ORIGINAL ARTICLES

Development and validation of the Nasal Obstruction Symptom Evaluation (NOSE) Scale[1]

MICHAEL G. STEWART, MD, MPH, DAVID L. WITSELL, MD, MHS, TIMOTHY L. SMITH, MD, MPH, EDWARD M. WEAVER, MD, MPH, BEVAN YUEH, MD, MPH, and **MAUREEN T. HANNLEY, PHD,** Houston, Texas, Durham, North Carolina, Milwaukee, Wisconsin, Seattle, Washington, and Alexandria, Virginia

OBJECTIVE: **The study goal was to validate a disease-specific health status instrument for use in patients with nasal obstruction.**

DESIGN, SETTINGS, AND PATIENTS: **The study consisted of a prospective instrument validation conducted at 4 academic medical centers with 32 adults with nasal septal deformity.**

METHODS: **Prospective instrument validation occurred in 2 stages. Stage 1 was the development of a preliminary (alpha-version) instrument of potential items. Stage 2 was a test of the alpha-version for item performance, internal consistency, and test-retest reliability; construct, discriminant, criterion validity, and responsiveness; and creation of the final instrument.**

RESULTS: **Items with poor performance were eliminated from the alpha-version instrument. In testing the final instrument, test-retest reliability was adequate at 0.702; internal consistency reliability was also adequate at 0.785. Validity was confirmed using correlation and comparison analysis, and response sensitivity was excellent.**

CONCLUSIONS: **The Nasal Obstruction Symptom Evaluation Scale is a valid, reliable, and responsive instrument that is brief and easy to complete and has potential use for outcomes studies in adults with nasal obstruction.** (Otolaryngol Head Neck Surg 2004;130:157-63.)

From the Bobby R. Alford Department of Otorhinolaryngology and Communicative Sciences, Baylor College of Medicine (Dr Stewart), the Division of Otolaryngology–Head and Neck Surgery, Department of Surgery, Duke University School of Medicine (Dr Witsell), the Department of Otolaryngology and Communication Sciences, Medical College of Wisconsin (Dr Smith), the Department of Otolaryngology–Head and Neck Surgery (EMW, BY), University of Washington School of Medicine (Drs Weaver and Yueh), and the Surgery and Perioperative Care and Health Services Research and Development Services (Drs Weaver and Yueh), and VA Puget Sound Health Care System, Seattle, WA, and the American Academy of Otolaryngology–Head and Neck Surgery Foundation (Dr Hannley).

Supported by Career Development Award CD-98318 from the Health Services Research and Development Service of the Veterans Health Administration, Department of Veterans Affairs (Dr Yueh).

Reprint requests: Michael G. Stewart, MD, MPH, Baylor College of Medicine, The Bobby R. Alford Department of Otolaryngology and Communicative Sciences, 6550 Fannin St, Suite 1727, Houston, TX 77030; e-mail, mgstew@bcm.tmc.edu.

[1]The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

0194-5998/$30.00

10.1016/j.otohns.2003.09.016

Surgical and medical treatments for nasal obstruction are a common part of otolaryngology practice. Although symptoms of nasal obstruction can have several etiologies, such as mucosal congestion, turbinate hypertrophy, adenoid hypertrophy, nasal mass, and others, deviation of the nasal septum is a very common cause. The definitive treatment for a deviated nasal septum is surgical correction, or septoplasty. However, there is limited evidence to document whether septoplasty is effective in improving symptoms of nasal obstruction.

There are 2 potential ways to assess outcome after nasal septoplasty: objective measures and subjective measures. The objective assessment of nasal obstruction is controversial, and there is no agreement on an accepted measurement tool.[1-4] There have been several prior studies of patients' subjective outcome after septoplasty.[5-9] However, many of those studies were retrospective, and the prospective studies have used either nonvalidated survey instruments or instruments that were not validated for the assessment of nasal obstruction. Most studies us-

ing objective and subjective outcomes have supported the concept that septoplasty is effective in achieving the outcomes measured, but there has been no prospective study using an outcomes instrument validated for nasal obstruction.

To perform a prospective assessment of subjective treatment outcomes, a validated outcome instrument is needed, which was the purpose of this study. The development and validation of the outcomes instrument, named the Nasal Obstruction Symptom Evaluation (NOSE) Scale, was conducted as part of a parallel prospective multicenter observational clinical study. This multicenter study was commissioned and funded by the American Academy of Otolaryngology–Head and Neck Surgery Foundation and coordinated under the auspices of its National Center for the Promotion of Research in Otolaryngology.

METHODS

We performed a multicenter prospective instrument validation study in two phases. The study was approved by the institutional review boards at all participating institutions, and data collection and follow-up were coordinated at the Duke Clinical Research Institute. Data analysis was carried out by one of the authors (M.G.S.) using SPSS version 10.0 statistical software (SPSS Inc, Chicago, IL).

Subjects

Patients were enrolled consecutively into the prospective protocol from the clinical practices of the authors and their partners, using strict inclusion and exclusion criteria. The enrollment period was July 1, 2002, through January 31, 2003. All septoplasty patients were eligible. Inclusion criteria for the study were age of at least 18 years; septal deviation consistent with presenting symptom of chronic nasal obstruction; symptoms lasting at least 3 months; and persistent symptoms after a 4-week trial of medical management, including either topical nasal steroids, topical or oral decongestants, or oral antihistamine–decongestant combinations. Exclusion criteria were sinonasal malignancy; radiation therapy to the head and neck; septoplasty performed with concurrent sinus surgery, rhinoplasty, or sleep apnea surgery; septoplasty performed as access to other sites; prior septoplasty, rhinoplasty, or turbinoplasty; history or clinical evidence of chronic sinusitis (using the Academy of Otolaryngology–Head and Neck Surgery Rhinosinusitis Task Force definition); septal perforation; craniofacial syndrome; acute nasal trauma or fracture in the past 3 months; nasal valve collapse; adenoid hypertrophy; sarcoidosis; Wegener's granulomatosis; uncontrolled asthma; pregnancy; and illiteracy. Limited demographic data were collected on patients who refused enrollment or were excluded.

An additional group of patients for between-group discrimination was collected as a convenience sample from the Baylor College of Medicine and Duke University sites. Patients who were seen in otolaryngology clinics for nonrhinologic complaints, such as hearing loss or hoarseness, were asked to complete the alpha-version of the NOSE Scale.

Phase 1: Instrument Development

The authors, all of whom have had training and experience in health services research, item selection, and instrument validation, reviewed all available disease-specific instruments designed to assess nasal symptoms, rhinitis, or sinusitis. Any individual items pertaining to nasal obstruction were extracted for further consideration. The authors then developed additional items reflecting other aspects of nasal obstruction as described by patients. These obstruction-related items were then grouped into a preliminary (alpha-version) instrument, which contained 10 items (Fig 1). All 10 items were scored using a 5-point Likert scale (not a problem, very mild problem, moderate problem, fairly bad problem, and severe problem) and were phrased, "Over the past one month, how much of a problem (was). . .?" In the alpha-version instrument, all aspects of nasal obstruction that were covered in the existing sinonasal instruments were included, and some new constructs related to nasal obstruction were added, such as panic and embarrassment.

Phase 2: Instrument Validation

A convenience sample of patients whose planned surgery date was at least 1 week after the enrollment visit were sent the alpha-version in-

Nasal Obstruction Symptom Evaluation (NOSE) Instrument

Physician AAO-HNS#: _ _ _ _ _ _ **Today's date:** _ _/_ _/_ _ _ _

➔ **To the Patient:** Please help us to better understand the impact of nasal obstruction on your quality of life by **completing following survey**. Thank You!

Over the past 1 month, how much of a problem were the following conditions for you?

	Please circle the most correct response				
	Not a problem	*very mild problem*	*moderate problem*	*fairly bad problem*	*severe problem*
1. Nasal congestion or stuffiness	0	1	2	3	4
2. Poor sense of smell	0	1	2	3	4
3. Snoring	0	1	2	3	4
4. Nasal blockage or obstruction	0	1	2	3	4
5. Trouble breathing through my nose	0	1	2	3	4
6. Trouble sleeping	0	1	2	3	4
7. Having to breathe through my mouth	0	1	2	3	4
8. Unable to get enough air through my nose during exercise or exertion	0	1	2	3	4
9. Feeling panic that I cannot get enough air through my nose	0	1	2	3	4
10. Embarrassment around friends and coworkers because I have trouble breathing through my nose	0	1	2	3	4
	Poor	*Fair*	*Good*	*Very good*	*Excellent*
11. In general, my health is	0	1	2	3	4

Please mark on this line how troublesome is your difficulty in breathing through your nose:

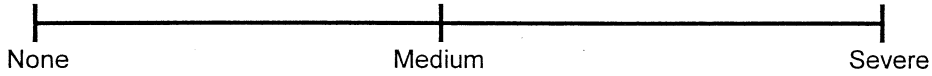

Fig 1. The alpha-version of the NOSE instrument. (The NOSE Scale © 2003, the American Academy of Otolaryngology–Head and Neck Surgery Foundation.)

strument again in 7 to 14 days, but before surgical intervention–to assess test-retest reliability. Reliability was assessed using the Goodman-Kruskal γ coefficient; a value of at least 0.70 was considered adequate test-retest reliability.[10]

A table was constructed with individual items in rows, and statistical attributes of the items, such as mean, median, range, standard deviation, and skewness, in columns. In addition, a bar graph of response frequency for each item was created, to

view the response distributions. Items with large floor or ceiling effects, or skewed distributions of responses (such a large proportion of patients reporting "not a problem"), were marked for possible elimination.

Internal consistency reliability was assessed by calculating the Cronbach α coefficient[10,11] and noting item–total correlations. Internal consistency reliability was considered adequate if at $\alpha \geq 0.70$[11]. Items that did not contribute to the overall internal consistency of the instrument were marked for possible elimination.

Content validity was ensured during design of the instrument in phase 1, as described earlier. Construct validity was assessed using a combination of several techniques: principal components factor analysis; item–item and item–total correlations; and between-group discrimination. Principal components factor analysis was performed on all 10 items using orthogonal varimax rotation of factors.[12] All factors with an eigenvalue of greater than 1.0 were retained for potential inclusion in the final rotated factor solution. Additional construct validity was assessed by constructing a table of item–item and item–total Spearman correlation coefficients; a significant level of association was set as a coefficient of $\geq$0.40. As different instrument lengths were explored—after individual items were considered for elimination—item–total correlations were also assessed, again using Spearman coefficients and the same level of significance. Between-group discrimination (discriminant validity) was assessed by comparing total scores between study patients and the sample of patients seen for nonrhinologic complaints. The Mann-Whitney *U* test was used to compare groups; a significant difference was expected and was defined as $P < 0.05$.

Criterion validity was assessed by comparing obstruction-specific items with a summary visual analog scale measuring "difficulty in breathing through your nose," which was included with the alpha-version instrument. Statistical analysis was performed using the Spearman correlation coefficient; a significant association was set at a coefficient of $\geq$0.40.

The results of individual item analysis were combined with reliability and validity analysis to construct a final version of the instrument by eliminating items that were redundant, that showed poor statistical distribution, and that did not contribute to overall reliability and validity. This final version of the instrument was then evaluated for response sensitivity.

Response sensitivity of the final instrument—scoring only the final included items—was assessed by calculating the standardized response mean and the effect size[11,13,14] and comparing those values with published standards. For both measures, a value of approximately 0.2 represents low sensitivity to change, 0.5 represents moderate sensitivity, and 0.8 represents high sensitivity to change.[14] Data were used from the 3-month postoperative data collection point. All patients had undergone nasal open septoplasty or submucous resection, after failure of medical therapy to relieve obstruction.

RESULTS

A total of 32 patients were enrolled in the study, from 4 sites of author affiliation (Baylor College of Medicine, Duke University, Medical College of Wisconsin, and University of Washington) (25 men and 7 women; mean age, 47.0 years; age range, 19 to 78 years). Twenty-one patients completed the test-retest portion of the validation process. There is no sample size or power calculation for psychometric validation, but a general rule of thumb is that 25 to 50 patients make an adequate sample for validation.

Initial item review of the alpha instrument demonstrated that items 9 and 10 ("panic" and "embarrassment") had poor statistical performance, with a large floor effect, a low median, and a large right (positive) skew, so those items were eliminated. Performing principal components factor analysis on the remaining 8 items revealed that items loaded on 2 major components: items 1, 4, 5, 6, 7, and 8 loaded onto a primary factor, and items 2 ("sense of smell") and 3 ("snoring") loaded onto a secondary factor. Analysis of internal consistency of those 8 items demonstrated adequate internal consistency ($>$0.70) but also demonstrated that items 2 and 3 each detracted from internal consistency (ie, removal of each item resulted in a higher value of Cronbach's α). Therefore items 2 and 3 were eliminated.

Table 1. Item–item and item–total correlations for the 5 item instrument

	Congestion	Obstruction	Trouble breathing	Sleeping	Exercise
Congestion					
Obstruction	0.665*				
Trouble breathing	0.599*	0.58*			
Sleeping	0.321	0.386	0.369		
Exercise	0.327	0.447*	0.646*	0.398	
Total	0.695*	0.753*	0.805*	0.738*	0.778*

Values shown are the Spearman correlation coefficient.
*Correlations greater than 0.40.

Next, interitem correlations were checked between the remaining 6 items. The correlation between items 5 ("trouble breathing through nose") and 7 ("having to breathe through mouth") was very high at 0.698, demonstrating that the 2 items were very close to measuring the same underlying construct. Testing removal of each item demonstrated that removal of item 7 resulted in a higher internal consistency than removal of item 5, so item 7 was removed. The instrument then contained 5 items: items 1, 4, 5, 6, and 8. Internal consistency reliability of this 5-item instrument was high at $\alpha = 0.785$.

Test-retest reliability was assessed using the total score on the 5-item instrument in a subgroup of patients (n = 21). The mean time between instrument administrations was 8.7 days. The reliability coefficient was adequate at $\gamma = 0.702$.

As part of the confirmation of construct validity, the interitem and item–total correlations are shown in Table 1. This analysis demonstrated several expected associations. First, the items each have a high correlation with the subscale total, which is expected because items with poor item–total correlation were eliminated. Next, the items related to obstruction, congestion, and trouble breathing had high levels of correlation with each other but poor correlation with the item on sleeping. The item on trouble breathing during exercise had a high correlation with the item on trouble breathing through the nose, but weak or absent correlation with items on congestion and obstruction, and no correlation with the item on sleeping. These findings confirm that the items are measuring related but distinct individual concepts that nevertheless form a unified construct.

Table 2. Spearman correlation coefficients between the visual analog summary scale and individual items

Item	Content	Coefficient
1	Nasal congestion or stuffiness	0.64*
2	Nasal blockage or obstruction	0.55*
3	Trouble breathing through nose	0.76*
4	Trouble sleeping	0.21
5	Unable to get air through nose during exercise	0.47*

*Correlations greater than 0.40.

The group of nonrhinologic patients consisted of 12 patients, and they had a mean raw score (on the 5-item instrument) of 2.17, whereas the 32 study patients had a mean raw score on the same instrument of 13.13 (Mann-Whitney U test, $P < 0.001$). Therefore the instrument demonstrated excellent between-group discrimination.

Criterion validity was assessed by measuring the association of obstruction items with the single summary visual analog scale. The values are shown in Table 2, demonstrating that items assessing obstruction correlate well with the visual analog scale, and the item assessing sleeping difficulty did not correlate well. This analysis confirms the criterion validity of the instrument.

To assess response sensitivity of the instrument, we calculated the standardized response mean and effect size for 21 patients who completed the instrument at the 3-month postoperative follow-up period. The standardized response mean for the 5-item instrument was 1.66, which indicates very high sensitivity to change. The effect size was 2.65, which also indicates very high sensitivity.

Otolaryngology–
Head and Neck Surgery
February 2004

Nasal Obstruction Symptom Evaluation (NOSE) Instrument

→ **To the Patient:** Please help us to better understand the impact of nasal obstruction on your quality of life by **completing the following survey**. Thank You!

Over the past 1 month, how much of a problem were the following conditions for you?

Please circle the most correct response

	Not a problem	*very mild problem*	*moderate problem*	*fairly bad problem*	*severe problem*
1. Nasal congestion or stuffiness	0	1	2	3	4
2. Nasal blockage or obstruction	0	1	2	3	4
3. Trouble breathing through my nose	0	1	2	3	4
4. Trouble sleeping	0	1	2	3	4
5. Unable to get enough air through my nose during exercise or exertion	0	1	2	3	4

Fig 2. The final version of the NOSE instrument. (The NOSE Scale © 2003, the American Academy of Otolaryngology–Head and Neck Surgery Foundation.)

Because of the scoring algorithm for each item, the range of raw scores on the final instrument was from 0 to 20. The instrument was then scaled to a total score of 0 to 100 by multiplying the raw score by 5. Because of item wording, a score of 0 means no problems with nasal obstruction and a score of 100 means the worst possible problems with nasal obstruction.

The final version of the instrument is shown in Figure 2.

DISCUSSION

Although global quality-of-life and health status instruments are an important part of health status assessment, for many conditions the changes in health status are too subtle or disease specific to be assessed using the content of a global instrument. Therefore disease-specific health status instruments are needed.[15] This has been shown to be true in many diseases, including visual loss from cataracts[16] and hearing loss.[17]

We completed the validation of a disease-specific instrument designed to assess nasal obstruction: the NOSE Scale. The instrument is brief and easy to complete, with minimal respondent burden. This is important if the instrument is going to be given repeatedly in prospective trials. It is also reliable, valid, and responsive to change in clinical status, as demonstrated with the data presented here. This means that scores on the instrument remain consistent when the underlying patient's status does not change (ie, there is little random or spurious error in the assessment), that the instrument is measuring what is supposed to measure, and that the scores on the instrument do respond when the patient's underlying status changes.

Like many similar instruments, the NOSE Scale was validated for use in groups of patients, not individual patients. Therefore it could be used for

comparing disease-specific health status between groups of patients before and after treatment, or used to compare the effects of different treatments, for example, medical versus surgical therapy. Similarly, it could be used to assess differences in outcome when different surgical techniques are used. It could also be used to compare symptom severity between different groups of patients, for example, those with and without nasal polyps. However, it was not designed to be used with individual patient data or to predict outcome in individuals.

The brevity of the instrument does not detract from its sensitivity. In fact, studies have shown that shorter instruments might be more sensitive to change in clinical status than longer instruments.[13,14]

The NOSE Scale could also be used with a global or generic quality of life instrument, to assess the relative impact of the specific disease on different aspects of global quality of life. When comparing the impact of different disease states, those global data can be helpful. However, the item content for many diseases is so unique that a disease-specific instrument is usually needed to be able to assess the impact of disease treatment.

There are several other validated instruments that are available for use in rhinology. These include the Chronic Sinusitis Survey (CSS),[18] the Rhinoconjunctivitis Quality of Life Questionnaire (RQLQ),[19] the Rhinosinusitis Disability Index (RSDI),[20] the Sino-Nasal Outcome Tool (SNOT-20),[21] and the Allergy Outcome Survey (AOS).[22] The CSS, RSDI, and SNOT-20 were all designed to be used in patients with chronic sinusitis, and not purely nasal obstruction. The RQLQ and AOS were primarily designed for patients with allergic rhinitis and conjunctivitis. All of these instruments are valid and reliable and have been used successfully in prospective outcomes studies on chronic infectious sinusitis and on allergic rhinitis.

We believe that the NOSE Scale is a valuable and unique addition to this armamentarium of valid and reliable instruments that can be used for outcomes research in rhinology.

REFERENCES

1. Katz S, Schmelzer B, Vidts G. Treatment of the obstructive nose by CO2-laser reduction of the inferior turbinates: technique and results. Am J Rhinol 2000;14:51-5.
2. Pirila T, Tikanto J. Unilateral and bilateral effects of nasal septum surgery demonstrated with acoustic rhinometry, rhinomanometry, and subjective assessment. Am J Rhinol 2001;15:127-33.
3. Passali D, Lauriello M, Anselmi M, et al. Treatment of hypertrophy of the inferior turbinate: long-term results in 382 patients randomly assigned to therapy. Ann Otol Rhinol Laryngol 1999;108:569-75.
4. Jessen M, Malm L. The importance of nasal airway resistance and nasal symptoms in the selection of patients for septoplasty. Rhinology 1984;22:157-64.
5. Siegel NS, Gliklich RE, Taghizadeh F, et al. Outcomes of septoplasty. Otolaryngol Head Neck Surg 2000;122:228-32.
6. Stewart EJ, Robinson K, Wilson JA. Assessment of patient's benefit from rhinoplasty. Rhinology 1996;34:57-9.
7. Samad I, Stevens HE, Maloney A. The efficacy of nasal septal surgery. J Otolaryngol 1992;21:88-91.
8. Arunachalem PS, Kitcher E, Gray J, Wilson JA. Nasal septal surgery: evaluation of symptomatic and general health outcomes. Clin Otolaryngol 2001;26:367-70.
9. Dinis PB, Haider H. Septoplasty: long-term evaluation of results. Am J Otolaryngol 2002;23:85-90.
10. DeVellis RF. Scale development: Theory and applications. Newbury Park: Sage Publications; 1991.
11. Aday LA. Designing and conducting health surveys: a comprehensive guide, 2nd ed. San Francisco: Jossey-Bass Publishers; 1996.
12. Lipsey MW. Design sensitivity: Statistical power for experimental research. Newbury Park: Sage Publications; 1990.
13. Gliklich RE, Hilinski JM. Longitudinal sensitivity of generic and specific health measures in chronic sinusitis. Qual Life Res 1995;4:27-32.
14. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28:632-42.
15. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care 1989;27:S217-32.
16. Damiano AM, Steinberg EP, Cassard SD, et al. Comparison of generic versus disease-specific measures of functional impairment in patients with cataract. Med Care 1995;33:AS120-30.
17. Stewart MG. Outcomes and patient-based hearing status in conductive hearing loss. Laryngoscope 2001;111:11 1-21.
18. Gliklich RE, Metson R. Techniques for outcomes research in chronic sinusitis. Laryngoscope 1995;105:387-90.
19. Juniper EF, Guyatt GH. Development and testing of a new measure of health status for clinical trials in rhinoconjunctivitis. Clin Exp Allergy 1991;21:77-83.
20. Benninger MS, Senior BA. The development of the Rhinosinusitis Disability Index. Arch Otolaryngol Head Neck Surg 1997;123:1175-9.
21. Piccirillo JF, Merritt MG, Richards ML. Psychometric and clinimetric validity of the 20-item Sino-Nasal Outcome Test (SNOT-20). Otolaryngol Head Neck Surg 2001;126:41-7.
22. Kemker BJ, Corey JP, Branca J, et al. Development of the Allergy Outcome Survey for allergic rhinitis. Otolaryngol Head Neck Surg 1999;121:603-5.