Article

Internal consistency and interrater reliability of the Brazilian version of Martín-Bayarre-Grau (MBG) adherence scale

Samara Ramalho Matta[1,*], Thiago Botelho Azeredo[2], Vera Lucia Luiza[3]

[1]Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro, RJ, Brasil,[2] Observatório de Vigilância e Uso de Medicamentos, DEFARMED/FF/UFRJ,[3]Departamento de Política de Medicamentos e Assistência Farmacêutica, NAF/DCB/ ENSP/Fiocruz, Rio de Janeiro, RJ, Brasil

This paper aims to analyze the measurement equivalence aspects (internal consistency and interrater reliability) of a Brazilian version of Martín-Bayarre-Grau (MBG) adherence questionnaire as part of its cross-cultural adaptation. Item-total correlation and Cronbach's alpha coefficients were used as internal consistency estimates. Stability was evaluated through test and retest comparison and expressed through intraclass correlation coefficient (ICC) and kappa with quadratic weighting. ICC for the overall scale was 0.81, indicating an "almost perfect" agreement. However, some cases of "poor" and "slight" agreements were found while analyzing individual items. The translated version of the MBG questionnaire showed good homogeneity (alpha 0.78), higher than cutoff points suggested in the literature. The scale has proved capable of measuring the level of adherence to treatment in hypertensive and/or diabetic patients in a reliable way.

Uniterms: Adherence to medication. Reproducibility of results. Questionnaires/study. Martín-Bayarre-Grau/study/aspects

INTRODUCTION

Poor adherence to chronic treatment affects the health of individuals and has economic consequences to health systems, which cover populations with high prevalence of chronic diseases (WHO, 2003).

Among methods applied to investigate adherence, patient interviews are widely used because they are easy to apply and have low cost, in spite of their limitations (Osterberg, Blaschke, 2005; Garfield *et al*., 2011; Nguyen, La Caze, Cottrell, 2014). Interviews can be conducted using questionnaires that are previously validated, developed for this purpose or translated.

If one opts to translate a questionnaire, a formal procedure of cross-cultural adaptation should be followed. This process culminates with the study of psychometric properties of the adapted scale (Reichenheim, Moraes, 2007). In this final stage of adaptation, measurement equivalence between versions is analyzed through

reliability and validity assessment (Reichenheim, Moraes, 2007), generating information on the scale's suitability to the application context.

Despite the importance of knowing these properties, a systematic review shows that data concerning internal consistency and test-retest reliability are available only for a relatively small number of adherence measures (Garfield *et al*., 2011).

The Cuban Martín-Bayarre-Grau (MBG) questionnaire (Alfonso, Vea, Ábalo, 2008) was selected for the cross-cultural adaptation because it covers the range of dimensions involved in the concept of adherence proposed by WHO (2003), which emphasizes the active role of the patient in the treatment as fundamental to adherence to long-term therapies. The questionnaire includes twelve questions with five-point Likert type response options, addressing three dimensions: compliance with treatment, personal implication and doctor-patient mutual respect. It is a quick application questionnaire, useful in health services settings.

This paper aims to analyze measurement equivalence aspects (internal consistency and interrater reliability) of a Brazilian version of Martín-Bayarre-Grau (MBG)

*Correspondence: S. R. Matta. Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro. Rua Professor Carlos Wenceslau, 343 – Realengo - 25715-000 - Rio de Janeiro - RJ, Brasil. E-mail: samara.matta@ifrj.edu.br

adherence questionnaire as part of its cross-cultural adaptation.

METHODS

Reliability analyses (internal consistency and stability – interrater reliability) were performed as part of the pilot study "The medicine at home program as public medicine distribution model – analyzing the implementation in the city of Rio de Janeiro" – RECASA. The RECASA program consisted mainly in the delivery of antihypertensive and antidiabetic medicines to enrollees at home.

This study was conducted in 2011 and analyzed the implementation of this governmental medicines provision model. The pilot study was conducted in December 2010 through a test-retest application of the questionnaire in face-to-face interviews at patients' home.

Sample size for the pilot study was calculated assuming simple random sampling from a finite population. We opted for the worst scenario, since outcome variables were unknown. Feasibility to conduct the pilot study in a short time was also considered. A sampling error of 20% and 5% significance level were used, resulting in a sample of 25 individuals.

A second sample size was calculated to ensure pilot study sample adequacy to a reliability study. An expected intraclass correlation coefficient (ICC - main interrater reliability estimate for this study) was set at 0.8 against a minimum of 0.5. Two observations were considered (test and retest) and a significance level of 5% and power of 80% were used to generate a sample size of 22 individuals. The Winpepi program (http://www.brixtonhealth.com/pepi4windows.html) was used for this estimate. Given the proximity of this number with the full sample necessary to the pilot, the ICC was calculated based on the 25 individuals interviewed.

Criteria for inclusion of individuals in the pilot sample were: to have been diagnosed with hypertension (HT) and/or diabetes (DM) and be under prescribed treatment; to be 18 years old or older; in the case of DM patients, using oral antidiabetic medication. A reference health care facility provided a patients list for the random selection. This health care facility was chosen because of its location in a neighborhood comprising a diversity of socioeconomic levels and schooling, as well as easy access.

The questionnaire was applied with the aid of a vignette in order to facilitate patients' recollection of response options (Likert scale). At the end of the first interview (test), the best day to conduct the second interview (retest) was set, keeping an interval ranging from five to seven days between interviews. Two typists independently entered questionnaire information in test and retest databases. Databases were then compared, corrected and merged.

Internal consistency was estimated by calculating item-total correlation and Cronbach's alpha coefficients for the test and the retest, using the SPSS 8.0 program. Interrater reliability was estimated by calculating intraclass correlation coefficient (ICC) between test and retest total scores. In addition, kappa with quadratic weighting was used to analyze individual items' test-retest level of agreement. ICC and kappa were calculated using VassarStats application (http://faculty.vassar.edu/lowry/kappaexp.html), using a 95% confidence interval.

Cutoff points for inferring adequate internal consistency and interpreting of interrater reliability coefficients were set at 0.70 for Cronbach's alpha (Streiner, Norman, 2003) and defined in ranges proposed by Landis and Koch (1977) for ICC and Kappa: >0 (poor); 0 to 0.20 (slight); 0.21 to 0.40 (fair); 0.41 to 0.60 (moderate); 0.61 to 0.80 (substantial); and, 0.81 to 1.00 (almost perfect).

The research project on which this study nests was approved by the Research Ethics Committee of the Sérgio Arouca National School of Public Health and the Civil City Department of Health and Defense of Rio de Janeiro through protocols CAAE 0157.0.031.000-09 and CAAE 0257.0.314.000-09, respectively.

RESULTS

During telephone contacts, main challenges were problems in the telephone book, refusals and several additional calls. However, most visits without prior appointment were successful. Thirty people were interviewed due to the need for replacement to ensure the minimum 25 test and retest interviews.

Most respondents were female (60%), married (40%), average age was 62 years (SD 8.1 years) and 40% were employed in the private sector (Table I). Refusals on retest did not cause major changes in the profile of the subjects included in the study (Table I).

Most respondents in the test (76%) and the retest (72%) showed 'partial adherence' considering Alfonso, Vea and Ábalo (2008) classification. The average score of the final MBG adherence scale showed values to the test (32.4, SD 7.9 points) close to the retest (33.04; SD 8.5 points), indicating that the instrument should have good agreement level in reliability tests (Table II).

Cronbach's alpha in the retest (0.79) was slightly higher than in the test (0.78) and values obtained excluding

TABLE I - Selected characteristics of pilot respondents. Rio de Janeiro Municipality, 2010

Variables	n	%
Gender (n=25)		
Male	10	40
Female	15	60
Age in years (n= 25)		
45 – 55	6	24
56 – 66	13	52
67 – 77	6	24
Marital status (n=25)		
Single	9	36
Married	10	40
Widow or divorced	6	24
Ethnicity (n=24)		
White	6	25
Not-white	18	75
Main occupation (n=25)		
Private sector employee	10	40
Retired	7	28
Family care	5	20
Employer, self-employed, volunteers	3	12
Average number of medicines in use (SD)	3.8 (SD 2,1)	-

each item followed this pattern of slight superiority in the retest. The corrected item-total correlation average was 0.41 for the test and 0.45 for the retest, and the values obtained for item D were the lowest in both test and retest. The intraclass correlation coefficient for the total score was 0.81 (95% CI 0.62 to 0.91). Kappa with quadratic weighting varied from 0.09 (slight agreement) to 0.96 (almost perfect agreement) (Table III).

DISCUSSION

The internal consistency of our adapted version may be considered high. Although it was lower than that of the original scale (0.89) (Alfonso, Vea, Ábalo, 2008), it was compatible with the internal consistency level usually found and deemed appropriate for other measures (>0.7) (Nguyen, La Caze, Cottrell, 2014; Osterberg, Blaschke, 2005). Also, the MBG Portuguese version Cronbach's alpha was higher than other Portuguese adherence scale versions, such as Morisky-Green test (0.66) and Brief Medication Questionnaire (0.73) (Ben, Neumann, Mengue, 2012). Furthermore, the MBG scale's internal consistency would not increase significantly with the exclusion of any item, indicating all items contribute to the homogeneity of the scale. Other scales subject to cross-cultural adaptation to Portuguese had alpha higher than 0.8 (Imaginário et al., 2014; Monteiro, Tavares, Pereira, 2012). However, these studies applied larger sample sizes, which increase of Cronbach's alpha value.

The original scale average item-total correlation was superior to 0.5, which was considered a good level of internal consistency (Alfonso, Vea, Ábalo, 2008). In our study, average item-total correlations stood at less than 0.5 in the test (0.41) and retest (0.45).

Corrected item-total correlation coefficients indicate the correlation of an item with the total scale when that item is omitted. Literature suggests values over 0.2 show a good level of correlation (Streiner, Norman, 2003).

Items D and H showed the lowest values for item-total correlations. If item D was excluded, Cronbach's alpha in the test would not suffer alteration and it would increase slightly in the retest. Furthermore, agreement between test and retest was slight for item H and fair for item D. These items contribute poorly to the scale internal consistency and reliability. These items performed better in the original scale regarding item-total correlation and Cronbach's alpha; interrater reliability was not estimated

TABLE II - Adherence score in test-retest of Portuguese version of Martín-Bayarre-Grau (MBG) scale. Rio de Janeiro Municipality, 2010

Adherence level*	Test (n=25)		Retest (n=25)	
	n	%	n	%
Total	6	24	7	28
Partial or none	19	76	18	72
Final score	Average (SD) 32.4 (7.9)		Average (SD) 33.0 (8.5)	

*Cutoff points (Alfonso, Vea, Ábalo, 2008): total adherence (38 a 48 points), partial adherence (18 a 37 points), no adherence (0 a 17 points).

TABLE III - Internal consistency and interrater reliability for the Portuguese version of Martín-Bayarre-Grau (MBG) scale. Rio de Janeiro Municipality, 2010

Items*		Corrected item-total correlation coefficient		Cronbach's Alpha if item is excluded		Kappa with quadratic weighting (CI 95%)	Agreement classification
		Test	Retest	Test	Retest		
a	Takes medications as scheduled.	0.33	0.59	0.77	0.77	0.65 (0.16-1.00)	Substantial
b	Takes all prescribed doses.	0.35	0.23	0.77	0.80	0.44 (0.00-1.00)	Moderate
c	Follows dietary guidance.	0.44	0.47	0.76	0.78	0.61 (0.19-1.00)	Substantial
d	Attends medical appointments.	0.08	0.17	0.78	0.80	0.34 (---)	Fair
e	Exercises as recommended.	0.61	0.54	0.73	0.77	0.70 (0.40-0.99)	Substantial
f	Fits dosage schedule in routine activities.	0.45	0.56	0.76	0.77	0.74 (0.17-1.00)	Substantial
g	You and your doctor share decisions about your treatment plan.	0.58	0.79	0.74	0.73	0.43 (0.12-0.74)	Moderate
h	Complies with treatment plan without any family or friends supervision.	0.16	0.35	0.78	0.79	0.09 (---)	Slight
i	Sticks to the treatment plan without great effort.	0.24	0.33	0.78	0.80	--- (---)	Poor
j	Uses reminders for the treatment.	0.62	0.52	0.73	0.77	0.96 (---)	Almost Perfect
k	You and your doctor discuss how to comply with your treatment plan.	0.63	0.41	0.73	0.79	0.49 (0.08-0.91)	Moderate
l	Were you able to give your opinion.	0.40	0.46	0.76	0.78	0.70 (0.40-0.99)	Substantial

* English version of items was not obtained through a formal cross-cultural adaptation process. Portuguese version is as follows: (a) Toma as medicações no horário estabelecido; (b) Toma todas as doses indicadas; (c) Segue as regras da dieta; (d) Vai às consultas marcadas; (e) Realiza os exercícios físicos indicados; (f) Encaixa os horários do remédio nas atividades do seu dia a dia; (g) O (a) Senhor (a) e seu médico decidem juntos o tratamento a ser seguido; (h) Cumpre o tratamento sem supervisão de sua família ou amigos; (i) Leva o tratamento sem grandes esforços; (j) Faz uso de lembretes para realização do tratamento; (k) O (a) Senhor (a) e seu médico discutem como cumprir o tratamento; and (l) Tem a possibilidade de dar a sua opinião no tratamento que o médico prescreveu. CI = confidence interval

for the original scale (Alfonso, Vea, Ábalo, 2008).

Problems of general meaning of those items had already been identified in the process of semantic equivalence assessment (Matta, Luiza, Azeredo, 2013), which may explain the low reliability of those items.

ICC for the adapted scale indicates an almost perfect test-retest agreement, according to Landis and Koch (1977) criteria, and lands over the threshold of adequate reliability (ICC>0.7) reported for other adherence measures (Garfield *et al.*, 2011). Although kappa for some items indicates poor test-retest agreement, most items showed substantial agreement and some almost perfect agreement. We can conclude that the adapted scale has an adequate interrater reliability.

Adopting kappa as an estimate of agreement on ordinal data has important limitations, as it does not convey vital information on the structure of agreement. This information is crucial when, for example, two observers classify each individual in an ordinal scale and a low kappa value is obtained (Imaginário *et al.*, 2014; Monteiro, Tavares, Pereira, 2012). In this scenario, one loses less information by adopting ICC for continuous scale as an estimate of reliability (Sim, Wright, 2005); this was done in our study. A more detailed study of the agreement structure for each individual item would require adoption of a larger sample size, which would result in narrower confidence intervals, favoring the interpretation of the meaning of Kappa (Sim, Wright, 2005).

In general, we can state that the adapted version of the MBG questionnaire has good homogeneity, higher than the cutoff points suggested in the literature for item-total correlation and Cronbach's alpha. The questionnaire showed adequate levels of internal consistency and interrater reliability and was able to measure in a reproducible way the level of adherence to treatment in hypertensive and diabetic patients. Studies on construct validity are recommended to complete the measurement equivalence assessment between the original MBG instrument and its translated version. Furthermore, further comparison studies with clinically relevant outcomes (criterion validity) should be conducted in order to define cutoff points suitable for use in epidemiological studies and in clinical practice.

ACKNOWLEGMENTS

AUTHOR'S CONTRIBUTIONS

SRM, TBA and VLL participated in the conception, design and drafting of the manuscript. SRM performed and TBA collaborated with data review. TBA and VLL prepared the English version and supervised SRM's master's degree thesis. All authors revised the final version of the manuscript.

REFERENCES

ALFONSO, M.L.; VEA, H.D.B.; ÁBALO, J.A.G. Validación del cuestionario MBG (Martín-Bayarre-Grau) para evaluar la adherencia terapéutica en hipertensión arterial. *Rev. Cubana Salud Públ.,* v.34, n.1, 2008.

BEN, A.J.; NEUMANN, C.R.; MENGUE, S.S. Teste de Morisky-Green e Brief Medication Questionnaire para avaliar adesão a medicamentos. *Rev. Saúde Públ.*, v.46, n.2, p.279-89, 2012.

GARFIELD, S.; CLIFFORD, S.; ELIASSON, L.; BARBER, N.; WILLSON, A. Suitability of measures of self-reported medication adherence for routine clinical use: A systematic review. *BMC Med. Res. Methodol.,* v.11, n.149, 2011.

IMAGINÁRIO, S.; JESUS, S. N.; MORAIS, F.; FERNANDES, C.; SANTOS, R.; SANTOS, J.; AZEVEDO, I. Motivação para a Aprendizagem Escolar: Adaptação de um Instrumento de avaliação para o Contexto Português. *Rev. Lusófona de Educação,* v.28, n.28, p.91-105, 2014.

LANDIS, J.R.; KOCH, G.G. The measurement of observer agreement for categorical data. *Biometr.,* v.33, n.1, p.159-74, 1977.

MATTA, S.R.; LUIZA, V.L.; AZEREDO, T.B. Adaptação brasileira de questionário para avaliar adesão terapêutica em hipertensão arterial. *Rev. Saúde Públ.*,v.47, n.2, p. 292-300, 2013.

MONTEIRO, S.; TAVARES, J.; PEREIRA, A. Adaptação portuguesa da escala de medida de manifestação de bem-estar psicológico com estudantes universitários-EMMBEP. *Psicol. Saúde Doenças,* v.13, n.1, p. 66-77, 2012.

NGUYEN, T.M.U.; LA CAZE, A.; COTTRELL, N. What are validated self-report adherence scales really measuring?: a systematic review. *Brit. J. Clin. Pharmacol.,* v.77, n.3, p.427-445, 2014.

OSTERBERG, L.; BLASCHKE, T. Adherence to Medication. *New Engl. J. Med.,* v.353, n.5, p.487-497, 2005.

REICHENHEIM, M.E.; MORAES, C.L. Operacionalização de adaptação transcultural de instrumentos de aferição usados em epidemiologia. *Rev. Saúde Públ.,* v.41, n.4, p.665-73, 2007.

SIM, J.; WRIGHT, C.C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys. Ther.,* v.85, n.3, p.257-68, 2005.

STREINER, D.L.; NORMAN, G.R. Health measurement scales. *In: A practical guide to their development and use.* Oxford: Oxford University Press, 2003.

WORLD HEALTH ORGANIZATION. WHO. *Adherence to long-term therapies*: evidence for action. Geneva: WHO, 2003.