancer.

B. External Validility is the degree to which the inferences drawn from a study can be generalized to a broader population beyond the study population. Example[2]: A number of studies on white males in developed countries show that current smoking status increases the risk of fatal coronary heart disease. It remains to be judged whether these findings in white males can be generalized to other populations such as the males in Pakistan. However, based on the known mechanisms of tobacco smoke components, it would not be unrealistic to assume that smoking is equally harmful for Pakistani males or females.

C. Measurement Validity is the degree to which a test actually measures what it is designed to measure, The process involves comparison with a technique known to be accurate (the gold standard). The validity of a measurement of body weight, for example, can be checked by calibrating the scale with standard weights. Similarly, a laboratory test must be appraised to establish validity, which is measured by the sensitivity and specificity of the test. Sensitivity is the ability of the test to detect correctly those individuals who have a disease. While specificity is the ability to detect correctly those individuals who do not have the disease.

What threatens validity?

Internal Validity

Internal validity may be impaired by selection bias, information bias, uncontrolled confounding, or an unduly small study sample.

External Validity:

External validity may be impaired by a small study sample, and/or improper selection of sample.

Validity of a measure:

Validity of a measure can be affected by observer and instrument bias. In order to improve validity, the researcher needs to be aware of the factors that threaten validity of a study and should take them into account during study design and analysis of results. Is internal validity more important than external validity (generalizability)?

Internal validity is much more important than generalizability since the inferences drawn from a study cannot be generalized unless they are valid. While designing a study one may be tempted to enhance generalizability of the study findings to a larger population. To achieve this one may try to enlarge the

study sample to include subjects who are more representative of the whole population. However, it is usually much more difficult to get cooperation of a random sample of participants from an entire population and to collect accurately complete information from them. A very large sample from the entire population also increases the likelihood of confounding and/or bias. In order to enhance validity, it is mare appropriate to restrict the study population to individuals who are comparable in other respects for the outcome under study and on whom complete and accurate information can be obtained.

Example[3,4]: In 1976 a large scale prospective cohort study was initiated in the USA to evaluate the health effects of different contraceptive practices among women. The researchers considered selecting a large random sample of all women in USA. However, they were concerned about such a design because they felt that it would be difficult to obtain detailed and accurate medical information from all the women by mail questionnaire alone. Also complete long term follow up of more than 100,000 women from 50 states would be logistically unfeasible. The sample was, therefore, restricted to all married, female registered nurses between the ages of 30 and 55 years residing in 11 states with a large number of registered nurses. The choices of nurses increased the likelihood of obtaining accurate medical data by mail, as well as chances of complete follow up, resulting in better internal validity. The question, however, arises whether findings from this study can be generalized to other population groups and largely depends upon the specific research question.

What is Reliability?

Reliability refers to the degree to which a measurement procedure can be reproduced. Lack of reliability may arise from differences between observation, or instruments of measurements, or instability of the attribute being measured[5]. A beam scale can measure body weight with great precision (that is with great reliability), on the other hand a questionnaire designed to measure quality of life is more likely to produce values that vary from one occasion to the next.

Can reliability exist without validity?

Reliability can be present without validity. For example in a rural area of Pakistan if an illiterate mother is asked the ages of her children she may report the same ages every time she is asked. However, the ages reported might not be the exact ages of her children.

What factors affect reliability?

Three main sources can result in errors of measurement[6]:

Observer Variability refers to variability in measurement that is due to the observer and includes such things as choice of words in an interview. For example if two different persons record the blood pressure of an individual the value recorded by them may vary. Also the value can differ if one person records the blood pressure for the second time in the same individual, irrespective of biologic variability. Subject Variability refers to intrinsic biologic variability in the study subjects due to such things as fluctuations in mood or circadian rhythms: For example, the blood pressure of an individual varies before and after exercise. Instrument Variability refers to variability in the measurement due to fluctuating environmental factors or the instrument used for measurement. If two different' blood pressure recording instruments are used (an aneroid and a mercury sphygmomanometer), the reading in the same individual may vary.

How can one increase reliability?

There are several ways to enhance reliability[6]:

Standardizing the measurement methods for example all study protocols should include operational definitions and precise instructions for recording measurements. Training and supervision of observers improves consistency of measurement techniques. Instruments can be designed to reduce variability. Variation in the way human observers make measurements can be eliminated with automatic mechanical devices and self administered questionnaires. The impact of random error of any source is reduced by repeating the measurement and using the mean of two or more readings. Thus in order to ensure a sound investigation, draw meaningful conclusions and interpret results correctly, it is essential

to consider the validity and reliability of the data and study design. Validity is essential for inference as this indicates whether the conclusions drawn are based on the accurate information. Reliability helps ensure that the same results would be obtained if the same methods are used again.

References

1. Abramson, ill. Making aense of data - a self-instruction manual on the interpretation of epidemiological data. latEd. NewYork, Oxford University Press, 1988.
2. Hannekena, C.H. and Buring, J.E. Edited by S.L. Myrent. Epidemiology in medicine. 1st Ed. Boston, Little Brown and Company, 1987.
3. Hennenkens, C.H., Speizer, F.E., Rosner, B., et al Use of permanent hair dyes and cancer amongregistered nurses. Lancet, 19791:1390-93.
4. Stampfer, M.J., Willet, W.C., Colditz, O.A., etaL Aproapectiveatudy of postmenopauaal estrogen therapy and coronary heart disease. N.Engl.J.Med., 1985;313:1044-49.
5. Last, J.M. (Ed). A dictionary of epidemiology. 2nd Ed. New York, Oxford Univeraity Press, 1988.
6. Hulley, S.B. and Cummings. SR. (Ed). Deaigning clinical reaearcb - an epidemiologic approach. latEd. Baltimore, Williams and Wilkina, 1988.