Comparison of the Pediatric Voice Handicap Index With Perceptual Voice Analysis in Pediatric Patients With Vocal Fold Lesions

Kaalan Johnson, MD; Susan Baker Brehm, PhD, CCC-SLP; Barbara Weinrich, PhD, CCC-SLP; Jareen Meinzen-Derr, PhD; Alessandro de Alarcon, MD, MPH

Objective: To compare a subjective patient/family-derived voice handicap survey with an expert observer-derived method of evaluating voice disturbance in pediatric patients with vocal fold lesions (VFLs).

Design: Retrospective review.

Setting: Tertiary care referral center.

Patients: Thirty-eight children with VFLs referred for voice evaluation.

Main Outcome Measures: Pediatric Voice Handicap Index (pVHI) scores and Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) scores. Percentages for CAPE-V (100-point scale) and pVHI (92-point scale) were calculated for direct comparisons. Relationships between pVHI scores and CAPE-V scores were investigated using the Spearman rank correlation.

Results: Thirty-eight patients with VFLs (median age, 8.3 years; age range, 4.2-17.2 years; 63% males) were included from a database of more than 600 children and evaluated between November 15, 2005, and June 15, 2010. The median CAPE-V overall score was 30.3 (range, 1-67), and the normalized total pVHI score was 29.3 (range, 0-73) (P=.90). The Spearman rank correlation showed significant fair correlations between CAPE-V overall and functional pVHI and between CAPE-V strain and breathiness, and the pVHI total, functional, but none higher than p=0.44 (P=.03). The correlation was higher in males for CAPE-V loudness to total pVHI (p=0.40, P=.04) and in females for CAPE-V breathiness (p=0.58, P=.03) and strain (p=0.55, P=.04) to total pVHI.

Conclusions: The CAPE-V and the pVHI are useful tools in the measurement of voice outcomes in children with VFLs. There are fair correlations between the CAPE-V and the pVHI, and they likely evaluate important yet different aspects of voice disturbance. Significant gender differences in these correlations should be further evaluated in future studies.


VOCAL FOLD NODULES ARE widely considered the most common physical examination finding associated with pediatric dysphonia, affecting 5% to 35% of children.1,2 Vocal fold nodules may comprise up to 40% of a tertiary pediatric referral voice practice.3 The resulting dysphonia may have a significant effect on patient quality of life,4 and, thus, accurate evaluation and effective treatments are of paramount importance.

Voice assessment measures are often used during the evaluation of pediatric dysphonia and generally include aerodynamic and acoustic measures, videostroboscopy and endoscopy, clinician-derived perceptual assessment measures (such as the GRBAS [grade, roughness, breathiness, asthenia, strain] and the Consensus Auditory-Perceptual Evaluation of Voice [CAPE-V]), and patient- or proxy-derived measures of dysphonia impact (such as the Pediatric Voice Outcome survey, the Pediatric Voice-Related Quality-of-Life Survey, and the Pediatric Voice Handicap Index [pVHI]).5-9 Establishing a single criterion standard outcome measure in voice assessment has been difficult because many of these measures assess different aspects of the voice disorder. Several research groups have looked at the relationship between the perceptual evaluation and the patient-perceived effect on voice-related quality of life.10 We have explored this relationship by evaluating the roles of the pVHI and the CAPE-V in patients with voice disturbance after airway reconstruction, where only fair correlations were demonstrated between clinician-based (CAPE-V) and patient-(or proxy-) derived (pVHI) measures. To our knowl-
edge, this study is the only examination of this relationship in pediatric patients. A similar evaluation in children with dysphonia secondary to vocal fold nodules has not been performed.

The objective of this study was to examine the relationship between a subjective proxy-derived voice assessment measure (pVHI) and an expert observer–derived method of evaluating perceptual voice disturbance (CAPE-V) in pediatric patients with vocal fold nodules or vocal fold lesions (VFLs).

**METHODS**

**PATIENTS**

We performed a retrospective database review of prospectively gathered data on pediatric patients (<18 years old) presenting to the Center for Pediatric Voice Disorders at Cincinnati Children’s Hospital Medical Center, Cincinnati, Ohio, and obtaining a diagnosis of bilateral vocal nodules or bilateral VFLs between November 15, 2005, and June 15, 2010. Bilateral VFLs were defined as symmetrical or near-symmetrical, benign, or true VFLs (which may consist of nodules, small cysts, or reactive lesions; often difficult to distinguish clinically in our experience). Patients with papillomas, reflux laryngitis, obvious polyps, large cysts, or other clearly defined vocal fold abnormalities were excluded. Patients were excluded if they were 18 years or older or if their medical record contained inadequate data for comparison. This study was approved by the institutional review board at Cincinnati Children’s Hospital Medical Center.

**DATA SOURCES**

Deidentified patient data reviewed from the voice database included demographics, primary and secondary diagnoses, historical data on smoke exposure and hydration status, and pVHI and CAPE-V scores from the initial encounter of each patient at the voice clinic. Initial clinical CAPE-V data were obtained and recorded by the evaluating clinician in accord with the clinical protocol for CAPE-V utilization published by Kempster et al12 in 2009. For this study, audio recordings of these evaluations were later deidentified and randomized to allow for blinded evaluation by 2 licensed/certified speech-language pathologists (S.B.B. and B.W.) with extensive experience in voice disorders. The rating values were averaged to obtain final CAPE-V data for analysis, similar to previous CAPE-V research protocols used by Solomon et al.13 Data for internal consistency were recorded for the 2 raters by rescoring a random sample of 30% of the voice recordings for intrarater evaluation using the CAPE-V more than 1 week after the initial ratings.

**STATISTICAL ANALYSIS**

Continuous variables are reported as median, range, and interquartile range. Categorical variables are reported as absolute numbers and proportions. Because the CAPE-V overall score is recorded on a 100-point visual analog scale and the total pVHI score is recorded on a 92-point scale, total pVHI values were normalized to a percentage by dividing the total score by 92. These normalized values were then used for comparisons between total pVHI and CAPE-V overall scores using the Wilcoxon signed rank test. Correlations between overall categories and subgroup domains were examined using the Spearman rank correlation, with statistical significance for all analyses set at P = .05. Statistical analysis was performed using a commercially available software program (SAS version 9.1; SAS Institute, Inc, Cary, North Carolina).

**RESULTS**

Fifty-five patients with VFLs were identified from a database of more than 600 children between January 1, 2004, and December 31, 2010 (Table 1). Two patients were excluded for having grossly incomplete data and 1 for being 18 years or older. Fourteen patients were excluded because they lacked just 1 vocal assessment measure (ie, they were not cooperative with providing CAPE-V sentences). The remaining 38 patients had a median age of 8.3 years (age range, 4.2-17.2 years), with 63% (n=24) being males and 21% (n=8) having a medical diagnosis recorded, most commonly gastroesophageal reflux disease (11%, n=4). Correlations between voice assessment measures and medical diagnoses were not analyzed because the diagnoses were purely historical, with relatively small numbers in each group.

**CAPE-V AND pVHI COMPARISONS**

No significant difference was noted between the median CAPE-V overall score of 30.3 (range, 1-67) and the nor-
Table 3. Correlations Between CAPE-V Domains and pVHI Domains

<table>
<thead>
<tr>
<th>CAPE-V Domain</th>
<th>Total</th>
<th>Functional</th>
<th>Physical</th>
<th>Emotional</th>
<th>Talkative</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.27</td>
<td>0.38</td>
<td>0.09</td>
<td>0.31</td>
<td>0.13</td>
<td>0.23</td>
</tr>
<tr>
<td>Roughness</td>
<td>0.10</td>
<td>0.02</td>
<td>0.60</td>
<td>0.08</td>
<td>0.46</td>
<td>0.18</td>
</tr>
<tr>
<td>Breathiness</td>
<td>0.28</td>
<td>0.36</td>
<td>0.10</td>
<td>0.31</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>Strain</td>
<td>0.08</td>
<td>0.02</td>
<td>0.50</td>
<td>0.06</td>
<td>0.50</td>
<td>0.24</td>
</tr>
<tr>
<td>Pitch</td>
<td>0.40</td>
<td>0.44</td>
<td>0.21</td>
<td>0.44</td>
<td>0.02</td>
<td>0.24</td>
</tr>
<tr>
<td>Strain</td>
<td>0.36</td>
<td>0.34</td>
<td>0.25</td>
<td>0.29</td>
<td>0.18</td>
<td>0.26</td>
</tr>
<tr>
<td>Pitch</td>
<td>0.04</td>
<td>0.005</td>
<td>0.20</td>
<td>0.006</td>
<td>0.90</td>
<td>0.16</td>
</tr>
<tr>
<td>Strain</td>
<td>0.01</td>
<td>0.005</td>
<td>0.20</td>
<td>0.006</td>
<td>0.90</td>
<td>0.16</td>
</tr>
<tr>
<td>Pitch</td>
<td>0.11</td>
<td>0.18</td>
<td>-0.05</td>
<td>0.21</td>
<td>0.32</td>
<td>-0.08</td>
</tr>
<tr>
<td>Strain</td>
<td>0.50</td>
<td>0.26</td>
<td>0.80</td>
<td>0.20</td>
<td>0.06</td>
<td>0.60</td>
</tr>
<tr>
<td>Loudness</td>
<td>0.14</td>
<td>0.09</td>
<td>0.08</td>
<td>0.19</td>
<td>0.10</td>
<td>-0.04</td>
</tr>
<tr>
<td>Total</td>
<td>0.40</td>
<td>0.80</td>
<td>0.26</td>
<td>0.13</td>
<td>0.60</td>
<td>0.80</td>
</tr>
</tbody>
</table>

The Spearman rank correlation showed significant fair correlations between CAPE-V overall and the functional domain of pVHI (ρ=0.38, P=0.02) and between total pVHI and the breathiness (ρ=0.40, P=0.01) and strain (ρ=0.36, P=0.02) domains of the CAPE-V (Table 3). Domain comparisons within the scales showed significant fair correlations between the functional domain of the pVHI and the CAPE-V overall, roughness, breathiness, and strain domains and between the emotional domain of the pVHI and the CAPE-V breathiness domain (ρ=0.34-0.44, P≤0.03).

Age was not significantly correlated with CAPE-V or pVHI totals or domains (P>0.2 for all); however, CAPE-V loudness was negatively correlated with the age of the child (ρ=−0.32; P=0.03). Only 12 participants (32% of the total) were 12 years or older. There were no significant differences (P>0.3) noted in any of the assessment scores when stratifying by these age groups (younger vs older).

Table 2 for complete comparisons.

The intraclass correlation coefficient for the same rater at 2 different time points indicated a high degree of repeatability regarding the different domains. The domains of strain and loudness had the highest coefficients (0.87), followed by pitch (0.84), roughness (0.79), and breathiness (0.62). The overall score had an intraclass correlation coefficient between the 2 time points (same rater) of 0.77.

The intraclass correlation coefficients between the 2 different CAPE-V raters indicated a slightly lower degree of repeatability. The domain of loudness had the highest coefficient (0.80), followed by strain (0.73) and roughness (0.67). Pitch and breathiness had the lowest coefficients (0.36 and 0.26, respectively). The overall score had an intraclass correlation coefficient between the 2 raters of 0.62.

GENDER-RELATED COMPARISONS

Correlation of the 2 scales was notably different between the genders, with the highest correlations noted on gender-specific comparisons (Figure and Table 4). For female patients, CAPE-V overall correlated with functional pVHI (ρ=0.56, P=0.04), and total pVHI correlated with breathiness and strain (ρ=0.55-0.58, P≤0.04). CAPE-V breathiness had the highest correlations with pVHI, reaching significance in correlation with total pVHI and all its domains except physical (ρ=0.53-0.62, P≤0.05). In males, significant correlations were seen between CAPE-V loudness and the total and emotional pVHI domains (ρ=0.40-0.44, P=0.04).

The assessment of dysphonia in pediatric patients with VFLs remains a significant challenge for the pediatric voice professional. There is no single criterion standard measure for voice outcomes assessment. Multiple scales are available for the evaluation of patient-derived voice-related quality of life and clinical-derived perceptual evaluation of voice.
Each of these tools carries its own inherent strengths and deficiencies that must be accounted for when using them in a complete voice evaluation. In our current practice, the CAPE-V seems to provide a reliable measure of clinician-derived perceptual voice assessment, which did not differ significantly regarding severity in patients with VFLs from a validated proxy-derived measure of voice-related quality of life/handicapping (pVHI), and this remained true when analyzed separately by gender. This is in contrast to the airway reconstruction voice population, which showed a significantly higher CAPE-V overall percentage than normalized total pVHI percentage (50.5% vs 32.6%, P < .001).11

Contrasting the airway reconstruction population provides an excellent illustration of how significant long-standing voice disturbances may be perceived less severely by the patient or proxy than relatively mild voice disturbances in an otherwise healthy patient population.

Overall correlations between clinician-derived CAPE-V assessments and proxy-reported pVHI assessments in the whole study population were fair (ρ = 0.34-0.44), which is consistent with the only study, to our knowledge, comparing clinician-derived and patient-derived voice measures in adults10 (ρ = 0.34-0.64) and with a previous initial report of this finding in children after airway reconstruction (ρ = 0.32-0.53).11 Given the good intraclass correlation data for the CAPE-V, these low overall correlations suggest significant variability in the perspective and priorities of the patient proxy reporting the pVHI. A reliable, relatively objective clinician-derived measure of voice disturbance may, in some cases, be substantially different from the patient- or proxy-derived measure given the context and priorities of a specific patient. This underscores the importance of understanding these differences when formulating an accurate vocal assessment and therapeutic recommendation for the dysphonic patient.

**GENDER-RELATED COMPARISONS**

Correlations between clinician-derived and proxy-reported scales improved in specific domains when analyzed by gender. When male patients were analyzed separately, no significant correlations were found between the CAPE-V and the pVHI except for CAPE-V loudness, which reached significance with a correlation coefficient of ρ = 0.40 (P = .04). The CAPE-V loudness domain correlated most strongly with the emotional domain of the pVHI. All other correlations between scales in male patients did not approach significance and, in some cases, showed negative correlation coefficients. This finding suggests that patient proxy reports of abnormalities in the volume of voice production (loudness) are more congruent with clinician (and perhaps societal) expectations of the male voice. Conversely, all other measures of male vocal quality seem to carry significantly different expectations of perceived normalcy between patient proxy report and clinician assessment.

In female patients, the correlations were moderate and notably higher than in male patients, with significant correlations ranging from ρ = 0.53 to 0.62 (P ≤ .05). The physical domain of the pVHI did not correlate with any CAPE-V domains in females (or in males for that matter), but functional and emotional pVHI domain correlations in females were sufficient to account for significant total pVHI correlations with the breathiness and strain domains of the CAPE-V. The CAPE-V overall trended toward but did not reach significance in correlation with the total pVHI in females (ρ = 0.49, P = .08). In contrast, male patients’ CAPE-V overall correlation with total pVHI was ρ = −0.04 (P = .85). Clinician assessments of the female voice, especially in the domains of breathiness and strain, seem to much more closely mirror the expectations of patient proxy reports than in the male voice of patients with VFLs.

Further study is indicated to better understand the limitations of these instruments and how they may be affected by patient/proxy (or societal) expectations of voice. It is clear that gender has an important role in this discussion and must be carefully evaluated when considering vocal assessment data.
LIMITATIONS OF THIS STUDY

This study has the limitations common to all retrospective reviews, with incomplete and imperfect data sources. These data benefited from inclusion in a database that standardized many of the protocols for data collection and deidentification. There were still, however, missing data due to either omission or patient noncompliance. Limitations of the scales themselves have been discussed in previous works.1-4

The present findings should also be interpreted understanding the potential for increased type I error, given the number of multiple correlation tests conducted. The goal of these analyses was a transparent and thorough evaluation of the data to consider areas that may warrant further study, not to produce data in support of pervasive conclusions, given the low correlations overall, multiple tests, and preliminary nature of this review. All the correlations are reported to allow independent interpretation of the available data.

In conclusion, the CAPE-V and the pVHI are useful tools in the measurement of voice outcomes in children with VFLs. This study showed no significant differences in severity between CAPE-V and pVHI scores and only fair overall correlations between these assessments. However, when gender is accounted for in the analysis, the correlations strengthened in males for the loudness domain and in females for the breathiness and strain domains. The strengthening of these correlations regarding gender may reflect the impact of societal expectations for voice on perception by parental proxy and expert clinicians and warrants further study. The CAPE-V and pVHI likely evaluate important yet different aspects of voice disturbance, and the noted gender differences in scale correlations should be further evaluated in future studies.

Submitted for Publication: May 3, 2011; final revision received August 13, 2011; accepted September 20, 2011. 
Correspondence: Alessandro de Alarcon, MD, MPH, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, MLC 2018, Cincinnati, OH 45229-3039 (alessandro.dealarcon@cchmc.org).

Author Contributions: Drs Johnson and de Alarcon had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Johnson, Brehm, Weinrich, Meinzen-Derr, and de Alarcon. Acquisition of data: Johnson, Brehm, Weinrich, and de Alarcon. Analysis and interpretation of data: Johnson, Brehm, Weinrich, Meinzen-Derr, and de Alarcon. Drafting of the manuscript: Johnson, Meinzen-Derr, and de Alarcon. Critical revision of the manuscript for important intellectual content: Johnson, Brehm, Weinrich, Meinzen-Derr, and de Alarcon. Statistical analysis: Meinzen-Derr and de Alarcon. Administrative, technical, and material support: Brehm, Weinrich, and de Alarcon. Study supervision: de Alarcon.

Financial Disclosure: None reported.

Previous Presentation: This article was presented at the American Society of Pediatric Otolaryngology meeting; May 1, 2011; Chicago, Illinois.

REFERENCES