



QUESTIONING THE QUESTIONNAIRE: TESTING VALIDITY AND RELIABILITY OF QUESTIONNAIRES

Medicine

Dr. Varun Mohan Malhotra*	Professor, Department of Community Medicine, Adesh Institute of Medical Sciences and Research, Bathinda, Punjab 151001 *Corresponding Author
Dr Aishwarya Kapoor	Demonstrator, Department of Microbiology, Adesh Institute of Medical Sciences and Research, Bathinda, Punjab 151001.
Dr. Rashpreet Kaur	Assistant Professor, Department of Management and Hospital Administration, Adesh University, Bathinda, Punjab 151001.

ABSTRACT

The questionnaire is the most commonly used instrument for data collection, especially in quantitative bio-medical research. Post-graduates in medical schools in India, generally are well aware of the significance and techniques of sampling including the calculation of sample size. However, the development of the questionnaire as an instrument for data collection has remained backstage. The article discusses the concept, scope and techniques for testing the validity and reliability of questionnaires as measurement tools in biomedical research.

KEYWORDS

Questionnaire, Reliability, Validity

INTRODUCTION

A questionnaire is a predetermined set of questions used to collect data [1]. Questionnaires can be filled out by the participants or completed by the interviewer (also called interview schedule) based on the responses of the respondent. The process may be face-to-face, through telephone, post or email. The questions may be open- or close-ended, dichotomous or multiple choice, asked on a Likert or slider scale, or a combination of these; the objective is to capture unbiased, usable and desired information from a substantial number of respondents in the most cost-effective manner. Developing a questionnaire is an important task in any biomedical research because a questionnaire has the potential to influence the internal and external validity of the study [2]. Although the sampling i.e. sampling technique and calculation of sample size are, generally given the importance it deserves, the process of development of questionnaire as the research instrument often remains backstage. This article discusses the concept, scope and techniques of testing the validity and reliability of questionnaires for bio-medical research[3,4].

Concepts of Validity and Reliability

Validity expresses the degree to which an instrument captures what it purports to measure. It is the accuracy e.g., a thermometer recording the temperature, accurately. It should be appreciated that in research, the mistake may be at a different plane; attempting to measure temperature by say a sphygmomanometer is a crude example but may happen because how of measurement are not standardized. In addition, our measuring instrument (e.g., glucometer) can potentially err in either direction i.e. it may show a non-diabetic to be hyperglycemic or an uncontrolled diabetic as normal. Sensitivity i.e., the ability of the test (or question in a questionnaire) to correctly diagnose those having the characteristic (e.g., disease), and specificity is the ability of the test (or question) to correctly diagnose those without the characteristic. Sensitivity and specificity are the two pillars of validity. In the context of the validity of a questionnaire, several types e.g. face validity, content validity, criterion validity and construct validity have been described, and are discussed later.

Reliability refers to the degree to which the results obtained can be replicated i.e. the questionnaire gives the same results when re-applied in a different time frame or different recorder, under similar circumstances. Three aspects of reliability, namely stability, equivalence, and homogeneity are relevant to the questionnaire and are mentioned later. The oft-repeated figure of 'target practice' depicts the concepts of validity and reliability aptly and is reproduced in figure 1.

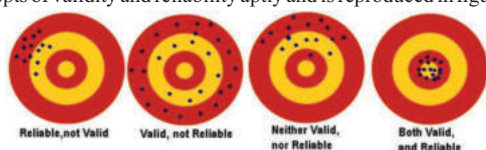


Figure 1: Diagrammatic representation of Validity and Reliability

Verifying Validity of a Questionnaire: All questionnaires designed by researchers should undergo scrutiny and rigors to establish their validity. Validity limits the systematic or built-in error in the questionnaire.

The validity of a questionnaire has been further divided into sub-groups. Although there is no consensus regarding the division, we would discuss validity under 3 major sub-groups (i) Translational or representational validity which is further divided into the face- and content-validity (ii) Criterion-related validity with two sub-sub groups of concurrence and predictive validity, and (iii) Construct validity which encompasses subtypes of convergence, discriminant, known-group and factorial validity.

Translation Validity: Face and content validity are established when an expert or experts on the research subject review the questionnaire and conclude that it measures the characteristic or trait of interest (face validity), and the questionnaire is comprehensive enough to capture all domains of the concept (content validity) [5,6]. For example, the questionnaire regarding the quality of life developed by WHO (WHOQOL-BREF) has incorporated the domains of physical health, psychological aspects, environmental issues and social relationships to capture the quality of life, comprehensively. Both face- and content validity has been criticized for being too subjective. More recently, various rate scales have been developed to bring more objectivity to content validity.

Criterion validity is assessed when one is interested in determining the relationship of scores on a test to a specific criterion. It is a measure of how well questionnaire results stack up against another instrument or predictor [7]. It is conceptually akin to comparing a new diagnostic test against a 'gold-standard' test to detect the sensitivity, specificity, and predictive value of the new test. The challenge, in research, maybe that the 'gold standard' itself has not been identified. If the process of comparison can be completed concurrently, we are doing 'concurrence validity'. However, the ability of the questionnaire to correctly forecast a future health event depicts its predictive validity. For instance, a researcher may use a questionnaire to elucidate current cardiovascular complications of diabetic patients and compare the results with the results of investigations like EKG, ECHO Stress test etc. to establish concurrence validity. On the other hand, establishing predictive validity i.e., the ability to forecast events is a resource-intensive exercise and may not be feasible in most situations. However, questionnaires developed by international, national health organizations (e.g. WHO) and health institutions (e.g. Medical schools) should undertake exercises to detect the predictive validity of the questionnaire. This adds to the robustness of the questionnaire and assists further research by providing a validated tool.

Construct Validity. Construct validity is the degree to which an instrument measures the trait or theoretical construct that it is intended

to measure [8]. It differs from criterion-related validity as there is no criterion for comparison; hence the process utilizes a hypothetical construct for comparison. It is the most difficult measure of validity. Construct validity of a study instrument is established using one or more of the following sub-types.

Convergent Validity (or lack of it) is achieved if comparable results (convergence) about the same concept are obtained when the construct is measured in different ways [7]. For example, in a study on determinants of obesity in urban adolescents, we compared the screen time as reported by the participants (self-reporting) with actual screen time (timed by parents) in pilot testing of the questionnaire to test the 'accuracy' of the self-reporting. Using the same logic, discriminant validity provides evidence through two tests that measure dissimilar concepts, and show results that show a negative correlation. [7]. For example, in the study of adolescent obesity, higher screen time showed an inverse relationship with outdoor recreational activities.

Known-group Validity is tested by exposing the questionnaire to a known group i.e., a group with the already established attribute of the outcome (e.g. diseased) and comparing the results with a group in whom the attribute is absent (non-diseased). Since the attribute of the two groups of respondents is known, a valid questionnaire would have a high percentage of 'True positives', and 'True negatives' [9]. For example, a questionnaire to identify the risk of suicide should measure higher risk when applied to clinical cases of depression (known group) as compared to cases without depression.

Factorial validity: This validates the contents of the construct by employing the statistical model called factor analysis [10]. An example will clarify the concept better. In WHOQOL_BREF, seven questions designed to measure a physical health domain will tend to show similar results to one another when compared to answers to three questions asked to measure the social-relationship domain.

Testing Reliability of a Questionnaire. Reliability is the extent to which a questionnaire, test, observation or any other measurement tool produces the same results on repeated trials. It reflects the stability or consistency of results over time and across raters. It is important to understand that lack of reliability may arise from differences between observers (inter-observers' variation) or instruments of measurement or instability of the attribute being measured (e.g. diurnal variation in blood pressure). Reliability is usually assessed in three ways; test-retest reliability, alternate-form reliability and internal consistency reliability.

Test-retest correlation (or stability) is evidence of the temporal stability of the measuring instrument. This occurs when the same or similar scores are obtained with repeated testing under similar circumstances, in a different time frame. It is the most used test for the reliability of questionnaires. Test-retest reliability, however, needs fulfilment of two assumptions: firstly, there is no real variation in the characteristics in time (e.g. a patient may improve or deteriorate later), and secondly, the variation is not subjective (perception about the usefulness of a vaccine) [11].

Alternate-form Reliability (or Equivalence): Alternate form refers to the amount of agreement between two (or more) research instruments. For example, two different questionnaires (parallel forms) having similar (but not identical) questions (say questions in different wording, or reshuffled order) are administered together. A high degree of correlation between the two results confirms reliability [12]. Another way to test equivalence is by testing individuals with identical instruments, but with a different interviewer (inter-observer reliability). This test is more suitable when the questionnaire is semi-structured with open-ended questions where the interviewer has more potential to influence results, consciously or subconsciously.

Internal -consistency reliability, also called homogeneity is most often established by the split-half technique. We randomly divide questions purporting to measure the same construct into two sets. The two sets are separately administered during the pilot study to a sample of individuals. high correlation (usually calculated as Cronbach's alpha) between two sets estimates homogeneity [13].

CONCLUSION

Designing a questionnaire as a measurement tool is an important process in the planning stage of research. Unfortunately, its development and its ability to capture what it is expected to do gets

back-stage in biomedical research. It is recommended that the significance of validity and reliability of study instruments is ensured by the young researchers as well as research supervisors. The questionnaire deserves to undergo a process of 'examination' so that it is 'not questioned' later; lest the result of the research may be 'questioned'.

REFERENCES:

1. Kember, David, and Doris YP Leung. Establishing the validity and reliability of course evaluation questionnaires." *Assessment & Evaluation in Higher Education* 33.4 (2008): 341-353.
2. Taherdoost, H. Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/ Survey in a Research SSRN Electron. (2018): 28-36.
3. Norland-Tilburg, E. V. Controlling error in evaluation instruments. *Journal of extension* 28.2 (1990): 23-41.
4. Strategic Information Consultants. What Are You Really Measuring? Reliability and Validity in Questionnaire Design. Data Analysis Australia, 2017. Available at www.daa.com.au/analytical-ideas/questionnaire-validity/ (Last accessed June 03, 2022)
5. Sangoseni, Olaide, Madeleine Hellman, and Cheryl Hill. Development and validation of a questionnaire to assess the effect of online learning on behaviors, attitudes, and clinical practices of physical therapists in the United States regarding evidenced-based clinical practice." *Internet Journal of Allied Health Sciences and Practice* 11.2 (2013): 7.
6. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. *Res Nurs Health*. 2006 Oct;29(5):489-97. doi: 10.1002/nur.20147.
7. Bolarinwa OA. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Niger Postgrad Med J*. 2015 Oct-Dec;22(4):195-201. doi: 10.4103/1117-1936.173959.
8. Anderson JL, Sellbom M. Construct Validity of the DSM-5 Section III Personality Trait Profile for Borderline Personality Disorder. *J Pers Assess*. 2015 Sep-Oct;97(5):478-86. doi: 10.1080/00223891.2015.1051226.
9. Hofman CS, Lutomski JE, Boter H, Buurman BM, de Craen AJ, Donders R, et al. Examining the construct and known-group validity of a composite endpoint for The Older Persons and Informal Caregivers Survey Minimum Data Set (TOPICS-MDS): A large-scale data sharing initiative. *PLoS One*. 2017 Mar 15;12(3):e0173081. doi: 10.1371/journal.pone.0173081.
10. Motl RW, Dishman RK, Trost SG, Saunders RP, Dowda M, Felton G, et al. Factorial validity and invariance of questionnaires measuring social-cognitive determinants of physical activity among adolescent girls. *Prev Med*. 2000 Nov;31(5):584-94. doi: 10.1006/pmed.2000.0735.
11. Singh AS, Vik FN, Chinapaw MJ, Uijtendwilligen L, Verloigne M, Fernández-Alvira JM, et al. Test-retest reliability and construct validity of the ENERGY-child questionnaire on energy balance-related behaviours and their potential determinants: the ENERGY-project. *Int J Behav Nutr Phys Act*. 2011 Dec 9;8:136. doi: 10.1186/1479-5868-8-136.
12. Stevenson JD Jr. Alternate form reliability and concurrent validity of the PPVT-R for referred rehabilitation agency adults. *J Clin Psychol*. 1986 Jul;42(4):650-3. doi: 10.1002/1097-4679(198607)42:4
13. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011 Jun 27;2:53-55. doi: 10.5116/ijme.4dfb.8dfd.