Screening for Voice Disorders in Older Adults (RAVI)—Part III: Cutoff Score and Clinical Consistency

***Leandro Pernambuco, †‡§Albert Espelt, and ¶Kenio Costa de Lima,** **Paraíba, Brazil, †Barcelona, Spain, ‡Bellaterra, Spain, §Madrid, Spain and ¶Rio Grande do Norte, Brazil*

Summary: Aim. The aim of the study was to determine the cutoff score and clinical consistency of "Screening for Voice Disorders in Older Adults" (RAVI—Rastreamento de Alterações Vocais em Idosos).
Study Design. This is a prospective, nonrandomized, cross-sectional diagnostic study.
Methods. A sample of 301 subjects, including both sexes, aged 60 and more, and all of whom were living in either a community or an institution, was studied. To determine which subjects had or did not have voice problems, we used a composite reference standard (auditory-perceptual analysis of sustained vowel phonation, auditory-perceptual analysis of connected speech, and vocal self-assessment). The best cutoff score was identified using the receiver operating characteristic (ROC) curve analysis. The clinical consistency indicators were co-positivity, co-negativity, positive and negative predictive values, positive and negative likelihood ratio, and test efficiency. The significance level was 5%.
Results. The area under the ROC curve was 0.763 (95% confidence interval: 0.706–0.821), and the best cutoff score for determining which older adults had or did not have a voice disorder was 2. All clinical consistency indicators were satisfactory: co-positivity (79%), co-negativity (60%), predictive positive value (51%), negative predictive value (84%), positive likelihood ratio (2.01), negative likelihood ratio (0.34), and test efficiency (69%).
Conclusions. RAVI has satisfactory indicators of clinical consistency and is able to determine which older adults have voice disorders by a cutoff score of 2. The use of RAVI as a screening tool is recommended to help determine the prevalence of voice disorders in older adults.
Key Words: voice disorders–aging–epidemiology–validation studies–diagnostic studies.

INTRODUCTION

Voice disorders (VD) in older adults can lead to a lack of communication efficiency.[1–4] This lack of efficiency has a negative impact on their functionality and quality of life, and is associated with difficulties in socialization, autonomy, and well-being.[1–4] Therefore, older adults with VD often experience feelings of intolerance, discrimination, and isolation; they often have psychological disorders, such as anxiety or depression.[1–8]

Despite the known evidence that VD has a negative functional and psychosocial impact on older adults, the effect of this health condition on public health is still underestimated. The available clinical assessment procedures depend on auditory-perceptual training, access to technology, increased lead time, and more complex analysis capabilities,[9,10] which are difficult to reproduce in the epidemiological context.

A recent systematic review[11] reported the following: (1) many studies determined the prevalence of VD in older adults based on data from individuals who sought or had access to health care; (2) older people do not seek treatment for VD because of either misinformation regarding the treatment or accepting the health condition as part of the natural aging process; and (3) tools with psychometric properties were unavailable, and clinical consistency indicators and adequate discriminatory power to diagnose VD at the epidemiological level were unavailable for older adults.

To determine which individuals have or do not have VD,[12,13] some studies suggest the use of voice-related self-assessment questionnaires.[14–19] However, these tools are limited because they are not designed for the specificities of VD in the older adult population and were not originally meant for diagnostic purposes. Additionally, these questionnaires generally have questions that can only be answered by people who perceive some degree of VD in themselves or who have already been diagnosed with VD. To be used for screening purposes and to ensure the interpretations generated by the results of these tools are equal, valid, and reliable, it would be necessary to provide the appropriate psychometric adaptations, review the entire validation process before setting cutoffs, and reproduce them with the purpose of screening.[20] However, none of the current tools accomplish this.

To address this, we developed the "Screening for Voice Disorders in Older Adults" (RAVI—Rastreamento de Alterações Vocais em Idosos),[21,22] the only available tool to date that facilitates epidemiological diagnosis of VD in older adults. The RAVI allows researchers and clinicians to determine the prevalence of this health condition in large populations and to early screen those individuals who need referral to other procedures to confirm or refute the VD diagnosis.[21,22]

RAVI is a fast, risk-free, low-cost, and easy-to-use tool that has adequate evidence of validity and reliability.[21,22] However, for RAVI to be utilized in population surveys and health services, it is necessary to obtain clinical consistency indicators and determine its discriminatory power through cutoff scoring. To determine these properties, it is necessary to compare RAVI scores with the results of VD diagnosis reference standards. Therefore,

Accepted for publication March 4, 2016.
From the *Department of Speech, Language and Hearing Sciences, Universidade Federal da Paraíba (UFPB), Paraíba, Brazil; †Agència de Salut Pública de Barcelona (ASPB), Servei d'Atenció i Prevenció a les Drogodependències, Barcelona, Spain; ‡Departament de Psicobiologia i Metodologia de les Ciències de la Salut, Facultat de Psicologia, Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain; §CIBER de Epidemiologia y Salud Pública, Instituto de Salud Carlos III, Madrid, Spain; and the ¶Programa de Pós-graduação em Saúde Coletiva (PPGSCol-UFRN), Universidade Federal do Rio Grande do Norte (UFRN), Rio Grande do Norte, Brazil.
Address correspondence and reprint requests to Leandro Pernambuco, Jardim Universitário, S/N, Castelo Branco, João Pessoa, PB, 58051-900, Brazil. E-mail: leandroape@globo.com
Journal of Voice, Vol. 31, No. 1, pp. 117.e17–117.e22
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.03.003

the aim of this study was to determine the discriminatory power of RAVI through cutoff scoring and investigate the clinical consistency indicators.

METHODS

This is a prospective, nonrandomized, and cross-sectional diagnostic study. Sample size calculation considered a prevalence of 29.1% of VD in older adults[1] and a 95% confidence interval (CI) range ≤10%, which resulted in an expected 350 individuals. Because the prevalence of VD in the elderly has significant variability across previous studies,[11] the sample size was also monitored by the stabilization of a sensitivity indicator throughout data collection, which resulted in a final sample size of 301 older adults.

The samples had the following characteristics: female (n = 211, 70.1%), aged between 60 and 79 years (n = 260, 86.4%), residents in the metropolitan area of a northeastern Brazilian city (n = 236, 78.4%), low level of education (n = 210, 69.8%), and without private health insurance (n = 239, 79.4%).

Because no single diagnostic test is considered the reference standard in the field of voice diagnosis,[21] clinical consistency was investigated by comparing the RAVI result with a composite reference standard[23,24] by evaluating the auditory-perceptual analysis of sustained [ε] vowel phonation, auditory-perceptual analysis of connected speech (counting 1 to 20), and vocal self-assessment.

The recording of voice samples took place in a quiet environment (≤50 dB), with the volunteer sitting comfortably in a chair. To capture samples, the researcher positioned a microphone headset (Logitech, Model H10, Newark, CA, USA) at a distance of 10 cm from the mouth of the subject with a 45° pickup angle. The microphone was connected to a laptop computer (Lenovo, Model G470, Itu, SP, Brazil), and the samples were recorded with *PRAAT* software version 5.4.08 (available http://www.praat.org), at a sample rate of 44,100 Hz, mono, and saved as a waveform file (WAV). Later, the vocal samples were edited using the *Audacity* software (version 2.0.5, available at http://audacity.sourceforge.net/download/windows). Three seconds of the central emission of the sustained vowel sound was kept to be considered the moment of greater phonation stability. In addition, the reduced time of vowel emission helps avoid inattention or judges' fatigue and lessens the chance of distraction by signal redundancy.[25] For normalization of the sustained vowel and connected speech samples, *Audacity* normalize function was used to keep the peak levels between −6 and 6 dB.

The vocal samples underwent auditory-perceptual analysis based on the vocal deviation scale (VDS).[26,27] The VDS uses a visual analog scale consisting of a 100-mm horizontal line, on which judges rate the global impression of the overall severity of voice deviation.[23,24] The extreme left side of the 100-mm horizontal line represents the absence of vocal deviation and the extreme right represents the maximum vocal deviation.[26,27]

These judgments were made independently by three voice specialists for over 5 years; they did not know the identity of the subject. To assess intra-observer agreement, 61 (20.26%) subjects were randomized and their voice samples were re-evaluated without the judges' knowledge. To aid in the reliability of the VDS evaluation, two additional voice-specialist judges defined four anchor stimuli according to each task and sex, considering vocal markers such as accent, vocal tension, and focus resonance. These two judges did not participate in the VDS evaluation.

The score obtained using VDS (continuous variable) was converted into the following categories (ordinal variable) according to specific cutoff values for Brazilian older adults:[27] normal variability of vocal quality (values ranging from 0 mm to 35.6 mm), mild to moderate vocal deviation (from 35.7 mm to 51.1 mm), moderate vocal deviation (from 51.2 mm to 74.3 mm), and intense vocal deviation (from 74.4 mm to 100 mm).

The final classification of each voice sample was obtained by consensus between the three assessments in accordance with the following criteria: similar category in the three evaluations, most frequent category in the three evaluations, or intermediate category when the three evaluations were different. The intra-observer agreement was calculated using the interclass correlation coefficient (ICC).

The vocal self-assessment asked the question: "How do you rate the quality of your usual voice, day to day?" The following were the possible answers: excellent, very good, good, fair, or bad.[28,29] For analysis purposes, the "fair" and "bad" categories were classified as "negative evaluations" and the others as "positive assessments" (author's criteria).

The subjects were classified into the group with voice disorders (WVD) or without voice disorder (WOVD) according to the combinations of clinical criteria that generated the composite reference standard (Figure 1).

The discriminatory power of the RAVI was determined by analysis of the receiver operating characteristic (ROC) curve, which plotted the relationship between the true positive and false positive numbers. The authors analyzed the area under the curve (AUC), whose value varied between 0 and 1, with better results being closer to 1. Through the ROC curve analysis, it was possible to identify the best RAVI cutoff score for determining which older adults had VD. The best cutoff score is usually the one that promotes a better balance between true positives and true negatives, but according to the instrument's main goal it may be necessary to prioritize the true positives (screening tools) or true negatives (diagnostic confirmation tools).[12,30] Because RAVI is a screening tool, it prioritizes the best result related to the number of true positives.

Because of the absence of a reference standard for vocal assessment, this study adopted the term "clinical consistency" rather than accuracy, as well as co-positivity and co-negativity rather than sensitivity and specificity, respectively, as recommended by the literature.[31,32] In addition to co-positivity and co-negativity, the other clinical consistency indicators calculated were positive and negative predictive values, positive and negative likelihood ratios, and test efficiency. The significance level was 5%.

RESULTS

Figure 2 demonstrates the distribution of responses to each of the RAVI's questions. Dry throat and phlegm in the throat were the most frequently mentioned, followed by itchy throat and a general feeling of discomfort caused by the voice. These