Validation of an Outcomes Instrument for Tonsil and Adenoid Disease

Michael G. Stewart, MD, MPH; Ellen M. Friedman, MD; Marcelle Sulek, MD; Andrew deJong, MD; Gregory F. Hulka, MD; Marilyn H. Bautista, MPH; Susan E. Anderson, RN

**Objective:** To design and validate a disease-specific health status instrument—the Tonsil and Adenoid Health Status Instrument—for use in children with tonsil and adenoid disease.

**Design:** Prospective psychometric and clinimetric instrument validation in 3 stages.

**Settings:** A tertiary academic pediatric specialty hospital and a tertiary academic hospital, in 2 different cities.

**Patients/Other Participants:** Children with tonsil and adenoid disease presenting for evaluation and treatment (n=224).

**Intervention/Method:** Prospective instrument validation. Stage 1 consisted of initial item testing, reduction, and subscale construction; stage 2, reliability and validity testing, factor analysis, and final item reduction; and stage 3, responsiveness analysis.

**Main Outcome Measures:** Test-retest and internal consistency reliability; content, construct, and criterion validity; orthogonal principal components factor analysis; and response sensitivity analysis.

**Results:** Factor analysis and item analysis confirmed 6 distinct subscales measuring different constructs (aspects) of disease-specific health status that are affected by tonsil and adenoid disease: eating and swallowing, airway and breathing, infections, health care utilization, cost of care, and behavior. For each subscale, the Tonsil and Adenoid Health Status Instrument demonstrated excellent test-retest reliability ($r=0.72-0.88$) and internal consistency reliability (Cronbach $\alpha=.73-.87$). Content validity was ensured during the design process. Construct validity was demonstrated by means of convergent and divergent validity with a global quality-of-life instrument (the Child Health Questionnaire, version PF28). Criterion validity was also satisfactory. Finally, the instrument was appropriately sensitive, with high standardized response means and effect sizes.

**Conclusions:** The Tonsil and Adenoid Health Status Instrument is a valid, reliable, and sensitive instrument with 6 distinct subscales. This instrument has significant utility for outcomes research in children with tonsil and adenoid disease.

MATERIALS AND METHODS

This was a multicenter prospective instrument validation study in 3 phases. This study was approved by the Baylor College of Medicine Institutional Review Board for Human Subjects Research (Houston, Tex). All analyses were performed with SPSS 7.0 statistical software (SPSS Inc, Chicago, Ill).

Phases 1, 2, and 3 have been partially described in another publication.29

PHASE 1

An expert group identified 31 individual concepts related to health status in T&A disease and constructed questions (items) for each individual concept, along with summary items concerning the overall impact of each dimension on health status. Items were all constructed with the use of a 5-part Likert scale (not a problem, very mild problem, moderate problem, fairly bad problem, and severe problem), and were phrased, “How much of a problem is ....” The alpha-version of the instrument was structured into a telephone interview format, with the use of the principles of Sudman and Bradburn.30 The instrument asked the parent to recall the previous 6 months.

Inclusion criteria were age of 2 to 16 years and diagnosis of any combination of the following: recurrent tonsillitis, recurrent pharyngitis, chronic tonsillitis, at least 2 episodes of peritonsillar abscess, tonsil hypertrophy, airway obstruction, obstructive sleep pattern, obstructive sleep apnea, and hypopnea. Exclusion criteria were diagnosis of possible malignant neoplasm of the tonsil or adenoid; emergency surgery (eg, for peritonsillar abscess); adenoidectomy alone, performed for treatment of otologic disease; marked immunodeficiency (such as human immunodeficiency virus infection, severe combined immunodeficiency disorder, iatrogenic immunodeficiency from treatment of a malignant neoplasm, etc); complete cleft of the secondary palate; and non–English-speaking primary caretaker. 

Recurrent tonsillitis or pharyngitis was defined as 3 or more episodes of infection in 12 months. Chronic tonsillitis was defined as persistent symptoms of tonsillitis (ie, odynophagia, sore throat, dysphagia, fever, and cervical adenopathy) for at least 3 months. Obstructive sleep pattern and obstructive sleep apnea or hypopnea were defined by either a characteristic history of obstructive sleep pattern—such as witnessed apnea, or witnessed loud snoring combined with respiratory disturbance—or a tape recording of the child sleeping that demonstrates the sleep disturbance or a polysonogram. Tonsil hypertrophy was defined by the attending otolaryngologist examination as bilateral tonsil size of at least 3 of 4 points on a widely used standardized 4-point scale.3

Telephone interviews were performed by 2 of us (M.G.S. and M.H.B.) with extensive experience in interview techniques, and 2 otolaryngology resident physicians who were trained and validated in telephone interview techniques. Items that had to be repeated or caused the parents to ask for further explanation were noted by the interviewer.

Item responses from the alpha-version instrument were entered into a spreadsheet in SPSS. Initial item reduction was performed by sequential statistical analysis, including individual item analysis, internal consistency reliability, construct validity, item-item and item-subgroup correlations, and factor analysis. All statistical analyses were performed with SPSS version 7.0 statistical software.31

Next, a large table was constructed with individual items in rows, and several columns containing the results of individual item analysis, internal consistency reliability, item-item correlation, item-subgroup importance correlation, and item–summary scale correlation, all graded on a subjective 0 to 3 scale. The following scale was used to assess the item’s “performance” on the test of interest: 0 indicates poor performance; 1, marginal performance; 2, good performance; and 3, excellent performance. Another column contained the number of times a subject had difficulty answering the item. Items with adequate score distribution, high internal consistency reliability, high item-item and item-subgroup importance correlation, and low amount of respondent difficulty were selected for inclusion in the beta-version of the instrument; overall, 18 items were selected. Of note, no items had significantly different results for the individual analyses. Items with poor internal consistency typically also had poor subgroup importance correlation, poor score distribution, etc. The items on the beta-version instrument were rewritten into a format for self-completion (rather than interview) by the method of Aday.30

With the use of the 18 remaining items after initial item reduction, confirmatory factor analysis was performed to assess the grouping of items into subgroups. Principal components factor extraction was performed by means of orthogonal varimax rotation of factors.31 All factors with an eigenvalue greater than 1.0 were included in the final

QOL in children,22-26 but there are no instruments that measure disease-specific health status in children with tonsil and adenoid (T&A) disease. Because T&A disease affects children with different communication skills, the parent who is the primary caretaker is surveyed as a proxy for the affected child. In general, using a proxy for health status assessment is discouraged in adult patients,27 but it is a common and accepted practice in the pediatric population.20,24,28

In this report, we describe in detail the validation of a disease-specific health status instrument for use in children with T&A disease and, in addition, describe potential clinical and research uses for the instrument.

RESULTS

In phase 1, a total of 34 interviews were conducted; the mean age of affected children was 6.8 years (median, 5.5 years; range, 2–15 years). There were 18 boys and 16 girls, and the ethnic distribution was as follows: white, 17 (50%); African American, 6 (18%); Hispanic, 3 (9%); Asian American, 2 (6%); and no data available, 6 (18%).

Individual item analysis was performed to assess the mean, median, range, variance, and distribution skewness of responses for all 37 items. For optimum discriminatory power, item variance should be relatively high,
PHASE 2

Between July 1, 1997, and June 30, 1998, at the Baylor College of Medicine and Duke University School of Medicine (Durham, NC) sites, consecutive parents of eligible children were given the beta-version instrument and a validated global QOL instrument, the Child Health Questionnaire version PF28 (CHQ-PF28).25,29 The first analysis completed in phase 2 was factor analysis, to allow the construction of independent subscales. Factor analysis was performed with orthogonal varimax rotation of factors32,33. Several solutions were calculated. All factors with an eigenvalue greater than 1.0 were analyzed to minimize unexplained variance; solutions using 5, 4, 3, and 2 total factors were calculated to assess the degree of unexplained variance in each model. On the basis of the data from phase 1, we anticipated that a solution with approximately 5 factors would explain an adequate amount of variance.

Using these results, we then constructed subscales. Items were scored from 0 to 4, and items were summed to obtain the subscale raw score. Subscale raw scores were scaled to a minimum of 0 and a maximum of 100, by means of the following formula: scaled score = [(raw score−min score)/(max score−min score)] × 100, where max score indicates the maximum possible subscale score, and min score, the minimum possible subscale score.

Test-retest reliability was assessed in a subgroup of patients scheduled for adenotonsillectomy by repeating the administration of the beta-version of the instrument at the time of surgery, 2 to 6 weeks after the initial completion. Patients with surgery scheduled for earlier than 2 weeks after the initial visit were not used in the assessment of test-retest reliability. Test-retest reliability for subscales was assessed by means of the Goodman-Kruskal γ coefficient between test administrations.34

Internal consistency reliability for subscales containing at least 2 items was assessed by calculating the Cronbach α coefficient30,34 and noting item-total correlations. Items that did not contribute to the internal consistency of a subscale were noted for potential deletion.

Construct validity was assessed by means of (1) a multitrait, multi-item correlation matrix with the CHQ-PF28;30,34 (2) item-subscale and subscale-subscale correlations; and (3) between-group discrimination. Constructs assessed on the global CHQ instrument included global health; physical functioning; role and social limitations—emotional or behavioral; role and social limitations—physical; bodily pain or discomfort; behavior; mental health; self-esteem; parental impact—emotional; and parental impact—time.35 Because of the nonparametric nature of the data, Spearman correlation coefficients were used throughout. A preliminary “expected correlation” matrix was created a priori by means of subscales from the T&A instrument and the CHQ-PF28 for the purpose of analyzing the results. For instance, health care utilization would be expected to correlate with parental impact—time and not with mental health. Furthermore, the “behavior” item on the instrument should correlate with the behavior subscale on the CHQ-PF28, and so on. Between-group discrimination was assessed by comparing subscale scores between children who had documented sleep-disordered breathing (SDB) and children with other indications for surgery. Children with known SDB should have significantly worse airway and breathing subscale scores than children without SDB. Furthermore, children with SDB should not necessarily have significantly different health care utilization subscale scores than children without SDB. This was examined by comparing subscale scores between groups with the Mann-Whitney test.

Criterion validity was also assessed.39 Correlation coefficients were compared between subscale scores and the objective indicators of (1) infections (eg, antibiotics prescribed), (2) breathing and airway (eg, tonsil size and results of polysomnogram), and (3) health care utilization (eg, number of telephone calls and number of physician visits) gathered from chart abstraction. There were no reliable criterion measures to use for validation of the behavior or swallowing subscales.

PHASE 3

Between August 1, 1998, and December 31, 1999, at the Baylor site, a consecutive sample of children undergoing adenotonsillectomy was studied prospectively to assess the response sensitivity of the instrument. Entry criteria were the same as for phases 1 and 2 of the study. Parents were given the T&A Health Status Instrument before the child’s surgery, then again at least 6 months after the child’s surgery, since the instrument measures the preceding 6-month period. Response sensitivity after surgical treatment was assessed by calculating the standardized response mean and the effect size34,36 and by comparing these values with published standards.

Several variables were assessed to explore relationships between individual items and the subscale: the α coefficient with the item deleted, the corrected item-total correlation, and the squared multiple correlation.35,34

The initial α coefficients for the 4 subscales were between 0.67 and 0.81. After items that tested poorly were eliminated, the α coefficients for the subscales in the alpha-version instrument were as follows: infection, α = .74; health care utilization, α = .83; airway and breathing, α = .80; and swallowing and eating, α = .72.

To assess construct validity, items from each subgroup were correlated with (1) other items from the same subgroup, (2) the overall “importance” item for that sub-
group, and (3) a summary item for health impact. Item-item, item-subgroup, and item-summary correlations were evaluated in correlation matrices by means of Spearman item, item-subgroup, and item-summary correlations were accounted for 75.3%

Table 2. Interitem Correlations for the Airway and Breathing Items From Phase I Version of the Instrument*

<table>
<thead>
<tr>
<th>Items</th>
<th>Snoring</th>
<th>Apnea</th>
<th>Mouth Breathing</th>
<th>Noisy Breathing</th>
<th>Falling Asleep</th>
<th>Speech Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loud snoring</td>
<td>0.45</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
<td>0.25</td>
<td>0.02</td>
</tr>
<tr>
<td>Apnea</td>
<td>0.45</td>
<td>0.54</td>
<td>0.37</td>
<td>0.17</td>
<td>0.09</td>
<td>0.14</td>
</tr>
<tr>
<td>Mouth breathing</td>
<td>0.44</td>
<td>0.54</td>
<td>0.71</td>
<td>0.09</td>
<td>0.09</td>
<td>0.14</td>
</tr>
<tr>
<td>Noisy breathing</td>
<td>0.44</td>
<td>0.37</td>
<td>0.71</td>
<td>0.09</td>
<td>0.09</td>
<td>0.14</td>
</tr>
<tr>
<td>Falling asleep</td>
<td>0.24</td>
<td>0.17</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.03</td>
</tr>
<tr>
<td>Speech problems</td>
<td>0.02</td>
<td>0.05</td>
<td>0.14</td>
<td>−0.05</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

*Spearman correlation coefficients of 0.40 or more are in boldface type.

Table 3. Principal Component Factors Identified From Confirmatory Factor Analysis of Phase I Version of the Instrument and Items That Primarily Load Onto Each Factor

<table>
<thead>
<tr>
<th>Factor No.</th>
<th>Item Content</th>
<th>Eigenvalue</th>
<th>% Variance</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chronic illness</td>
<td>5.87</td>
<td>25.5</td>
<td>9, 10, 11, 26, 33, 34</td>
</tr>
<tr>
<td>2</td>
<td>Swallowing</td>
<td>3.62</td>
<td>15.8</td>
<td>29, 31</td>
</tr>
<tr>
<td>3</td>
<td>Repeated care</td>
<td>2.82</td>
<td>12.3</td>
<td>14, 19, 20, 24</td>
</tr>
<tr>
<td>4</td>
<td>Infection</td>
<td>1.78</td>
<td>7.7</td>
<td>9, 19, 21, 23</td>
</tr>
<tr>
<td>5</td>
<td>Breathing</td>
<td>1.70</td>
<td>7.4</td>
<td>2, 3, 4</td>
</tr>
<tr>
<td>6</td>
<td>Loud snoring</td>
<td>1.52</td>
<td>6.6</td>
<td>1</td>
</tr>
</tbody>
</table>

*Spearman correlation coefficients of 0.40 or more are in boldface type.

Table 2. Item Importance and Item Summary Scale Correlation Coefficients for Phase I Version of the Instrument*

<table>
<thead>
<tr>
<th>Item No. and Subject</th>
<th>Subgroup Importance</th>
<th>Summary Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Loud snoring</td>
<td>0.32</td>
<td>0.15</td>
</tr>
<tr>
<td>2. Apnea</td>
<td>0.46</td>
<td>0.01</td>
</tr>
<tr>
<td>3. Mouth breathing</td>
<td>0.35</td>
<td>−0.24</td>
</tr>
<tr>
<td>4. Noisy breathing</td>
<td>0.49</td>
<td>−0.14</td>
</tr>
<tr>
<td>5. Falling asleep</td>
<td>0.01</td>
<td>−0.06</td>
</tr>
<tr>
<td>6. Speech problems</td>
<td>0.12</td>
<td>−0.08</td>
</tr>
<tr>
<td>9. Acute tonsil infection</td>
<td>0.56</td>
<td>0.48</td>
</tr>
<tr>
<td>10. Streptococcal sore throat</td>
<td>0.49</td>
<td>0.36</td>
</tr>
<tr>
<td>11. Chronic tonsil infection</td>
<td>0.49</td>
<td>0.45</td>
</tr>
<tr>
<td>12. Mouth odor</td>
<td>0.18</td>
<td>0.25</td>
</tr>
<tr>
<td>14. Acute ear infection</td>
<td>0.39</td>
<td>0.13</td>
</tr>
<tr>
<td>15. Chronic ear infection</td>
<td>0.13</td>
<td>−0.33</td>
</tr>
<tr>
<td>17. Acute sinus infection</td>
<td>−0.06</td>
<td>0.13</td>
</tr>
<tr>
<td>19. Physician trips</td>
<td>0.59</td>
<td>0.40</td>
</tr>
<tr>
<td>20. Physician calls</td>
<td>0.66</td>
<td>0.42</td>
</tr>
<tr>
<td>21. Antibiotics for 3 wk</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td>22. Repeated antibiotics</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>23. Cost of medications</td>
<td>0.33</td>
<td>0.24</td>
</tr>
<tr>
<td>24. Cost of physician</td>
<td>0.42</td>
<td>−0.02</td>
</tr>
<tr>
<td>25. Side effects of medications</td>
<td>0.26</td>
<td>0.12</td>
</tr>
<tr>
<td>26. Missed school</td>
<td>0.60</td>
<td>0.38</td>
</tr>
<tr>
<td>28. Swallowing</td>
<td>0.34</td>
<td>0.14</td>
</tr>
<tr>
<td>29. Poor appetite</td>
<td>0.85</td>
<td>0.35</td>
</tr>
<tr>
<td>30. Losing weight</td>
<td>0.41</td>
<td>−0.08</td>
</tr>
<tr>
<td>31. Not growing</td>
<td>0.45</td>
<td>−0.01</td>
</tr>
<tr>
<td>33. Behavior problems</td>
<td>0.52</td>
<td>0.11</td>
</tr>
<tr>
<td>34. School problems</td>
<td>0.48</td>
<td>0.13</td>
</tr>
</tbody>
</table>

As previously discussed, a shorter version of the instrument (the beta-version) was then created and tested in a separate patient population (n=158) for phase 2 of the study. There were 59% boys and 41% girls, and mean age was 5.8 years (median, 5.0 years; range, 2–15 years).

Repeated factor analysis of data from patients in phase 2 demonstrated that a 5-factor solution yielded the highest explained variance (70% vs 64% for 4 factors and 56% for 3 factors). Factor analysis with total variance, along with examination of the scree plot, demonstrated that the eigenvalue of the sixth factor was 0.84, confirming that a 5-factor solution was ideal. Individual items were grouped into subscales according to their principal factors for the remainder of the analysis. The behavior item did not load onto any of the 5 principal factors but was believed to be important, so it was included as a separate item and subscale. On the basis of factor loadings and content, the 6 subscales were named airway and breathing (4 items), infections (4 items), eating and swallowing (2 items), health care utilization (5 items), cost of care (2 items), and behavior (1 item).

Initial assessment after phase 2 indicated that internal consistency reliability was also high for all subscales. Individual item analysis disclosed that the 2 items related to cost could be combined into a single item (the item–item correlation coefficient was 0.95). In addition, an item concerning repeated acute infections had poor internal consistency and low item-total correlation with other subscale items, and therefore it was deleted. Similarly, an item related to missed days of school had poor
internal consistency and low item-total correlation and was also deleted. The internal reliability coefficients increased for both subscales after those items were deleted. Therefore, after elimination of those 3 items, the subscales contained the following numbers of items: airway and breathing, 4 items; infections, 3 items; eating and swallowing, 2 items; health care utilization, 4 items; cost of care, 1 item; and behavior, 1 item.

Cronbach α coefficients of at least .70 are considered adequate for group comparisons,34 and the coefficients for subscales with at least 2 items were all adequate.29 Test-retest reliability was also very strong for all subscales. Reliability coefficients greater than 0.70 are considered acceptable,34 and the subscale coefficients were as follows: airway and breathing, γ = 0.80; infections, γ = 0.74; eating and swallowing, γ = 0.84; health care utilization, γ = 0.78; cost of care, γ = 0.72; and behavior, γ = 0.88. In all subscales with at least 2 items, the Spearman correlation coefficients between repeated administrations were even higher than the reliability coefficients (p = 0.74-0.89).

Validity was not measured by a single test, but rather was inferred from a compilation of evidence of different types of validity. Content validity was ensured during the design phase of the instrument.29 In the assessment of construct validity, a multitrait, multi-item correlation matrix was created, and several expected associations were identified. All correlation coefficients were negative, since the T&A instrument and CHQ are scored in opposite directions (higher scores indicate poorer disease-specific health status but better global QOL). For instance, the utilization subscale correlated significantly with the parental impact–time subscale on the CHQ (p = -0.42; P = .001) but not with unrelated subscales such as mental health or self-esteem. The infection subscale correlated significantly with the bodily pain (P = .003) subscales. The airway and breathing subscale correlated with the physical functioning (p = -0.52; P < .001) and global health (p = -0.47; P < .001) subscales on the CHQ. The behavior item strongly correlated with the behavior subscale on the CHQ (p = -0.61; P < .001), and also with the self-esteem (p = -0.53; P < .001), mental health (p = -0.50; P < .001), and parental impact–emotional (p = -0.50; P = .002) subscales. The eating and swallowing subscale measured a construct not explored on the CHQ, and, as expected, no significant correlations were identified.

Construct validity of the T&A instrument was further demonstrated by the strong convergent and divergent validity shown in the item-subscale correlations: highly significant correlations were noted between items and their related subscale and nonsignificant correlations between items and nonrelated subscales. For instance, none of the items on the infections subscale correlated with items on the eating and swallowing subscale. Of course, there were some associations across subscales (for instance, infections were associated with increased health care utilization and increased cost of care, and cost of care was independently associated with health care utilization, etc).

As a final test of construct validity, the instrument demonstrated a strong ability for between-group discrimination. Airway and breathing subscale scores were significantly higher in a group of children with documented SDB than in a group without SDB (mean scores, 66.0 and 32.3, respectively; P < .001); other subscale scores did not differ between children with and without SDB (behavior subscale, P = .13; cost subscale, P = .38; utilization subscale, P = .30).

Criterion validity was demonstrated by assessing the correlations between appropriate subscale scores and objective clinical data available from 74 children. The number of documented infections in the previous 6 months correlated significantly with the infection (Spearman r = 0.55; P < .001) and utilization (r = 0.34; P = .004) subscales. Similarly, the correlations between the number of actual physician visits and the utilization (r = 0.32; P = .007) and cost (r = 0.26; P = .03) subscales were weaker but still statistically significant, even with a relatively small sample size. The sample of patients with objective polysomnogram data was inadequate for analysis of criterion validity of the airway and breathing subscale.

Finally, in phase 3 of the study, 62 children were enrolled and 32 completed the 6-month repeated version of the instrument. The instrument demonstrated very high levels of response sensitivity for 5 of the 6 subscales, as measured by both the standardized response mean and effect size. The standardized response means for each subscale were as follows: airway, 1.42; infections, 1.16; utilization, 1.38; eating and swallowing, 0.90; cost, 0.65; and behavior, 0.10. The calculated effect sizes were very similar, ranging from 1.49 (airway) to 0.14 (behavior). As a rule, standardized response mean and effect size values of approximately 0.2 represent low sensitivity to change, approximately 0.5 indicates moderate sensitivity, and around 0.8 indicates high sensitivity.35,36 So the instrument clearly shows high sensitivity to clinical change in all subscales except the behavior subscale.

A final version of the instrument is shown in the Figure. Items and their associated subscales are as follows: airway and breathing subscale, items 1, 7, 11, and 13; infection subscale, items 2, 8, and 9; health care utilization subscale, items 3, 4, 5, and 6; eating and swallowing subscale, items 12 and 14; cost of care subscale, item 10; and behavior subscale, item 15. As discussed previously, each subscale should be scored so that scores range from 0 (minimum score) to 100 (maximum score).

**COMMENT**

For many diseases, disease-specific health status instruments are necessary to assess changes in health status that are clinically important, but perhaps too subtle to be detected by means of a global QOL instrument.27 This has been demonstrated in patients with ocular cataracts,37 chronic sinusitis,35 and many other diseases. We have now completed validation of the T&A Health Status Instrument. The instrument was also designed to be comprehensive so that children with any T&A-related problems could be studied with the same instrument. Containing only 15 items, the instrument is easy to complete in a few minutes, and this low respondent burden makes the instrument ideal for multiple administra-
The T&A Health Status Instrument was validated for use in groups of children, not individual patients. Therefore, the instrument is very reliable and sensitive for comparing outcomes in groups (for instance, those treated with and without surgery), but not for predicting treatment outcome in an individual patient. This is an important point to remember when choosing a health status or QOL instrument for clinical or research use, because most health status instruments were validated for use in group comparisons.

The instrument described could be used to measure the patient- and family-based severity of "mild to moderate" T&A disease—for instance, those who do not seem to meet accepted surgical criteria for adenotonsillectomy but still seem to be affected. It might be that the disease-specific health status of these children is no better than that of children who meet an accepted surgical indication, such as a minimum number of infections per year.

The instrument could also be used to describe the natural course of T&A disease over time. Many clinicians believe that children outgrow their problems with T&A disease, and that watchful waiting is a good option in patients who do not have severe disease. However, there are few objective data to support this assertion. By periodically measuring the disease-specific health status of children who are treated with watchful waiting, we could better define the natural course of the disease. Of course, knowing the natural course of untreated disease would enable us to better assess the true effects of treatment.

Another important use of the instrument is for prospective measurement of the health status impact of medical or surgical treatment. Groups of patients with T&A disease could complete the instrument at presentation and then again after treatment. The instrument measures a 6-month period, so it should be used at least 6 months after treatment is completed. If the treatment is effective, then subscale scores should improve significantly; in fact, treatment efficacy could be compared between treatments by assessing relative improvement in patient-based health status. For instance, at 1 year, treatment of recurrent tonsillitis with antibiotics and watchful waiting could show health status improvement equivalent to that with tonsillectomy. Similarly, one could compare health status improvement after treatment of SDB with either continuous positive airway pressure or adenotonsillectomy. These prospective studies will be important in measuring the efficacy of adenotonsillectomy as treatment in affected children.

If a global QOL instrument is used in addition to the T&A instrument, then the changes in overall QOL after treatment could also be compared. However, in general, global instruments are much less sensitive to treatment effects than are disease-specific instruments. One benefit of using global instruments is that global QOL can be compared (ie, benchmarked) against global QOL in other diseases. While these comparisons between different diagnoses can be flawed (for instance, by the presence of other comorbid disease), these data can provide useful insight into the relative global burden of a particular disease.

The instrument is scored into subscales, as described previously. These subscale scores can quantify the degree of impact of different aspects of T&A disease in any given population or sample. For instance, one group of patients may be primarily affected with airway or breathing problems, whereas another group of patients may have infectious problems. The subscales are each scaled so that scores range from 0 (no impact) to 100 (maximum impact). Therefore, effective treatment should result in lower subscale scores; similarly, more effective treatments should result in larger numerical improvement in scores. Although individual subscale scores could in theory be added to obtain a "total" T&A score, that is not recommended. For a total score to be valid, the relative contribution of each subscale would have to be equivalent; for example, if there were 3 subscales, then each must make up 33.3% of the total score value—and those relative impacts are not known. It is preferable that individual subscale scores be used for analysis and interpretation.

The results obtained from prospective studies of children with T&A disease, such as those described herein, should help primary care and specialty physicians re-
evaluate current treatment indications and protocols for this prevalent problem, and should help ensure improved health status and QOL for affected children in the future.

Accepted for publication July 13, 2000.

This study was supported by grant R03-HS09829 from the Agency for Health Care Policy and Research, Rockville, Md (Dr Stewart).

Presented in part at the American Society of Pediatric Otolaryngology meeting, Palm Desert, Calif, April 29, 1999.

Corresponding author and reprints: Michael G. Stewart, MD, MPH, Baylor College of Medicine, One Baylor Plaza (NA-102), Houston, TX 77030 (e-mail: mgstew@bcm.tmc.edu).

REFERENCES