

## Reliability of screening tests for health-related problems among low-income elderly

Confiabilidade de testes de triagem para problemas de saúde de idosos pobres

Confiabilidad de las pruebas de detección de problemas de salud en ancianos pobres

Valéria Teresa Saraiva Lino <sup>1</sup>  
Margareth Crisóstomo Portela <sup>1</sup>  
Luiz Antônio Bastos Camacho <sup>1</sup>  
Nádia Cristina Pinheiro Rodrigues <sup>1,2</sup>

### Abstract

Screening tests for health problems can identify elderly people who should undergo the Comprehensive Geriatric Assessment, enabling the planning of actions to prevent disability. The aim of this study was to analyze the inter-rater reliability (IRR) of self-assessment questions (SAQ) and performance tests (PT) recommended in Brazil, in a sample of low-income elderly people, through an exploratory study performed with 165 elderly assessed by two professionals on different days. IRR was evaluated using the intraclass correlation coefficient (ICC) for continuous variables and the kappa statistic for categorical ones. The IRR for the PT (muscle strength, mobility body mass index, vision) was excellent and presented ICC values greater than 0.75. By contrast, the IRR for SAQ (urinary incontinence, self-perceived health and hearing impairment) was intermediate. Only the fall-related item presented a good IRR. In this study single SAQ had poor reliability when compared to PT, suggesting the necessity of revision of subjective self-assessment items with low reproducibility before implementation.

Triage; Self-Assessment; Aged

### Resumo

Testes de triagem de problemas de saúde podem identificar idosos aptos à avaliação geriátrica ampla, possibilitando o planejamento de intervenções, prevenindo-se o declínio funcional. O objetivo deste trabalho foi analisar a confiabilidade interavaliadores (CI) de questões de autoavaliação (QAA) e testes de desempenho (TD) recomendados no Brasil, em uma amostra de idosos pobres. Realizou-se estudo exploratório em que dois profissionais aplicaram os testes em dias diferentes a 165 idosos. Utilizou-se o coeficiente de correlação intraclassa (CCI) e a estatística kappa para as variáveis contínuas e categóricas, respectivamente. O CCI para os TD (força, índice de massa corporal, mobilidade, visão) foi excelente (> 0,75). As QAA de incontinência urinária, autopercepção de saúde e deficiência auditiva apresentaram kappa com valores intermediários. Apenas a QAA relacionada à queda teve boa CI. Neste estudo, QAA tiveram baixa CI quando comparadas a TD, sugerindo que itens subjetivos com pouca reprodutibilidade deveriam ser revisados antes da sua implementação.

Triagem; Autoavaliação; Idoso

<sup>1</sup> Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.

<sup>2</sup> Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil.

#### Correspondence

V. T. S. Lino  
Centro de Saúde Escola Germano Sinval Faria, Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz.  
Rua Leopoldo Bulhões 1480, Rio de Janeiro, RJ 21041-210, Brasil.  
valerialino@ensp.fiocruz.br

## Introduction

The increase in the proportion of Brazil's elderly population has contributed to the non-communicable disease epidemic with concomitant increases in morbidity and mortality<sup>1</sup>. The consequence of these demographic and epidemiological changes is an increase in the occurrence of disability. Older people are likely to have multiple conditions interacting in different ways, leading to difficulties in performing important tasks<sup>2</sup>. The Comprehensive Geriatric Assessment (CGA), a time-demanding multidimensional diagnostic process, is aimed at detecting the biological, psychological and social disorders that affect this group<sup>3</sup> but due to the high demand for care, it is not applicable to all elderly. A challenge is to establish a rapid examination to detect health problems which could direct the application of CGA and postpone disability.

In Brazil, the Ministry of Health recommends the use of the Rapid Multidimensional Evaluation of Elderly Persons in primary care, containing items related to multiple capabilities<sup>4</sup>. However, most questions have not been validated, and some of them address matters whose appropriateness is questionable for a screening test.

This work is concerned with identifying some rapid and reliable screening tests already applied in Brazil to select those elderly patients that should be submitted to the CGA.

The reliability of measurements may vary depending on the context<sup>5</sup>, differing among population subgroups. Furthermore, the higher subjectivity of the test, the greater the possibility of variation in the interpretation of the results by different observers<sup>6</sup>.

Due to the importance of psychometric and sociocultural considerations when administering an instrument this study aimed to assess the inter-rater reliability (IRR) of health-related tests in a low-income elderly community.

## Methods

### Study and sample

This research was part of an exploratory study aimed at developing a strategy of rapid assessment of the elderly. It was performed at the primary healthcare unit of the Oswaldo Cruz Foundation (Fiocruz), in Manguinhos, in the city of Rio de Janeiro, Brazil. In Manguinhos, most houses have a single room, monthly family income is usually lower than a the minimum wage, and more than 50% of residents have no more than an elementary school education<sup>7</sup>.

The non-probabilistic sample was drawn from users that were 60 years or older who received care from the Family Health Team. Individuals with advanced cognitive and sensorial deficits or impaired locomotion were excluded. The sample size was calculated using the prevalence of depression, estimated at 20% in the elderly, as a reference<sup>8</sup>. A kappa coefficient of 0.6 with a 95% confidence interval was used to generate a conservative sample size of 180 individuals, estimated using the WinPepi application, version 2009 (<http://www.brixtonhealth.com/pepi4windows.html>).

### Procedures

Data collection occurred from June to December 2010. Three health professionals were responsible for the standardization of the techniques used in the study during a three-hour meeting. Assessments were independently administered in two sessions. First a geriatrician applied the CGA and seven to 15 days later, either a psychomotor specialist or a social worker applied the tests whose IRR would be assessed. This interval aimed to avoid memory bias in favor of higher reliability and not to exhaust the patient after the long application of CGA.

Both assessments included rapid performance tests (PT) and self-assessment questions (SAQ) recommended for screening of health problems in Brazil<sup>9,10</sup> (Figure 1). Self-perceived health (SPH) is a measure of health associated with disability in the elderly<sup>11</sup> and could be indicative to referring them to the CGA.

### Statistical analysis

IRR was evaluated using the intra-class correlation coefficient (ICC) for continuous variables and the

Figure 1

Performance tests (PT) and self-assessment questions (SAQ).

PT	SAQ
Mobility (timed up and go) <sup>11</sup>	Urinary incontinence
Vision (Snellen scale)	Hearing
Strength (grasping force)*	Self-perceived health
Body mass index	Falls

\* Grasping force was assessed with a Crown dynamometer (dinamômetro Crown manual, Oswaldo Filizola), using Fried et al.'s<sup>20</sup> criteria, utilizing the best result of two trials obtained with the dominant hand.

kappa statistic for categorical variables, adjusting for prevalence and asymmetries with the Prevalence-Adjusted Bias-Adjusted Kappa (PABAK) technique<sup>12</sup>. The classification of IRR followed the recommendations of Landis and Koch: kappa values greater than 0.75 or below 0.40 represent excellent or poor agreement, respectively<sup>13</sup>. Values between these levels denote intermediate to good agreement. We adopted the same rule for the ICC. SPSS version 13 (SPSS Inc., Chicago, U.S.A.) was used to perform statistical analyses.

### Study ethics

The Ethics Research Committee of the Sergio Arouca National Public Health School/Fiocruz (report number 126/10) approved the research. All participants signed an informed consent form, which pledges anonymity and confidentiality of the information.

### Results

The first and the second sessions evaluated 185 e 165 individuals respectively. Five were excluded due to visual or cognitive impairment. No significant differences in sociodemographic characteristics were detected between the elderly who completed the study and those who

did not. The majority of participants were female (73.0%) and there was a slight majority of single or widowed participants (Table 1).

For the PT items, IRR was excellent ( $K > 0.75$ ). By contrast, for SAQ items IRR was intermediate ( $0.75 > K > 0.40$ ), except for the fall-related item, in which IRR was good (Table 2).

### Discussion

The IRR between the PT items were excellent, unlike measures assessed by the SAQ. The possibility of low reliability for subjective measures can occur, despite adequate training of examiners. Sociocultural factors, psychological issues, memory lapses and lack of insight of informants lead to variation in how people communicate information about symptoms<sup>14</sup>. In addition, the reliability of the information may be compromised when there is no socially acceptable environment in which to talk about intimate questions<sup>15,16</sup>. The perception of "invasion of privacy" may lead the respondents to admit a problem in a first interview but deny it in the second one<sup>5</sup>.

All the above factors may have influenced our results. The SAQ are inherently subjective and susceptible to communication problems. The CGA is more likely to generate a context of confidence between the professional and the individual

Table 1

Sociodemographic data and health problems identified in the geriatric assessment (N = 180).

Variable	Estimate
Age in years: mean (SD)	73.1 (7.0)
Sex (female): n (%)	132 (73.0)
Years of education: mean (SD)	3.0 (2.9)
Married: n (%)	81 (45.0)
Mobility-TUG: mean (SD)	12.5 (9.2)
BMI (kg/m <sup>2</sup> ): mean (SD)	28.7 (5.4)
Two or more falls in the last 12 months: n (%)	42 (23.0)
Hearing loss: n (%)	81 (45.0)
Vision, Snellen scale (less than 0.5 for both eyes): n (%)	82 (46.0)
Strength below the reference value: n (%)	
Female	37 (28.0)
Male	20 (41.0)
Urinary incontinence: n (%)	40 (22.0)
SPH: n (%)	
Good	100 (56.0)
Regular	60 (33.0)
Bad	20 (11.0)

BMI: body mass index; SD: standard deviation; SPH: self-perceived health; TUG: timed up and go.

Table 2

Reliability of screening items (N = 165).

Variable	ICC (95%CI)	Kappa (95%CI)	Weighted kappa	Simple concordance (%)	Level of agreement
Performance tests					
Strength	0.89 (0.85-0.92)				Excellent
TUG	0.81 (0.68-0.88)				Excellent
BMI	0.94 (0.92-0.96)				Excellent
Vision	0.87 (0.82-0.91)				Excellent
Questions					
SPH		0.48 (0.36-0.60)	0.54 (0.40-0.68)	71.0	Intermediate
Falls		0.60 (0.46-0.75)		92.1	Good
Urinary incontinence		0.49 (0.34-0.65)		81.2	Intermediate
Hearing loss		0.54 (0.42-0.67)		77.4	Intermediate

95%CI: 95% confidence interval; BMI: body mass index; ICC: intraclass correlation coefficient; SPH: self-perceived health; TUG: timed up and go.

examined allowing him or her to assume difficulties such as urinary incontinence. Additionally, it is important to underline that our study was conducted with elderly people with little or no formal education, in which age and educational factors influence cognitive performance<sup>11</sup>.

The screening for hearing impairment with the use of a single question has been recommended in Brazil based on studies that have evaluated the sensitivity and specificity of the single item<sup>10,17</sup>, but only one study has examined its IRR and found a kappa coefficient of 0.65<sup>18</sup>. In relation to the SPH, two studies have assessed only test-retest reliability, but even not examining IRR these studies have identified discrepancies between assessments in racial and ethnic minorities, individuals with lower levels of education and the elderly<sup>11,19</sup>. Regarding urinary incontinence, the social stigma related to the problem may have contributed to our results<sup>15</sup>. Finally, falls, in the life of an elderly person lead to functional decline, what justifies more precise information about them, even by

individuals with low levels of education. In our study, it was the only subjective question with good IRR.

Our study has limitations. It is known that observers tend to change their approaches over time so that planning a periodical replication is necessary<sup>5</sup>. This has not occurred here. Furthermore, we assessed elderly people with low levels of education in a primary care setting and our results should be applicable only to similar populations.

Aging often results in insidious changes in functional capacity. Screening for health problems allow examining a large portion of elderly, indicating those that should be submitted to CGA. Our study revealed that single SAQ addressing SPH, urinary incontinence, and hearing loss had poor reliability in older adults. Although high item reproducibility does not guarantee high accuracy, it is clear that subjective self-assessments with low reliability should be reviewed before implementation.

## Resumen

*Las pruebas de detección de problemas de salud identifican a ancianos que deben someterse a una revisión médica completa en Geriatría, lo que permite la planificación de acciones de prevención de discapacidades. Se analizó la fiabilidad entre evaluadores (FEA) de preguntas de autoevaluación (PA) y pruebas de rendimiento (PR) que se recomiendan en Brasil, en un estudio exploratorio realizado con 165 ancianos de baja renta, evaluados por dos profesionales en diferentes días. La FEA se evaluó mediante el coeficiente de correlación intraclassa (CCI) para las variables continuas y el índice kappa para las categóricas. La FEA para el PR (fuerza muscular, índice de masa corporal de movilidad, visión) era excelente y presentan CCI con valores superiores a 0,75. Por el contrario, la FEA para las PA (incontinencia urinaria, salud autopercebida y auditiva) fue intermedia. Las PA sobre caídas presentaron una buena FEA. La escasa fiabilidad de las PA sugiere la necesidad de una revisión de los elementos subjetivos de autoevaluación con baja reproducibilidad antes de la implementación.*

*Triage; Autoevaluación; Anciano*

## Contributors

V. T. S. Lino and M. C. Portela contributed in the conception, design, drafting and revising the manuscript. L. A. B. Camacho and N. C. P. Rodrigues contributed to the conception, design, data analysis and revising the manuscript.

## Acknowledgments

The authors wish to thank Maria José Barbosa Lima, Mônica Bastos de Lima Barros and Soraya Atiê for their contributions in the critical revision of the text prior to publication. We would also like to thank Flavio Henrique Lino for assistance in writing the article.

## References

- Schmidt MI, Duncan BB, Azevedo e Silva G, Menezes AM, Monteiro CA, Barreto SM, et al. Chronic non-communicable diseases in Brazil: burden and current challenges. *Lancet* 2011; 377:1949-61.
- Sousa RM, Ferri CP, Acosta D, Guerra M, Huang Y, Jacob K, et al. The contribution of chronic diseases to the prevalence of dependence among older people in Latin America, China and India: a 10/66 Dementia Research Group population-based survey. *BMC Geriatr* 2010; 10:53.
- Rubenstein LZ, Alessi CA, Josephson KR, Trinidad Hoyl M, Harker JO, Pietruszka FM. A randomized trial of a screening, case finding, and referral system for older veterans in primary care. *J Am Geriatr Soc* 2007; 55:166-74.
- Ministério da Saúde. Avaliação multidimensional rápida da pessoa idosa. In: Departamento de Atenção Básica, Secretaria de Atenção à Saúde, Ministério da Saúde, organizador. *Envelhecimento e saúde da pessoa idosa*. Brasília: Ministério da Saúde; 2007. p. 48-9.
- Hasselmann MH, Lopes CS, Reichenheim ME. Measurement reliability in a study on family violence and severe acute malnutrition. *Rev Saúde Pública* 1998; 32:437-46.
- Gordis L. *Epidemiologia*. 4ª Ed. Rio de Janeiro: Revinter; 2010.
- Carvalho MAP, Pivetta F. The integrated territory of health care in Manguinhos: we are all apprentices. Rio de Janeiro: Fundação Oswaldo Cruz; 2012.
- Li C, Friedman B, Conwell Y, Fiscella K. Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. *J Am Geriatr Soc* 2007; 55:596-602.
- Pereira LSM, Gomes G. Avaliação funcional. In: Guimarães RM, Cunha UGV, editores. *Sinais e sintomas em geriatria*. São Paulo: Atheneu; 2004. p. 17-30.
- Valete-Rosalino CM, Rozenfeld S. Auditory screening in the elderly: comparison between self-report and audiometry. *Braz J Otorhinolaryngol* 2005; 71:193-200.
- Zajacova A, Dowd JB. Reliability of self-rated health in US adults. *Am J Epidemiol* 2011; 174:977-83.
- Abramson J. WINPEPI (PEPI-for-Windows): computer programs for epidemiologists. *Epidemiol Perspect Innov* 2009; 1:6.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-74.

14. Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Methods Med Res* 1998; 7:301-17.
15. Shaw C, Tansey R, Jackson C, Hyde C, Allan R. Barriers to help seeking in people with urinary symptoms. *Fam Pract* 200; 18:48-52.
16. Kraemer HC. Measurement of reliability for categorical data in medical research. *Stat Methods Med Res* 1992; 1:183-99.
17. Nondahl DM, Cruickshanks KJ, Wiley TL, Tweed TS, Klein R, Klein BEK. Accuracy of self-reported hearing loss. *Audiology* 1998; 37:295-301.
18. Tomioka K, Ikeda H, Hanaie K, Morikawa M, Iwamoto J, Okamoto N, et al. The Hearing Handicap Inventory for Elderly-Screening (HHIE-S) versus a single question: reliability, validity, and relations with quality of life measures in the elderly community, Japan. *Qual Life Res* 2013; 22:1151-9.
19. Crossley TE, Kennedy S. The reliability of self-assessed health status. *J Health Econ* 2002; 21:643-58.
20. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci* 2001; 56:M146-56.

---

Submitted on 14/Feb/2014  
Final version resubmitted on 01/Sep/2014  
Approved on 22/Sep/2014