EDITORIAL

Validating instruments of measure : Is it really necessary?

Lai PSM

Authors:

Lai Siew Mei Pauline, B.Pharm, PhD
Department of Primary Care Medicine, University of Malaya Primary Care Research Group (UMPCRG), Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
Tel: +60379493920
Email: plai@ummc.edu.my

In this April issue of the Malaysian Family Physician, there are two manuscripts on the validation of instruments. The first manuscript is the validation of the Malay version of the Berlin Questionnaire to identify Malaysian patients at risk for obstructive sleep apnea,[1] whilst the second manuscript is on the validation of the Malay version of the Diabetes Quality of Life for Youth Questionnaire.[2]

All instruments assessing patient reported outcomes have to be evaluated for its reliability and validity in the country prior to its use. The purpose of this is to ensure that the instrument used is measuring what it is supposed to measure. This is applicable to instruments that have been developed in English by other authors, and validated elsewhere; as well as self-developed instruments or those that have been modified (e.g. translated into another language or otherwise).

Instruments that have been developed in English and validated elsewhere

Many Malaysian researchers are estatic when they discover that an instrument which they would like to use has been developed in English and validated in countries like the United States, United Kingdom or Australia. A common mistake is to assume that this questionnaire is also suitable for use in Malaysia. Although English is widely spoken and understood by many Malaysians, the English used overseas may not necessarily be interpreted the same way in Malaysia. This is often due to the cultural differences that exist between different populations in different countries. An example is the Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO-41), which was designed by the International Osteoporosis Foundation.

Researchers will need to seek permission for use from the original author(s). Sometimes, the original authors will state that changes to the original instrument are not allowed. This should be mentioned when validating the instrument, as it will then determine whether the instrument can subsequently be modified or not.

Sample size

The sample size required for validation studies is usually based on a rule of thumb, i.e. 5-10 participants should be recruited for every item of the instrument, in order to conduct confirmatory factor analysis.[3] The Berlin Questionnaire has 10 items. Although Yunus et al did not perform confirmatory factor analysis, they recruited 150 participants to validate their instrument. The Diabetes Quality of Life (DQOL) instrument however, has 60 items. The number of participants required should be 300-600. In this manuscript, the authors only recruited 82 participants, which is the minimum requirement for a validation study.

It is sometimes difficult to satisfy this "rule of thumb", especially in cases where the condition is uncommon (e.g. children with epilepsy with normal cognitive function). This will then need to be stated as a limitation of the study,

where confirmatory factor analysis will not be conducted.

Statistical analyses used in validation

The reliability of the instrument is the extent to which an instrument of measure yields the same result when used repeatedly. This involves the agreement of measuring instruments over time, where a test-retest is conducted. The duration of test-retest may vary from two to six weeks. Results are then compared and correlated with the initial test to give a measure of stability.[24] Spearman's rho or Pearson's correlation can be used to determine the relationship between each item of test-retest. Any correlation coefficient value which equals to ±0.4 was considered as a moderate association and therefore acceptable.[35]

Internal consistency is a measure of reliability of different survey items intended to measure the same characteristic. It is used to determine whether all items in a multi-item scale measures the same concept. Internal consistency is assessed using Cronbach's α coefficient, where a value of >0.7 is considered good.[46] The effect of removing a single item on the Cronbach's α should also be determined. Corrected item-total correlations can be used to identify items which did not agree well with other items in the questionnaire. Item-total correlations should >0.2 to be considered as acceptable.[57]

Dimensionality of an instrument can be analysed using maximum likelihood factor analysis.

Factor analysis that shows an eigen value>1 indicates that there are more than one component in the instrument.[13] Construct validity seeks agreement between a theoretical concept and a specific measuring device or procedure. This can be divided into two sub-categories: convergent validity and discriminant validity.[68]

Convergent validity is the actual general agreement among ratings, gathered independently of one another, where measures should be theoretically related.[79] In the validation of the Berlin Questionnaire by Yunus et al, the authors assessed obstructive sleep apnea by asking all participants to attend an overnight level I polysomnogram. Convergent validity was not performed by Jalaludin et al. in the validation of the Diabetes Quality of Life for Youth Questionnaire.

Discriminant validity is to determine the lack of a relationship among measures which theoretically should not be related.[68] E.g. the QOL should be better in participants without back pain versus those with back pain. Both the manuscripts published did not assess discriminant validity.

Self-developed instruments or instruments which have been modified (translated into another language or otherwise)

It is very common for Malaysian researchers to translate a previously validated instrument to Bahasa Malaysia as per the two manuscripts published in this month's issue. The translation process itself must be rigorous and should be performed according to guidelines and standards for the translation and cultural adaptation of patient-reported outcomes.[810] The final translated version will need to be pilot tested to evaluate the clarity of the translated document, whilst preserving the original content and meaning. Similarly, when developing new instruments, the face and content validity will need to be established as described above.

It is therefore very encouraging to see the two manuscripts on validation of instruments in this issue. This shows that Malaysian researchers are more aware of the importance of instrument validation. Future effort should focus on publishing more manuscripts of this nature and to create a network for researchers to collaborate in this area. We should also create an archive for all instruments that have

been validated in Malaysia so that it will be easily accessible to researchers measuring patient reported outcomes.

References

1. Yunus A, Seet W, Mohamad Adam B, Haniff J. Validation of the Malay version of Berlin questionaire to identify Malaysian patients for obstructive sleep apnea. Malaysian Family Physician 2013;8(1):03-9

2. Mohamad Adam B, Jalaludin MY, Fuziah MZ, Hadhrami MH, Janet YH Hong, Jamaiyah H. Reliability and validity of the Malay translated version of diabetes quality of life for youth questionaire. Malaysian Family Physician 2013;8(1):10-7

3. Gorsuch RL. Factor analysis. 2nd ed. Hillsdale, N.J: L. Erlbaum Associates; 1983.

4. Colorado State University. Stability reliability and internal consistency. 1993 [cited 2009 31 March]; Available from: http://writing.colostate.edu/guides/research/relval/com2a3.cfm.

5. Archambault S. Spearman rho. Psychology Department, Wellesley College; 2000 [updated 31 July 2000; cited 2009 17 May]; Available from: http://www.wellesley.edu/Psychology/Psych205/spearman.html.

6. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16(3): 297-334.

7. Devellis RF. Scale Development. Sage Publications; 1991. p. 24-33.

8. Campell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix Psychological Bulletin1959. p. 81-105.

9. Carmines EG, Zeller RA. Reliability and viability assessment. CA: Thousand Oaks; 1991.

10. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005;8(2):94-104.