Validation of the Sino-Nasal Outcome Test 20 (SNOT-20) domains in nonsurgical patients

Melissa A. Pynnonen, M.D.,* H. Myra Kim, Sc.D.,# and Jeffrey E. Terrell, M.D.*

ABSTRACT

Background: *The objectives of this study were, first, to confirm the presence of multiple domains within the Sino-Nasal Outcome Test 20 (SNOT-20) using a medically treated population, and, second, to reanalyze data from this population to reveal incremental information. A prospective, randomized controlled trial was performed.*

Methods: *One hundred twenty-seven adults with chronic rhinitis or rhinosinusitis symptoms were treated with nasal saline irrigation or spray. Treatment outcome was quality of life measured with SNOT-20 scores, which were reanalyzed for this study with a factor analysis. Differences in change scores were compared.*

Results: *Factor analysis confirmed the presence of four domains: psychological function, sleep function, rhinological symptoms, and ear and/or facial symptoms. At 8 weeks after randomization, saline irrigation had significant effects on the rhinological symptom ($p = 0.01$) and sleep ($p = 0.01$) compared with saline spray, but no between-group difference was seen in psychological function or ear and/or facial symptom domains.*

Conclusion: *Subscales identified differences in the impact of two medical interventions on chronic sinonasal symptoms. Reporting subscale scores might improve the precision of the SNOT-20 instrument, allowing discrimination between various treatments and their differential impact on sinonasal quality of life.*

(Am J Rhinol Allergy 23, 40–45, 2009; doi: 10.2500/ajra.2009.23.3259)

Key words: Adult, factor analysis, humans, nasal saline irrigation, prospective studies, quality of life, rhinitis, rhinosinusitis, severity of illness index, treatment outcome

The Sino-Nasal Outcome Test 20 (SNOT-20) is one of the most widely used quality-of-life instruments for sinonasal conditions[1–3] and is intended for populations of people with rhinosinusitis, rather than simply rhinitis. The SNOT-20 is a self-administered multiple-choice 20-item test that is usually scored with a single summary score (0–5) without domains or subscales. This instrument assesses a broad range of health and health-related quality-of-life problem including physical problems, functional limitations, and emotional consequences, but unlike many of the quality-of-life instruments designed to measure rhinitis symptoms, this rhinosinusitis measure is not divided into subscales or domains related to these different areas.[4] For instance, there are not separate domains for nasal, eye, or ear symptoms; sleep quality or pain; or social or emotional concerns. As such, it is possible that the SNOT-20 might be unresponsive to small but important changes in health-related quality of life. For instance, an intervention that has a positive impact on nasal complaints but a similarly sized negative impact on sleep symptoms may not be associated with any changes on overall (summated) scores and may lead to a false impression of "no detectable effect."

Browne *et al.* recently proposed and performed some validation studies on a modified SNOT-20 to determine if, indeed, there was more than one domain or construct measured by the instrument.[5] The authors used a population of patients who had presented to an otolaryngologist in the United Kingdom and had completed pre- and postoperative SNOT-20 surveys related to surgery for nasal polyps or nonpolypoid chronic rhinosinusitis. Their validation studies supported dividing the SNOT-20 into four domains: a rhinologic, ear and facial symptoms, sleep, and psychological domain.[5] The new rhinologic domain contained five questions: need to blow nose, sneezing, runny nose, postnasal discharge, and thick nasal discharge. The ear and facial symptoms domain contained four questions: ear fullness, dizziness, ear pain, and facial pain/pressure. The sleep domain contained three questions: difficulty falling asleep, waking up at night, and lack of a good night's sleep. The psychological domain contained six questions: fatigue, reduced productivity, reduced concentration, frustration/restlessness/irritability, sadness, and embarrassment. Two questions (cough and waking up tired) were not classified into any of these domains.

Browne *et al.* found that dividing the SNOT-20 into four separate domains revealed new information: *viz.*, that both groups of patients, those with polyps as well as those without polyps, improved in the rhinologic, ear/facial, sleep, and psychological function domains, but that a substantial portion of the improvement and the total SNOT-20 scores in the group of patients with nasal polyps were due to improvement in the rhinologic domain.

The objectives of this study were to confirm the domains of SNOT-20 in a medically treated patient population, which is different from the surgically treated population studied by Browne *et al.*,[5] and, more importantly, to gain a better understanding of the effect of saline irrigation based on applying the domain scores to data from a randomized clinical trial comparing saline intervention with saline spray in patients

*From the *Department of Otolaryngology, University of Michigan Health System, Ann Arbor, Michigan, and #Department of Biostatistics and Center for Statistical Consultation and Research, University of Michigan, Ann Arbor, Michigan*
Address correspondence and reprint requests to Melissa A. Pynnonen, M.D., Department of Otolaryngology, University of Michigan Health System, 1904 Taubman Center, 1500 East Medical Center Drive, Ann Arbor, MI 48109-0312
E-mail address: pynnonen@umich.edu
Copyright © 2009, OceanSide Publications, Inc., U.S.A.

with chronic sinonasal symptoms. We hypothesized that patients with chronic sinonasal symptoms experience greater improvement in the rhinologic domain of the SNOT-20 with saline irrigation compared with saline spray.

METHODS

Subjects and Data Collection

The data for this study were obtained from a previously published clinical trial, comparing saline spray with saline irrigation.[6] The study participants were adults recruited from the general population with chronic sinonasal symptoms. No attempt was made to diagnose the cause of the sinonasal symptoms, but the inclusion criteria required nearly daily sinonasal symptoms for at least the preceding month and excluded subjects with acute upper respiratory tract infections. In contrast to the Browne study, this was an entirely nonsurgical population. Subjects were randomized to saline spray or saline irrigation treatment with measurements made at baseline and 2, 4, and 8 weeks after randomization. Outcome comparisons were based on the overall SNOT-20 summary scores.[6]

We did factor analysis of the SNOT-20 data to confirm the domains described by Browne *et al.*[5] We then compared the differences in domain change scores at 8 weeks between the two treatment groups: saline spray or saline irrigations. The study protocol for the clinical trial that gave rise to the SNOT-20 data had been reviewed and approved by the University of Michigan Health System Institutional Review Board.

Statistical Analysis

Validation of Subscales. To identify and validate the potential subscales, we did confirmatory factor analysis of the baseline responses from the SNOT-20. This was to assess the agreement between hypothetical factors that go to make up the instrument and the subscales designed to assess those factors. Factor analysis was done with a maximum likelihood method.[7,8] If the SNOT-20 is a valid measure for use with the subscales defined by Browne *et al.*,[5] the same factors should emerge from a factor analysis of this data, and items relating to a particular subscale should be grouped together within a single factor.

Internal Consistency. The SNOT-20 would be internally consistent if the responses to questions that contribute to the same subscale correlate well with each other. To assess the internal consistency of the SNOT-20 and its subscales, we used item-rest correlations and Cronbach's α.[9] Item-rest correlations assess the extent to which an item (question) is related to the remainder of its scale and should exceed 0.2. Furthermore, items should be more closely related to their own subscale than to the other subscales. Cronbach's α measures the overall correlation between items within a subscale. Reliability is considered acceptable for group comparisons when α exceeds 0.7.

Subscales and Comparison between Intervention Groups. After confirming the subscales defined by Browne *et al.*[5] in our sample, we used the subscales to assess the treatment effect on each of the different domains of sinonasal symptoms. The subscale scores were generated using a summative scoring method.[5] A two-sample *t*-test was used to compare the between-group differences in change scores of each of the four subscales. The change scores at follow-up times were calculated as the difference from preintervention (baseline) subscale scores, and a large positive change score corresponds to a large improvement in the specific sinonasal condition. In addition to mean change scores, standardized effect sizes were calculated by dividing the mean change scores by its standard deviation. The effect size allows comparisons between the two intervention groups across all four subscales. All statistical analyses were done using Stata 9.2 (Stata Corp., College Station, TX).

RESULTS

A total of 193 potential subjects were screened, and a total of 127 eligible subjects were enrolled and randomized to either the irrigation or the spray group. The groups were similar in age, sex, extent of education, ethnicity, smoking status, symptom burden, and symptom duration as well as baseline mean SNOT-20 summary scores. In contrast to the Browne *et al.*[5] study, there were fewer men in our study (42/127, 33%) but the mean age of the subjects (46.6 years) was the same as the Browne study (48 years and 62% men).

None of the 127 participants had any missing values for any of the 20 SNOT-20 items. Based on eigenvalues of >1.0, four factors were identified as reported by Browne *et al.*[5] with eigenvalues of 6.2, 1.8, 1.6, and 1.1. The four factors captured 67% of the total variance. After factors were rotated with an oblique promax rotation, we found nearly identical correlation (loading) between individual items and each corresponding factor with the exception of fatigue (Table 1). Factor 1 covered psychological function, and Browne *et al.*[5] included fatigue in this factor. However, fatigue was found to load on both psychological symptom factor and sleep symptom factor in our data. Because fatigue loaded higher with psychological symptom factor, we kept fatigue in psychological symptom subscale as suggested by Browne *et al.*[5] and this led to six items for the subscale: fatigue, reduced productivity, reduced concentration, frustration/restlessness/irritable, sad, and embarrassed. Factor 2 covered sleep function and included three items: difficulty sleep, wake at night, and lack of sleep. Factor 3 covered the rhinological symptoms and included five items: need to blow nose, sneezing, runny nose, postnasal discharge, and thick nasal discharge. Factor 4 covered ear and facial symptoms and contained four items: ear fullness, dizziness, ear pain, and facial pain/pressure. Similar to Browne *et al.*[5], two items were not classifiable: cough, which did not load on any factor, and wake up tired, which loaded on two factors. The final communality was 0.79 as measured by Kaiser-Meyer-Olkin sampling adequacy measure.

Overall internal consistency was good with item-rest correlations ranging from a minimum of 0.30 (blow nose) to a maximum of 0.74 (reduced concentration). Each item also was more closely related to its own subscale than the other subscales, shown by higher item-rest correlation with its own subscale than any subscales it is not included in. For example, the item blow nose had an item-rest correlation of 0.52 with the rhinological symptom scale without blow nose item, while its correlation was 0.15 with ear/facial symptoms subscale, 0.14 with sleep function subscale, and 0.17 with psychological

Table 1 **Sino-Nasal Outcome Test 20 questions loading to each of the four Sino-Nasal Outcome Test subscales**

Question	Psychological Function	Sleep Function	Rhinological Symptoms	Ear/Facial Symptoms
Need to blow nose	−0.12	−0.06	0.71	0.11
Sneezing	0.02	0.14	0.45	0.05
Runny nose	0.03	−0.10	0.73	0.09
Cough*	0.22	0.05	0.31	−0.12
Postnasal discharge	−0.01	0.17	0.49	−0.11
Thick nasal discharge	0.19	0.00	0.46	−0.02
Ear fullness	0.04	0.04	0.15	0.59
Dizziness	0.31	0.00	−0.02	0.55
Ear pain	0.18	0.11	−0.05	0.54
Facial pain/pressure	0.06	0.05	0.10	0.58
Difficulty sleep	0.25	0.56	0.11	0.03
Wake at night	−0.00	0.77	0.09	0.02
Lack of sleep	0.06	0.99	−0.06	−0.02
Wake up tired*	0.41	0.61	−0.08	0.09
Fatigue*	0.52	0.40	−0.07	0.13
Reduced productivity	0.85	0.12	−0.05	0.05
Reduced concentration	0.82	0.06	−0.01	0.14
Frustration/restless/irritable	0.62	0.18	0.21	−0.03
Sad	0.57	0.07	0.12	0.11
Embarrassed	0.45	0.01	0.20	−0.18

**Cough, wake up tired, and fatigue did not load clearly on any construct.*

function subscale. Cronbach's α for each subscale was greater than 0.7 showing high internal consistency across items within each of the four subscales: 0.72 for rhinological symptom subscale, 0.74 for ear/facial symptom subscale, 0.87 for sleep function subscale, and 0.88 for psychological function subscale.

We generated four subscales using identical items to the subscales formed by Browne *et al.*[5] The psychological subscale included six items with scores that may vary from 0 to 30, and rhinological symptom subscale included five items with scores that may vary from 0 to 25. The ear and/or facial symptoms subscale had four items with scores that may vary from 0 to 20, and the sleep function subscale had three questions with scores that may vary from 0 to 15. We also generated a five-item psychological subscale, after excluding item fatigue, which in our sample loaded on both psychological subscale and on sleep subscale. The distribution of the subscale scores showed some floor effect with 7, 13, 16, and 19% of patients scoring the minimum score of 0 (responses of "no problem" for every item in each subscale) for rhinological, ear/facial, sleep, and psychological subscale, respectively. On the other hand, little ceiling effect was seen with only 6% scoring the maximum sleep subscale score (responses of "problem as bad as it can be" for every item in the subscale), while 0% scoring maximum scores for other subscales (Fig. 1). On the contrary, 0% of the patients had the minimum score for SNOT-20 summary scale, with its distribution fairly symmetric.

Each of the four subscales showed somewhat different within and between saline irrigation and spray group comparisons. Table 2 shows the mean within-group change scores for each subscale at week 8, and Fig. 2 shows the outcome trends over time for each subscale with cross-sectional means at 2, 4, and 8 weeks. The rhinological symptom subscale showed significant improvement within both groups, but even greater improvements were seen in the irrigation group than in the spray group. Although sleep showed no improvement in the saline spray group, there was significant improvement seen in the irrigation group at each follow-up time, and thus there were significant between-group differences at each follow-up time. The ear and/or facial symptom subscale and the psychological function subscale showed significant improvements within each group, but no difference was seen between the groups. In summary, compared with saline spray, saline irrigation appears to improve related sleep function and rhinologic symptoms most significantly. In fact, the saline spray group did not show improvement in sleep function at any of the three follow-up times. Psychological function subscale result was nearly identical when the subscale was constructed using five items after excluding fatigue.

Relative differences across the subscales were compared using standardized effect size (Fig. 3). In saline irrigation group, all four subscales showed a medium effect size (effect size = 0.5) or larger, but at both week 4 and week 8, rhinological symptom subscale showed the largest improvement in saline irrigation group. In saline spray group, all subscales showed small to medium effect size except sleep, which showed minimal improvement.

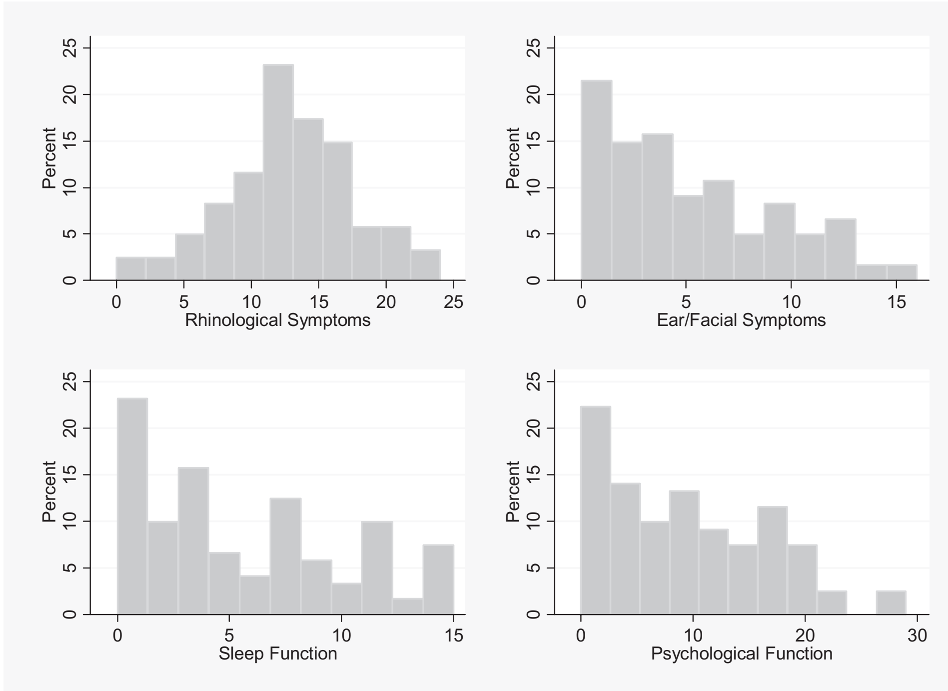

Figure 1. *Distribution of SNOT 20 subscales.*

Table 2 **Change from preintervention at 8 weeks follow-up time in Sino-Nasal Outcome Test 20 subscale scores for the saline spray group ($n = 61$) and the irrigation group ($n = 59$)**

Group	Rhinological	Ear/Facial	Sleep	Psychological
Spray	3.2 (4.5)*	2.0 (3.4)*	0.3 (4.2)	2.5 (6.3)#
Irrigation	5.4 (5.0)*	2.0 (3.4)*	2.3 (4.3)*	3.6 (6.8)*
p Value§	0.01	1.00	0.01	0.36
95% CI¶	2.23 (0.48, 3.98)	−0.001 (-1.25, 1.25)	2.02 (0.44, 3.60)	1.13 (−1.29, 3.55)

Cell values are mean (standard deviation), and a large magnitude corresponds to large improvement from preintervention.
Significant improvement based on paired t-test ($p < 0.001$).
#Significant improvement based on paired t-test ($0.001 \leq p < 0.05$).
§Based on comparisons between the saline spray and irrigation group using two-sample t-test at each week.
¶Differences in mean change scores between the irrigation group and saline spray group and its 95% confidence interval.

DISCUSSION

This study confirmed the findings of Browne *et al.*,[5] *viz.*, that the SNOT-20 measures more than one unique construct. Our data support the concept that the SNOT-20 can be effectively divided into domains of rhinologic, ear/facial symptoms, sleep symptoms, and psychological function. Interestingly, our factor analysis of the 20 questions of the SNOT-20 led to an almost identical grouping of questions, with only one item, fatigue, not loading well into the psychological functioning domain in our population. It is possible that this difference may be related to a different population being studied. Browne *et al.*[5] surveyed preoperative and postoperative patients aged ≥16 years who were undergoing surgical procedures to treat chronic rhinosinusitis, with or without nasal polyps. Our population consisted of adults who had bothersome nasal symptoms more days than not over the preceding 30 days and excluded patients who had undergone recent sinus surgery. Despite the population differences, the internal reliability measures of these new domains, as measured by calculating Cronbach's α, were strong, showing high internal consistency across items within each of the four domains. This very minimal difference on factor analysis between substantially different populations would support the generalizability of the SNOT-20 instrument when divided into these four domains.

Our study also confirms the construct validity of the SNOT-20 domains by showing that seven of eight patient domain scores improved over time in both the saline spray and the saline irrigation group, as one would expect with this relatively proven intervention. Moreover, the division of the SNOT-20 into four domains for the analysis of this saline intervention did yield more clinically meaningful data by

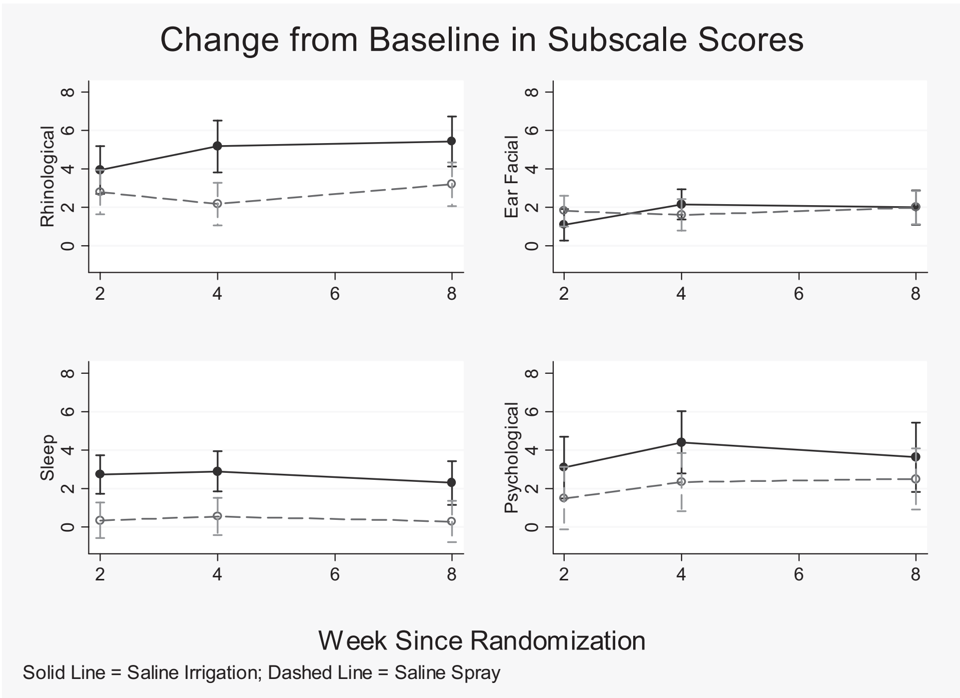

Figure 2. *Change from baseline in subscale score at 2, 4, and 8 weeks after intervention in the saline spray group and in the saline irrigation group.*

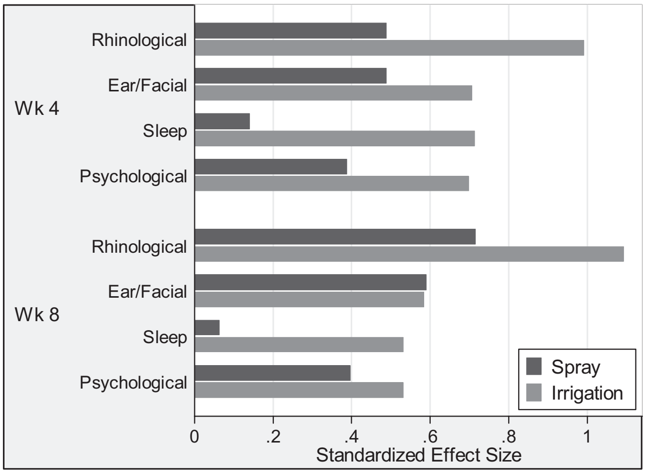

Figure 3. *Standardized effect size of each subscale at weeks 4 and 8 in the saline spray group and in the saline irrigation group.*

showing that the quality-of-life improvements of saline irrigations, compared with saline sprays, is primarily related to improvements in rhinologic symptoms and sleep symptoms rather than ear/facial symptoms or psychological domain improvements. Although this finding was not one of our *a priori* hypotheses, it is consistent with our clinical experience and observations.

These data indicate that the four domains of the SNOT-20 meet other psychometric properties of a valid questionnaire: responsiveness to change (as noted previously), and the ability to detect a clinically significant change. The patients undergoing nasal polypectomy in the study by Browne *et al.*[5] and the patients undergoing saline irrigations in our study all had standard effect sizes (change in score divided by standard deviation) of nearly 0.5 or greater, which is widely accepted as a significant change. Both studies showed the greatest standard effect sizes in the rhinologic symptoms, more so than the other domains. It was interesting to note that some of the greatest differences between the saline irrigation intervention and the saline spray group were noted in the changes in the sleep domain scores between the groups. This was not necessarily an expected finding, but it is a useful contribution to the literature on the benefits of saline irrigations from the perspective of both clinicians and patients.

An interesting observation in this study is the magnitude of differences in the SNOT-20 domain data reported in the saline irrigation group compared with patients who underwent surgical intervention for either nasal polyps or chronic rhinosinusitis in the United Kingdom. Subjects with chronic sinonasal symptoms from a general population who underwent 8 weeks of saline irrigations reported SNOT-20 domain scores that improved to a similar degree of magnitude as patients undergoing polypectomy and had standard effect scores on the rhinologic domain of twice those of patients with chronic rhinosinusitis who underwent surgery. A substantial number of population and methodological differences between the studies makes it difficult to draw definite conclusions, except to say that dividing the SNOT-20 instrument into domains appears to be valid in the assessment of both medical and surgical interventions for rhinosinusitis symptoms.

One limitation of this study was the lack of clinical diagnostic assignment for each of the subjects in our data set, so it is difficult to know what percentage of the subjects suffered from allergic rhinitis, inflammatory rhinosinusitis, infectious rhinosinusitis, or some other condition. Nevertheless, our entry criteria were standardized enough to include only patients with chronic nasal and/or sinus symptoms and excluded those who had had recent surgery or acute respiratory infections. Inclusion of a general population of symptomatic subjects, rather than a subspecialty population, aided in the generalizability of the intervention study, and, likewise, would seem to better corroborate the generalizable validity (with respect to different patient populations) of the Browne *et al.*[5] data, which were derived from a surgical population.

CONCLUSION

It appears that dividing the SNOT-20 into four subscales for rhinologic, ear/facial, sleep function, and psychological function domains is methodologically sound and clinically meaningful. Subscales were useful in identifying the different impacts of saline irrigation versus saline spray on sinonasal related quality of life. Reporting domain scores would improve the precision of the instrument and might allow clinicians to discriminate between the treatment effects in clinical studies that might otherwise be difficult to ascertain if the SNOT-20 were reported as a single score. Improved understanding based on the domain scores would also allow clinicians to more effectively counsel with patients about the quality-of-life impact of interventions for rhinosinusitis.

REFERENCES

1. Hessler JL, Piccirillo JF, Fang D, et al. Clinical outcomes of chronic rhinosinusitis in response to medical therapy: Results of a prospective study. Am J Rhinol 21:10–18, 2007.
2. Robinson S, Douglas R, and Wormald PJ. The relationship between atopy and chronic rhinosinusitis. Am J Rhinol 20: 625–628, 2006.
3. Piccirillo JF, Merritt MG Jr, and Richards ML. Psychometric and clinimetric validity of the 20-Item Sino-Nasal Outcome Test (SNOT-20). Otolaryngol Head Neck Surg 126:41–47, 2002.
4. van Oene CM, van Reij EJ, Sprangers MA, et al. Quality-assessment of disease-specific quality of life questionnaires for rhinitis and rhinosinusitis: A systematic review. Allergy 62:1359–1371, 2007.
5. Browne JP, Hopkins C, Slack R, et al. The Sino-Nasal Outcome Test (SNOT): Can we make it more clinically meaningful? Otolaryngol Head Neck Surg 136:736–741, 2007.
6. Pynnonen MA, Mukerji S, Adams M, et al. Nasal saline sprays versus saline irrigations for chronic sinonasal symptoms: A randomized controlled trial. Arch Otolaryngol 133: 1115–1120, 2007.
7. Harman HH. Maximum-liklihood method. In Modern Factor Analysis, 3rd ed. Chap. 10. Chicago: University of Chicago Press, 1976, 487.
8. Rencher AC. Factor analysis. In Multivariate Statistical Inference and Applications. Chap. 10. New York: Wiley, 1998, 559.
9. Streiner DL, and Norman GR. Reliability. In Health Measurement Scales: A Practical Guide to Their Development and Use, 2nd ed. Oxford, New York: Oxford University Press, 1995, viii, 231. □