OBJECTIVE: To validate a disease-specific health-related quality of life (HRQOL) instrument for children with obstructive sleep disorders (OSDs).

DESIGN: Prospective cohort study using a 6-item health-related instrument (OSD-6).

SUBJECTS: One hundred caregivers of patients with OSDs secondary to adenotonsillar hypertrophy (age range, 2-12 years) from 2 tertiary care, pediatric otolaryngology practices.

INTERVENTION: The OSD-6 was administered on initial presentation and 4 to 5 weeks after adenotonsillectomy. A subset of patients repeated the OSD-6 within 3 weeks after presentation to assess test-retest reliability.

MAIN OUTCOME MEASURES: Test-retest reliability, internal consistency, construct validity, and responsiveness to clinical change of the OSD-6 score.

RESULTS: Test-retest reliability was good (intraclass correlation coefficient = 0.74). Median OSD-6 score was 4.5 (0- to 6-point scale) with higher scores indicating poorer quality of life (QOL). Construct validity was demonstrated by the moderate correlation between OSD-6 score and global adenoid and tonsil-related QOL ($R = -0.62$), strong correlation between the OSD-6 change score and change in global adenoid and tonsil-related QOL ($R = -0.63$), and the moderate correlation between the change score and parent estimate of clinical change ($R = 0.40$). The mean change in OSD-6 score after adenotonsillectomy was 3.0 (95% confidence interval, 2.7-3.4). The mean standardized response was 2.3 (95% confidence interval, 1.9-2.7) indicating the instrument’s large responsiveness to clinical change. The change score was very reliable ($R = 0.85$).

CONCLUSIONS: The OSD-6 is a reliable, responsive, easily administered instrument. It is valid for detecting change after adenotonsillectomy in children with OSDs.

Arch Otolaryngol Head Neck Surg. 2000;126:1423-1429

T

HE TERM “obstructive sleep disorders” (OSDs) refers to the spectrum of sleep-disordered breathing that is severe enough to cause clinical symptoms. This includes children with obstructive sleep apnea and children with upper-airway resistance syndrome in which the respiratory distress index is often normal on standard polysomnographic testing. Obstructive sleep apnea is estimated to occur in 1% to 3% of pre-school-aged children. The incidence of upper-airway resistance syndrome is unknown, but it is likely more prevalent than obstructive sleep apnea. The leading cause of OSDs in children is adenotonsillar hypertrophy.

Treatment studies of OSDs in children have mainly relied on objective measures of effectiveness. However, these measures may be less relevant to the general public. Parents want their child to feel and function better and third-party payers want objective evidence that surgical interventions improve the quality of life (QOL) in general and health-related quality of life (HRQOL) more specifically. In the context of OSDs, HRQOL describes the net consequences of sleep-disordered breathing on the child’s daily activities, physical symptoms, social interactions, and emotional well-being.

Adenotonsillectomy is the most common treatment for OSDs in children and it is successful in alleviating the problem in most cases. However, no systematic outcome data exist concerning the effect of this procedure on a child’s QOL. In the absence of meaningful effectiveness data, the improvement in QOL for these children after adenotonsillectomy is unappreciated and skepticism regarding the benefits of this procedure continues. The pur-
SUBJECTS, MATERIALS, AND METHODS

The study protocol received approval of the Children’s Hospital and Regional Medical Center Institutional Review Board, Seattle, Wash, and was exempted from review by the Children’s Hospital of the King’s Daughter Institutional Review Board, Norfolk, Va.

SURVEY DEVELOPMENT

The Obstructive Sleep Disorders-6 survey (OSD-6) was modeled in a format analogous to the a 6-item health-related instrument for otitis media-6, a validated survey for assessing HRQOL in acute and chronic otitis media. The OSD-6 was composed of 6 domains, which reflected functioning of the child regarding the following: (1) physical suffering, (2) sleep disturbance, (3) speech and swallowing difficulties, (4) emotional distress, (5) activity limitations, and (6) level of concern of the caregiver relating to the patient’s OSD and associated symptoms. Each domain was represented by a question designed to reflect the global effect of an OSD-related symptom cluster for an individual child. The symptom clusters were created by assimilating information from literature review, opinions of pediatric otolaryngologists and pediatric otolaryngology nurse practitioners, and consultations with parents of affected children. In addition, the questionnaire included an open-ended response question to survey the subjects for pertinent items that might have been missed on our initial survey. Individual items on the survey were equally weighted. The mean survey score was calculated by summing the individual domain scores (domain score range, 0-6 with 0 indicating “no problem” and 6, “[it] couldn’t be worse”) and then dividing by 6, the total number of domains. A lower survey score indicated a better QOL.

SUBJECTS

Eligible subjects were a convenience sample of caregivers of patients between the ages of 2 and 12 years undergoing adenotonsillectomy for OSDs. Subjects were required to be English speaking to avoid semantic issues in interpretation of the domains and symptom clusters. Caregivers were excluded if their child had other adenotonsillar pathology, or if another procedure was to be performed on the same day. Demographic data (ie, age and sex) and tonsillar size were recorded for all patients, including patients whose caregiver refused enrollment or those who dropped out of the study. Children were diagnosed as having OSDs based on a combination of any of the following modalities as determined by the standard clinical practice of the attending physician: medical history, physical examination findings, x-ray film of the lateral aspect of the neck, nasopharyngoscopy, sleep videotape, and/or polysonomographic study. The diagnosis was determined by the attending physicians from pediatric otolaryngology clinics at Children’s Hospital and Regional Medical Center, Seattle, Wash, and Children’s Hospital of the King’s Daughter, Eastern Virginia Medical School, Norfolk.

DATA COLLECTION

The OSD-6 was given to the patient’s caregiver at the time of diagnosis. Caregivers rated the domains on a 0- to 6-point scale (no problem to [it] couldn’t be worse) based on how they felt the symptoms affected their child (Figure 1). In a subset of patients, a second survey was mailed to the caregiver 5 days after the first visit to measure test-retest reliability. If the mailed survey was not returned prior to surgery, a second survey was completed by the caregiver on the day of surgery if this date was within 3 weeks of the initial administration of the survey. The first and second survey had to be completed by the same person; otherwise these measurements were excluded from test-retest calculations.

In patients who underwent adenotonsillectomy and returned for the postoperative visit, the survey was re-administered 4 to 5 weeks after surgery, so that the potential for sampling any symptoms of the recovery period would be bypassed. The caregiver was additionally asked to rate the degree of clinical change experienced by their child on a −7 to 7 rating scale (with −7 indicating the child was much worse; 7, the child was much better). At all 3 time points, of this study was to design an instrument that would reflect important areas of function for children with OSDs and their caregivers, which could be completed by the child’s caregiver during a routine office visit. The instrument should be valid, ie, actually measure QOL in OSDs, be reproducible in stable subjects, and be responsive to important changes in patient status, even if the degree of change was small. By developing a meaningful instrument, we hoped to facilitate a larger outcome study of the effect of adenotonsillectomy on QOL for children with OSDs.

RESULTS

One hundred caregivers of children with OSDs meeting the eligibility criteria were enrolled in the study. Sampling bias was assessed to ensure the enrolled population was representative of those eligible. Demographic data from subjects who refused to participate and study enrollees who dropped (n = 11) were compared with the enrolled population and were found to be comparable. Demographic data, physical examination findings, and mean survey and the 6 domain scores were also compared by site (Seattle vs Norfolk) and were not significantly different. The median patient age was 6.2 years (age range, 2.1-12.9 years). 49% of the patients were female and 51% were male. In 48 (49%) of 97 patients diagnosis was based on medical history and physical examination findings alone; an additional diagnostic test was used in the remainder of these patients. Nasopharyngoscopy was performed in 3 patients (3.1%); x-ray films of the lateral aspect of the neck were obtained for 13 patients (13.4%). Sleep videotapes were obtained for 30 patients (30.9%); data from these videotapes were not objectively rated. Sleep studies were obtained for 6 patients (6.2%).

All caregivers were able to easily self-administer the questionnaire in several minutes after a brief explanation of its contents. No new symptoms were listed in the
open-ended response question administered with the initial survey. Forty-five caregivers completed a second survey within 3 weeks of the initial survey for test-retest calculations. Sixty-two caregivers completed a postintervention survey after adenotonsillectomy to assess responsiveness to clinical change. Mean preoperative domain and survey scores, as well as physical examination findings were compared between patients who attended the postoperative examination and those who did not. There were no differences in demographics, physical examination findings, or mean survey scores. The only difference in the 6 domain scores was the level of sleep disturbance. Patients who were noncompliant with the follow-up visit rated sleep disturbance as less severe (4.9 vs 5.9, P < .05) preoperatively than those who attended the postoperative visit.

The distribution of baseline domain and survey scores is shown in Figure 2 and Figure 3, respectively. Physical suffering, sleep disturbance, and speech or swallowing problems were moderate or greater (score = 6) for 88.0%, 89.0%, or 64.0% of the patients, respectively. Fifty-four percent of the caregivers rated their level of concern over their child's sleep disturbance in the 2 most severe categories ("very much," and [it] couldn't be worse). Seventy-six percent of the caregivers rated their level of concern as a moderate or worse problem. Median baseline survey responses for the 100 caregivers are given in Table 1. Sleep disturbance and caregiver concern were the highest rated items (median score = 6), whereas activity limitations were the least affected (median score = 3). Test-retest reliability was good (>0.7) for all survey domains except activity limitations (Table 1). The instrument demonstrated significant internal consistency with a Cronbach α of 0.80.

Construct validity was shown by demonstrating at least modest correlations for the a priori predictions between related constructs (Table 2). Demonstration of cross-sectional construct validity was shown by good cor-
Clinicians and others involved in decisions about health care may place different values on various health states than those who live with a condition. As decisions concerning resource allocation become increasingly stringent, it is important to understand the personal impact of diseases and their treatments beyond the standard medical morbidity or functional limitations so that this can be incorporated into the decision-making process. In the context of this study, it will be useful to use the OSD-6 to measure QOL changes in children who undergo adenotonsillectomy. We can then share this information with parents who are considering the procedure for their child as well as to demonstrate to third-party payers that this is a beneficial procedure for children with OSDs.

Measuring HRQOL involves the use of a self-administered survey or instrument, which contains questions that reflect important domains of function. The instrument must be valid, reliable, and responsive. It should make sense and adequately cover areas of relevance (face and content validity), measure what it is intended to measure (construct validity), yield relatively stable results when readministered to patients with stable disease (test-retest reliability), and quantify clinically meaningful QOL changes over time within a patient (responsiveness). A disease-specific instrument focuses on areas of function that pertain to a particular disease or condition. It is used to describe the effect of disease on individuals above and beyond the usual biomedical factors used to assess severity of the particular disease or condition.

Validation strategies of HRQOL instruments have been borrowed from clinical and experimental psychologists. The types of validity psychologists have emphasized include content, construct, and criterion validity. Content validity reflects the extent to which an instrument measures a particular construct or trait and is often determined by expert panels.

Construct validity reflects the extent to which the instrument measures what it is intended to measure and is determined by factor analysis or other statistical techniques. Criterion validity reflects the extent to which the instrument correlates with other measures of the same construct.

The responsiveness of the instrument to clinical change was assessed using data from all surgical patients in whom follow-up was available (n=62) and 12 patients who completed 2 surveys at longer than 3-week intervals, but did not undergo the procedure in the study period (Table 2). Fifty-five patients (74.3%) demonstrated a large degree of clinical change (change score ≥1.5), 1 patient had a moderate change (change score, 1.0-1.4), 6 patients (10.8%) had a small change (change score, 0.5-0.9), and 12 patients (16.2%) had a trivial change (change score, <0.5). Children who had a trivial change in status on the follow-up visit had a mean change score of 0.24 (95% CI, 0.13-0.32), whereas children who had a large change in status had a mean change score of 3.3 (95% CI, 3.1-3.6). The patients who underwent surgery had a mean change score of 3.0 (95% CI, 2.7-3.4); the nonsurgical patients had a mean change score of 0.35 (95% CI, 0.22-0.48). Excluding the nonsurgical patients, 55 (88.7%) of the 62 patients demonstrated a large degree of clinical change after surgery. The reliability of the change score for the OSD-6 was 0.86. The SRM for individual domain scores and the mean survey score demonstrated excellent responsiveness to clinical change (Table 3). An SRM greater than 0.8 represent a large degree of responsiveness to change and each of our 6 domain scores, as well as the SRM for the survey as a whole were well above this threshold.

Clinicians and others involved in decisions about health care may place different values on various health states than those who live with a condition. As decisions concerning resource allocation become increasingly stringent, it is important to understand the personal impact of diseases and their treatments beyond the standard medical morbidity or functional limitations so that this can be incorporated into the decision-making process. In the context of this study, it will be useful to use the OSD-6 to measure QOL changes in children who undergo adenotonsillectomy. We can then share this information with parents who are considering the procedure for their child as well as to demonstrate to third-party payers that this is a beneficial procedure for children with OSDs.

Measuring HRQOL involves the use of a self-administered survey or instrument, which contains questions that reflect important domains of function. The instrument must be valid, reliable, and responsive. It should make sense and adequately cover areas of relevance (face and content validity), measure what it is intended to measure (construct validity), yield relatively stable results when readministered to patients with stable disease (test-retest reliability), and quantify clinically meaningful QOL changes over time within a patient (responsiveness). A disease-specific instrument focuses on areas of function that pertain to a particular disease or condition. It is used to describe the effect of disease on individuals above and beyond the usual biomedical factors used to assess severity of the particular disease or condition.

Validation strategies of HRQOL instruments have been borrowed from clinical and experimental psychologists. The types of validity psychologists have empha-
sized are content and construct validity. Face and content validity evaluations tend to be rather subjective, but should ideally include systematic comparisons with other available instruments, widely accepted definitions, professional opinions, and information extracted from those affected by the condition.12 In evaluating construct validity, it is important to have a priori predictions about the direction and magnitude of the correlation between the instrument and related measures; otherwise it is easy to rationalize whatever correlation is observed. Validity is then strengthened or weakened according to whether the hypothesized relationships are supported or refuted.

This survey was validated in the 2- to 12-year-old age group. Parents or caregivers were used as proxy respondents, since children of this age cannot reliably answer questions on complex concepts.10,13-16 Adolescent patients were excluded because the prevalence of OSDs is highest in younger children and, by including adolescents, issues would have arisen concerning the appropriateness of proxy responses for this age group. The question format in which symptom clusters were grouped under the individual domains allowed the caregiver to rate age-appropriate symptoms for their child.

The OSD-6 had obvious face and content validity in its applicability to OSDs. Cross-sectional construct validity of the OSD-6 was demonstrated by confirmation of moderate a priori correlations between related domains (Table 2). In addition, the baseline survey score that was obtained on the initial clinic visit correlated well with the global QOL 11-point rating scale and moderately well with the physician estimate of the effect of the sleep disturbance on QOL. This may reflect that a caregiver who is living with a child with an OSD may be more acutely aware of the effect of this disorder on the child's QOL, than a physician is able to appreciate in a brief office visit. Physician assessment of disease severity may accurately reflect disease activity, but have less relevance to the functioning of the child.17-19

Table 1. Median Initial OSD-6 Survey Results and Test-Retest Reliability

<table>
<thead>
<tr>
<th>Survey Item</th>
<th>Median† Score</th>
<th>Reference Range</th>
<th>ICC‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical suffering</td>
<td>5</td>
<td>2-7</td>
<td>0.71</td>
</tr>
<tr>
<td>Sleep disturbance</td>
<td>6</td>
<td>1-7</td>
<td>0.81</td>
</tr>
<tr>
<td>Speech and swallowing problems</td>
<td>4</td>
<td>1-7</td>
<td>0.74</td>
</tr>
<tr>
<td>Emotional distress</td>
<td>4</td>
<td>1-7</td>
<td>0.70</td>
</tr>
<tr>
<td>Activity limitations</td>
<td>3</td>
<td>1-7</td>
<td>0.69</td>
</tr>
<tr>
<td>Caregiver concern</td>
<td>6</td>
<td>1-7</td>
<td>0.80</td>
</tr>
<tr>
<td>Survey score</td>
<td>4.5</td>
<td>1.5-7</td>
<td>0.86</td>
</tr>
<tr>
<td>Global QOL (visual analog)</td>
<td>5</td>
<td>0-10</td>
<td>0.74</td>
</tr>
</tbody>
</table>

*OSD-6 indicates 6-item health-related instrument to assess obstructive sleep disorders; QOL, quality of life; and ICC, intraclass correlation coefficient.
†N = 100 caregivers for the following 5 domains: physical suffering, sleep disturbance, speech or swallowing problems, emotional distress, and activity limitations. n = 99 for the caregiver concern domain, survey score, and global rating scale. Mean (± SD) time between responses, 11 (± 3.9) days. ‡n = 45 for the ICC.
sponses were clustered in the midrange of response options, since it was unlikely that the physician scheduling a surgical procedure would rate the patient’s problem as having no problem or “hardly a problem at all.” If the survey were administered to all patients being evaluated for adenotonsillectomy for OSUs (rather than those who were actually scheduled for surgery), heterogeneity of the physician responses might have resulted in better correlation.

Correlation was poor between clinical ratings of the degree of airway obstruction (tonsillar size, nasal obstruction, and speech quality) and the mean survey score; our initial assumption was that these factors would correlate with the degree of sleep disturbance and, therefore, the effect on the QOL. However, in retrospect, the clinical measures we used were likely not sensitive enough to distinguish between children with varying degrees of OSUs. For example, multiple factors may be involved in craniofacial morphometrics that contribute to a propensity for an OSU in addition to adenotonsillar hypertrophy. We did not have other external measures with which to correlate the survey results to further establish construct validity, since this validation was carried out using information obtained in routine clinical practice. Sleep studies were not obtained routinely; however, another recent validation study of an OSU survey showed only modest correlation with respiratory distress indexes obtained from nap studies. Moreover, a significant proportion of children with sleep-disordered breathing have normal findings on sleep studies and nap studies lack sensitivity in detecting sleep-disordered breathing in children.

In the patients who underwent adenotonsillectomy and in whom follow-up data were available, longitudinal construct validity was demonstrated by good correlation of the change in the mean survey score with the change in the global QOL rating scale and by moderate correlation with the degree of clinical change as reported by the caregiver. The small number (36 of 62) of caregivers from whom this response was elicited may have affected the latter correlation. Again, correlation was poor with the physician estimate of change in the effect of the sleep disturbance with change in the survey score.

Evidence for responsiveness is also additional evidence for an instrument’s validity. The OSD demonstrated excellent responsiveness as judged by criteria presented by Liang et al. However, to the extent that the intervention (adenotonsillectomy) resulted in improvement which is greater than the smallest clinically important difference, the estimate of the responsiveness may be inflated.

The OSD-6 has been shown to be a valid, reliable, and responsive evaluative instrument. We believe we are, indeed, measuring the changes in health status in children with OSUs, since the OSD-6 has obvious content validity for symptoms of sleep-disordered breathing. Therefore, we can recommend its use in clinical trials to assess HRQOL changes in patients with OSUs over time, whether due to no treatment, adenotonsillectomy, or other intervention that would alter patient status. However, the modest demonstration of cross-sectional construct validity may argue against the use of the instrument to distinguish among patients with OSU with varying levels of HRQOL. Ongoing studies will hopefully provide research evidence for the instrument’s validity in this regard.

Accepted for publication May 18, 2000.

From the Division of Pediatric Otolaryngology, Children’s Hospital and Regional Medical Center (Dr de Serres) and Department of Otolaryngology—Head and Neck Surgery, University of Washington (Drs de Serres and Gates), Seattle; Department of Otolaryngology—Head and Neck Surgery, Children’s Hospital of the King’s Daughter, Eastern Virginia Medical School, Norfolk (Dr Derkay); Department of Epidemiology, University of Washington School of Public Health (Dr Astley) and Department of Medicine, University of Washington School of Medicine (Dr Deyo), Seattle; and the Division of Pediatric Otolaryngology, State University of New York, Health Science Center at Brooklyn, and Long Island College Hospital, Brooklyn (Dr Rosenfeld). The authors have no commerical, proprietary, or financial interest in the products or companies described in this article.

This study was supported in part by National Research Service Award T32 DC00018 from the National Institutes of Health, Bethesda, Md (Dr de Serres).
Presented at the annual meeting of the American Society of Pediatric Otolaryngology, Palm Desert, Calif, April 29, 1999.

This project was done as a master’s thesis for the Department of Epidemiology, University of Washington School of Public Health, Seattle; September 1999.

Special thanks to Scott Manning, MD, Kathleen Sie, MD, and Andrew Inglis, MD, for allowing their patients to participate in this study. We thank Debra Phillips, RN, Amelia Morris, RN, and Lynn Golembiewski, RN, for their help in data collection. We would also like to acknowledge Dimitra Tampakopoulou, MD, for assistance in manuscript preparation, and Guy M. McKhann II, MD, for technical assistance.

Corresponding author: Lianne M. de Serres, MD, Department of Pediatric Otolaryngology, 3959 Broadway, Room 501N, The Babies and Children’s Hospital of New York, New York, NY 10032 (e-mail: LMD54@columbia.edu).

REFERENCES
