

## Cluster analysis and prediction of treatment outcomes for chronic rhinosinusitis

**Zachary M. Soler, MD, MSc,<sup>a</sup> J. Madison Hyer, MS,<sup>b</sup> Luke Rudmik, MD, MSc,<sup>c</sup> Viswanathan Ramakrishnan, PhD,<sup>b</sup> Timothy L. Smith, MD, MPH,<sup>d</sup> and Rodney J. Schlosser, MD<sup>a</sup>** *Charleston, SC, Calgary, Alberta, Canada, and Portland, Ore*

**Background:** Current clinical classifications of chronic rhinosinusitis (CRS) have weak prognostic utility regarding treatment outcomes. Simplified discriminant analysis based on unsupervised clustering has identified novel phenotypic subgroups of CRS, but prognostic utility is unknown.

**Objective:** We sought to determine whether discriminant analysis allows prognostication in patients choosing surgery versus continued medical management.

**Methods:** A multi-institutional prospective study of patients with CRS in whom initial medical therapy failed who then self-selected continued medical management or surgical treatment was used to separate patients into 5 clusters based on a previously described discriminant analysis using total Sino-Nasal Outcome Test-22 (SNOT-22) score, age, and missed productivity. Patients completed the SNOT-22 at baseline and for 18 months of follow-up. Baseline demographic and objective measures included olfactory testing, computed tomography, and endoscopy scoring. SNOT-22 outcomes for surgical versus continued medical treatment were compared across clusters.

**Results:** Data were available on 690 patients. Baseline differences in demographics, comorbidities, objective disease measures, and patient-reported outcomes were similar to previous clustering reports. Three of 5 clusters identified by means of discriminant analysis had improved SNOT-22

outcomes with surgical intervention when compared with continued medical management (surgery was a mean of 21.2 points better across these 3 clusters at 6 months,  $P < .05$ ). These differences were sustained at 18 months of follow-up. Two of 5 clusters had similar outcomes when comparing surgery with continued medical management.

**Conclusion:** A simplified discriminant analysis based on 3 common clinical variables is able to cluster patients and provide prognostic information regarding surgical treatment versus continued medical management in patients with CRS. (J Allergy Clin Immunol 2016;■■■:■■■-■■■.)

**Key words:** Chronic rhinosinusitis, sinusitis, cluster, quality of life, treatment, prediction, outcomes

Chronic rhinosinusitis (CRS) is a common chronic disease characterized by ongoing inflammation of the sinonasal mucosa. It significantly affects individual quality of life (QOL) and personal productivity, with substantial direct and indirect costs to society. The typical treatment paradigm for CRS involves comprehensive medical therapy to reduce inflammation, eliminate any pathogenic bacteria, and physically wash away mucus. Although many patients respond adequately to initial medical therapy, those with persistent disease despite medical therapy can be offered sinus surgery. Previous reports have suggested that patients in whom medical therapy fails and who undergo sinus surgery have better outcomes compared with those who continue only with medical therapy.<sup>1-3</sup> However, these studies typically report average outcomes for an entire cohort, and it remains possible that certain subsets of patients might derive more benefit from a particular treatment strategy than others. Understanding which patients are most likely to improve after surgery and which are better served with continued medical management would inform patient-, clinician-, and policy-level decision making.

Diagnostic criteria for CRS require persistence of cardinal sinonasal symptoms for a minimum of 12 weeks and objective evidence of mucosal inflammation.<sup>4</sup> Although all patients fulfilling these criteria would be considered to have CRS, it is widely believed that subsets of disease exist and might contribute to inconsistent responses to treatment and variable long-term clinical outcomes. Recently, we reported a study wherein unsupervised hierarchic clustering was used to separate a cohort of patients with CRS into clusters using baseline disease characteristics.<sup>5</sup> Clusters differed across a number of features, with patient-reported outcome measures driving much of the separation among clusters. Although medication use was found to differ among clusters, the study was cross-sectional in nature and thus was not designed to determine whether treatment outcomes differ among clusters. The goal of this follow-up study was to determine whether this clustering algorithm, when applied to a larger prospective cohort, would identify differences in outcomes between

From the Departments of <sup>a</sup>Otolaryngology–Head and Neck Surgery and <sup>b</sup>Public Health Sciences, Medical University of South Carolina, Charleston; <sup>c</sup>the Division of Otolaryngology–Head and Neck Surgery, Department of Surgery, University of Calgary; and <sup>d</sup>the Department of Otolaryngology–Head and Neck Surgery, Oregon Health Sciences University, Portland.

Z.M.S. and T.L.S. were supported for this investigation by a grant from the National Institute on Deafness and Other Communication Disorders (NIDCD), one of the National Institutes of Health, Bethesda, Maryland (R01 DC005805; PI/PPD: T.L.S.). Public clinical trial registration (<http://www.clinicaltrials.gov>) was #NCT01332136. Z.M.S. was also supported for this investigation by another grant from the NIDCD (R03 DC013651-01). T.L.S. is a consultant for IntersectENT, which is not associated with this manuscript. Z.M.S. is a consultant for Olympus, which is not affiliated with this manuscript. R.J.S. is supported by grants from OptiNose and IntersectENT, neither of which are associated with this manuscript. R.J.S. is also a consultant for Olympus and Arrinex, which are not affiliated with this study.

Disclosure of potential conflict of interest: Z. M. Soler has received research support from the National Institutes of Health (NIH) and has received consultancy fees from Olympus. T. L. Smith has received research support from the NIH/National Institute on Deafness and Other Communication Disorders (NIDCD) and has received consultancy fees and payment for development of educational presentations from Intersect ENT. R. J. Schlosser has received consultancy fees from Olympus and Meda and has received research support from Intersect, Entellus, and Optinose. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication June 23, 2015; revised October 19, 2015; accepted for publication November 17, 2015.

Corresponding author: Rodney J. Schlosser, MD, Department of Otolaryngology–Head and Neck Surgery, Medical University of South Carolina, 135 Rutledge Ave, MSC550, Charleston, SC 29425. E-mail: [schlossr@musc.edu](mailto:schlossr@musc.edu).

0091-6749/\$36.00

© 2015 American Academy of Allergy, Asthma &amp; Immunology

<http://dx.doi.org/10.1016/j.jaci.2015.11.019>

**Abbreviations used**

B-SIT:	Brief Smell Identification Test
CRS:	Chronic rhinosinusitis
CT:	Computed tomography
ESS:	Endoscopic sinus surgery
MCID:	Minimal clinically important difference
OR:	Odds ratio
QOL:	Quality of life
RSDI:	Rhinosinusitis Disability Index
SNOT-22:	Sino-Nasal Outcomes Test 22

those treated surgically and those who continue with only medical therapy.

## METHODS

### Study cohort

Data were derived from a multi-institutional prospective cohort study. Study participants were recruited from 4 tertiary medical centers across North America (Medical University of South Carolina, Stanford University, University of Calgary, and Oregon Health and Sciences University). Each participant had CRS, as defined by consensus criteria, with at least 12 weeks of cardinal symptoms and objective evidence of sinonasal inflammation on endoscopy or imaging. Patients were enrolled at the point in time at which appropriate medical management was considered to have failed and were offered endoscopic sinus surgery (ESS) as a treatment option. Failure of medical management required ongoing cardinal symptoms despite 2 or more weeks of broad-spectrum or culture-directed antibiotics, 5 or more days of oral steroids, and 1 or more month of topical steroids, with most patients exceeding minimums of prior treatment before surgery was considered. As per real-world clinical practice, patients were offered surgery but allowed to choose either to continue solely with medical management or to have ESS.

### Baseline assessments

All subjects provided written informed consent to participate in this observational study (Clinical Trial #NCT00799097; National Institutes of Health R01 DC005805; Pro 12409). Study coordinators administered baseline questionnaires that assessed demographic information (age, sex, and race) and medical comorbidities. Patients were considered to have medical comorbidities if they had been given a diagnosis by a physician in the past (asthma, aspirin sensitivity, chronic obstructive pulmonary disease, obstructive sleep apnea, fibromyalgia, and depression) or based on patient self-report (current smoking, alcohol intake, and prior sinus surgery). Patients were considered to have allergic rhinitis based on prior positive skin prick test responses or allergen-specific IgE antibody test results. Allergic fungal rhinosinusitis was defined by using classic Bent-Kuhn criteria.<sup>6</sup> At baseline, objective olfactory function was evaluated by using the Brief Smell Identification Test (B-SIT). Sinonasal endoscopy was performed by the treating physician and scored by using the Lund-Kennedy scale, with the presence or absence of polyps recorded.

All patients had computed tomographic (CT) scans performed after maximal medical therapy and before enrollment, and these were scored by using the Lund-Mackay system.<sup>4</sup> Physicians scoring sinonasal endoscopy and CT scan results were blind to all other patient-reported variables for the study's duration.

### Interventions

Patients chose to either continue solely with medical therapy or undergo ESS at baseline enrollment. Because this was an observational study, specifics of ongoing medical therapy for either group were not dictated by study protocol but occurred in a real-world fashion and were aimed at maximally controlling disease in line with patient preferences. For those undergoing

surgery, the specific procedure performed was left to the discretion of the treating surgeon based on CT findings and the overall clinical picture, as occurs in everyday clinical practice. All surgery was done by fellowship-trained rhinologists in accordance with commonly accepted principles of functional ESS.<sup>4</sup>

### Outcome assessments

Sinus-specific QOL was the primary outcome and was assessed first at baseline and then again at 6, 12, and 18 months after study enrollment. Sinus-specific QOL was assessed by using 2 different instruments: the 22-item Sino-Nasal Outcome Test-22 (SNOT-22) and Rhinosinusitis Disability Index (RSDI). SNOT-22 is a validated CRS-specific QOL instrument containing 22 questions (total score range, 0-110), with higher scores representing more severe effect. Total scores and individual domains were evaluated, including rhinologic, extranasal-rhinologic, psychiatric, ear/facial, and sleep domains, as has been described previously.<sup>7</sup> The RSDI is a validated 30-question survey comprised of 3 individual subscales to measure the effect of sinus disease on the physical, functional, and emotional domains (range, 0-120), with higher RSDI total and subscale scores representing a greater effect of disease. These instruments were chosen because they quantify the effect of each symptom specific to CRS, namely nasal congestion, nasal drainage, facial pain, and olfactory disturbance, as well as "extrarhinologic" manifestations of disease. These instruments are used worldwide and have been shown to discriminate between patients with CRS and healthy control subjects, as well as to identify significant differences after medical and surgical treatments.<sup>8</sup> Sleep quality was assessed by using the Pittsburgh Sleep Quality Index.<sup>9</sup> Depressed mood and anhedonia were assessed by using the Patient Health Questionnaire 2.<sup>10</sup> Patients were also asked how many days in the last 90 days they missed work.

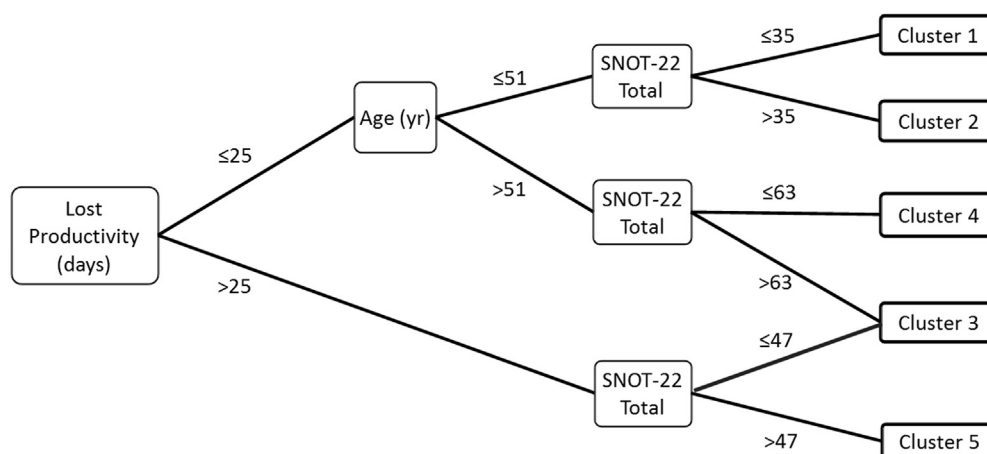
### Cluster determination

The input variables and hierarchic clustering methods used in the original cohort of 382 patients to derive clusters have been previously reported in detail.<sup>5</sup> Briefly, 103 variables encompassing demographics, comorbidity, objective CRS metrics, and patient-reported outcome measures were reduced to meaningful factors with a high degree of correlation. This permitted us to reduce 103 variables to 32. The Ward minimum-variance hierarchic method was then used to perform the cluster analysis. This analysis places subjects into groups or clusters suggested by the data and not defined *a priori*, and then simple discriminant analysis is performed to obtain simple algorithms for clinical use and future studies.

Cluster analysis was not performed again in this study. Rather, we used a discriminant analysis and simple clinical algorithm based on this clustering technique to classify patients into the 5 statistical clusters with a high likelihood of success. The clinical algorithm was used to place each patient from the current expanded cohort into one of 5 clusters based on age, SNOT-22 score, and lost productivity over 90 days (clusters 1-5, Fig 1).

### Statistical analysis

Descriptive statistics are presented for all demographic, comorbidity/exposure, and CRS severity measures. Differences between clusters at baseline on entry into the study were assessed by using  $\chi^2$  tests and ANOVA for categorical and continuous measures, respectively. *Post hoc t* tests were performed within each cluster based on linear mixed-effects models, which controlled for site and baseline QOL, to assess differences between medical and surgical treatments. Normality was examined in each analysis. When normality was questionable, appropriate transformations were considered. Logistic regression was then used to assess the odds of achieving a minimal clinically important difference (MCID) for each QOL measure after either medical or surgical treatments, controlling for baseline value and follow-up duration. An MCID was defined as an improvement from baseline to last follow-up of 9 or greater for SNOT-22 or 10.3 or greater for RSDI, as previously defined.<sup>8</sup> Statistical significance was assessed at an  $\alpha$  value of .05. For the secondary objective on likelihood of achieving an MCID, statistical significance was adjusted by using Bonferroni correction for multiple comparisons for simultaneously testing



**FIG 1.** Clinical algorithm used to define clusters. The above algorithm was used to classify patients into the 5 statistical clusters by using simple clinical measures. *Lost Productivity* is the number of work days missed in the last 90 days. Printed with permission from Medical University of South Carolina Rhinology.

10 hypotheses (5 clusters and 2 outcomes per cluster). All analyses were performed with SAS 9.4 software (SAS Institute, Cary, NC).

## RESULTS

### Study cohort

The study cohort included 690 patients ranging in age from 18 to 86 years. A total of 376 patients overlapped with the original cross-sectional cohort, and 314 were newly enrolled. Baseline differences in demographics, comorbidities, objective disease measures, and patient-reported disease severity are shown in [Table I](#). The patients were equally split between the sexes, and the average age was 58.0 years (SD, 15.9 years). Overall, 42% (289/690) of patients had undergone prior sinus surgery, and comorbidities, such as allergy (25% [172/690]), asthma (37% [255/690]), and nasal polyps (37% [254/690]), were common. As expected and previously reported,<sup>5,11</sup> differences existed between clusters for demographics, comorbidities/exposures, CRS severity measures, and patient-reported outcome measures similar to what was previously reported for the smaller cross-sectional cohort. Given the larger sample size of the current cohort, the power to detect differences increased, and additional baseline variables showed a significant difference, most notably olfactory function (B-SIT).

### Cluster 1

Cluster 1 was comprised of 87 patients, of whom 56 underwent surgery and 31 continued medical management. This group has the highest prevalence of male subjects (69%), was the oldest (65 years), and tended to have moderate CRS severity measures. At 6 months, the surgical group had significantly improved QOL based on both SNOT-22 and RSDI total scores compared with the medical group ([Fig 2](#)). However, this difference was no longer detectable at 12 months, and by 18 months, scores on both instruments were equivalent between treatment groups. There were no long-term differences in SNOT-22 domain scores or RSDI subscale scores between treatment groups ([Figs 3 and 4](#)). The odds of achieving an MCID were not significantly different between groups for either the SNOT-22 (odds ratio [OR], 2.0; 99.5% CI, 0.6-7.0) or RSDI (OR, 3.0; 99.5% CI, 0.7-13.4) instruments ([Table II](#)).

### Cluster 2

Cluster 2 (n = 233) was the largest, with 182 patients undergoing surgery and 51 continuing medical management. This group was split between the sexes and tended to be older (62 years), with some of the most severe CRS severity measures. The surgical group had significantly better SNOT-22 scores at the 6-, 12-, and 18-month time points, with similar robust differences for rhinologic and extranasal-rhinologic SNOT-22 domains. The RSDI total score was significantly better in the surgical group at 6 months but not at other time points. However, the RSDI physical score was significantly better at both 6 and 18 months in the surgical group compared with the medical group. The odds of achieving an MCID were higher in the surgical group for both the SNOT-22 (OR, 4.7; 99.5% CI, 1.5-15.4) and RSDI (OR, 5.5; 99.5% CI, 1.2-19.8) instruments.

### Cluster 3

Cluster 3 was the second largest cluster, with 215 total patients, of whom 161 elected surgery and 54 continued with solely medical management. This group was split between the sexes, was the youngest (37 years), and had the least severe CRS severity measures. The surgical group had significantly better QOL at the 6-, 12-, and 18-month time points compared with the medical group based on both SNOT-22 and RSDI total scores. Significant differences were also seen at every time point for the rhinologic, extranasal-rhinologic, and ear/facial domains of the SNOT-22 instrument, as well as for the RSDI physical, functional, and emotional subscores. Those patients in cluster 3 had the highest odds of achieving an MCID with surgery compared with medical treatment for both the SNOT-22 (OR, 16.9; 99.5% CI, 3.7-77.5) and RSDI (OR, 7.8; 99.5% CI, 2.2-28.6) instruments.

### Cluster 4

A total of 94 patients were classified into cluster 4, with 76 choosing to undergo surgery and 18 continuing solely with medical management. This group had two thirds female subjects, was relatively young (38 years), and had the worse CT and endoscopic measures. Improvements in QOL were seen in both patients undergoing surgery and in those continuing with medical

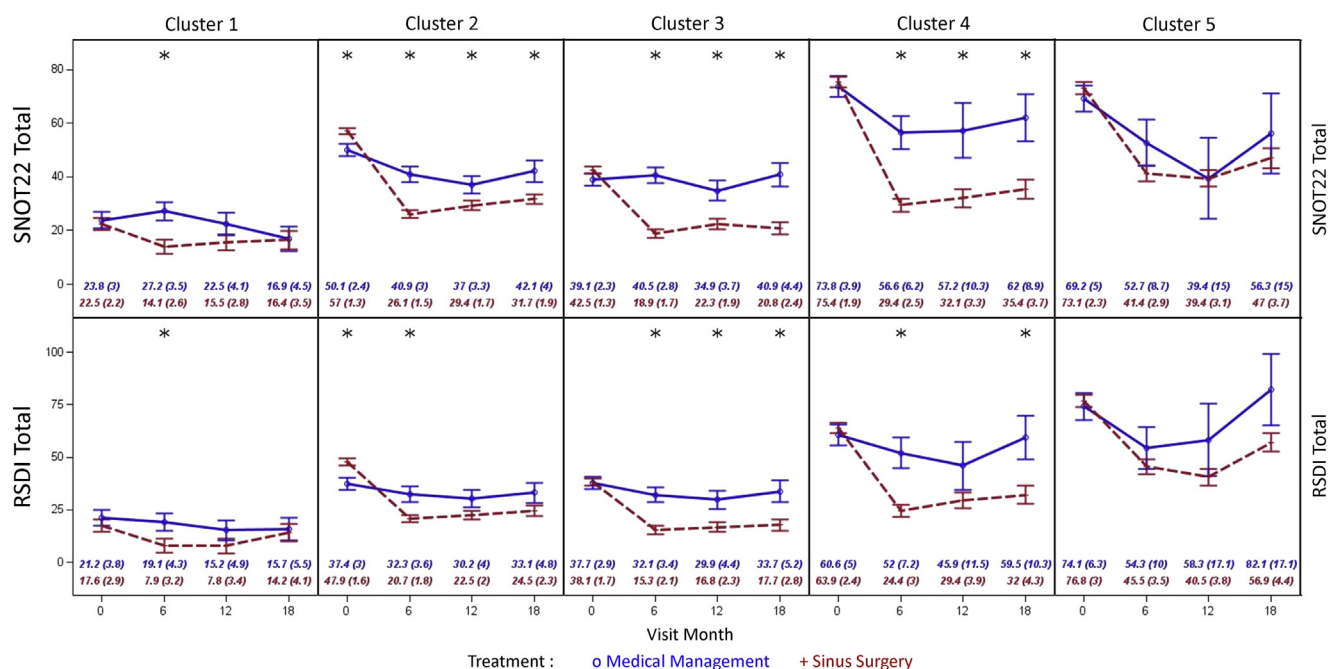
**TABLE I.** Baseline differences between clusters for the study cohort

Variable	Total (n = 690)	Cluster 1 (n = 87)	Cluster 2 (n = 233)	Cluster 3 (n = 215)	Cluster 4 (n = 94)	Cluster 5 (n = 61)	P value
Demographics							
Sex, no. (%)							<.001
Female	356 (51.6)	27 (31)	111 (47.6)	112 (52.1)	62 (66)	44 (72.1)	
Male	334 (48.4)	60 (69)	122 (52.4)	103 (47.9)	32 (34)	17 (27.9)	
Race, no. (%)							.019
African American	40 (5.8)	5 (5.7)	8 (3.4)	13 (6.1)	10 (10.6)	4 (6.6)	
White	586 (85.1)	78 (89.7)	207 (88.8)	171 (79.9)	75 (79.8)	55 (90.2)	
Other	63 (9.1)	4 (4.6)	18 (7.7)	30 (14)	9 (9.6)	2 (3.3)	
Age (y), mean (SD)	50.61 (15.05)	65.46 (7.8)	62.13 (7.5)	37.38 (9.72)	38.31 (8.73)	50.82 (11.89)	<.001
Comorbidities/exposures							
AFRS, no. (%)	19 (2.8)	1 (1.1)	4 (1.7)	8 (3.7)	5 (5.3)	1 (1.6)	.276
Allergic rhinitis, no. (%)	172 (24.9)	26 (29.9)	68 (29.2)	40 (18.6)	18 (19.1)	20 (32.8)	.02
Prior sinus surgery, no. (%)	289 (41.9)	32 (36.8)	114 (48.9)	75 (34.9)	41 (43.6)	27 (44.3)	.036
Asthma, no. (%)	255 (37)	24 (27.6)	90 (38.6)	72 (33.5)	44 (46.8)	25 (41)	.061
Polyps, no. (%)	254 (36.8%)	30 (34.5%)	94 (40.3%)	72 (33.5%)	42 (44.7%)	16 (26.2%)	.095
ASA intolerance, no. (%)	57 (8.3)	6 (6.9)	20 (8.6)	13 (6)	13 (13.8)	5 (8.2)	.241
COPD, no. (%)	31 (4.5)	7 (8)	15 (6.4)	4 (1.9)	1 (1.1)	4 (6.6)	.024
Depression, no. (%)	94 (13.6)	4 (4.6)	30 (12.9)	25 (11.6)	16 (17)	19 (31.1)	<.001
Fibromyalgia, no. (%)	25 (3.6)	1 (1.1)	11 (4.7)	4 (1.9)	2 (2.1)	7 (11.5)	.004
Oral steroid dependence, no. (%)	57 (8.3)	6 (6.9)	23 (9.9)	7 (3.3)	11 (11.7)	10 (16.4)	.005
Smoker (packs/d), mean (SD)	0.03 (0.17)	0 (0)	0.04 (0.18)	0.01 (0.09)	0.06 (0.25)	0.06 (0.28)	.048
Alcohol (g/wk), mean (SD)	25.01 (52.24)	46.36 (63.33)	31.87 (68.9)	13.82 (21.53)	21.15 (36.67)	13.7 (45.63)	<.001
Objective CRS severity measures, mean (SD)							
Endoscopy total	6.02 (3.89)	5.84 (3.92)	6.49 (3.97)	5.35 (3.78)	6.4 (4.03)	6.21 (3.47)	.027
CT total	11.81 (6.3)	11.8 (6.23)	12.65 (6.23)	10.41 (6.15)	12.79 (6.78)	12.02 (5.72)	.002
B-SIT total	8.76 (3.19)	8.87 (3.01)	7.99 (3.46)	9.49 (2.79)	8.9 (3.31)	8.58 (3)	<.001
Patient-reported outcome measures, mean (SD)							
SNOT-22 total	51.7 (20.44)	23.49 (9.35)	56.04 (14.51)	41.8 (13.44)	75.39 (9.3)	73.23 (12.62)	<.001
SNOT-22 (rhinologic)	16.15 (6.32)	9.45 (5.97)	17.54 (5.09)	14.03 (5.38)	21.64 (4.61)	19.28 (5.09)	<.001
SNOT-22 (extranasal-rhinologic)	8.17 (3.6)	5.13 (3.49)	9.09 (3.13)	6.79 (3.25)	10.54 (2.8)	10.15 (2.81)	<.001
SNOT-22 (ear/facial)	9.06 (5.32)	3.43 (3.16)	9.19 (4.71)	7.35 (3.79)	14.97 (3.69)	13.34 (4.78)	<.001
SNOT-22 (psychologic)	15.33 (8.44)	4.8 (4.1)	16.77 (6.73)	11.92 (6.66)	23.23 (5.24)	24.48 (5.13)	<.001
SNOT-22 (sleep)	13.49 (6.86)	5.24 (4.71)	15.02 (5.32)	10.57 (5.86)	19.87 (3.67)	19.79 (3.81)	<.001
RSDI total	45.59 (24.78)	19.76 (14.39)	46.52 (22.69)	38.32 (19.51)	63.54 (16.87)	76.97 (17.7)	<.001
RSDI (physical)	18.74 (9.35)	8.26 (6.01)	19.48 (8.12)	15.88 (7.42)	26.47 (6.61)	29.02 (6.49)	<.001
RSDI (emotional)	12.29 (8.95)	5.37 (5.32)	12.11 (8.57)	10.2 (7.73)	17.47 (7.57)	22.3 (8.11)	<.001
RSDI (functional)	14.57 (8.83)	6.13 (5.39)	14.94 (8.26)	12.24 (7.42)	19.61 (6.33)	25.65 (6.64)	<.001
PSQI total	9.42 (4.48)	5.97 (3.43)	9.83 (4)	7.87 (3.69)	11.91 (4.34)	14.54 (3.34)	<.001
SF-12 total	31.62 (2.58)	31.26 (2.31)	31.3 (2.5)	32.1 (2.55)	32.05 (2.98)	30.92 (2.37)	<.001
SF-12 (factor 1)	16.71 (1.96)	15.46 (1.83)	16.39 (1.81)	16.84 (1.91)	17.71 (2.09)	17.64 (1.35)	<.001
SF-12 (factor 2)	14.91 (2.15)	15.8 (1.71)	14.91 (2.15)	15.27 (1.99)	14.34 (2.24)	13.28 (2.14)	<.001
PHQ total	1.58 (1.63)	0.66 (1.15)	1.5 (1.54)	1.28 (1.36)	2.03 (1.56)	3.57 (1.76)	<.001
Productivity loss	9.27 (19.64)	1.26 (3.32)	3.33 (5.45)	5.52 (11.85)	4.8 (5.72)	60.93 (24.06)	<.001
Medication use in last 90 d, mean (SD)							
Oral antibiotics	17.04 (22.15)	15.74 (23.62)	17.78 (23.22)	12.99 (17.42)	16.67 (19.43)	30.88 (28.7)	<.001
Oral antihistamines	23.27 (35.25)	14.77 (30.1)	27.04 (36.45)	18.14 (32.68)	29.32 (37.06)	29.67 (39.3)	.002
Oral decongestants	19.06 (30.84)	9.41 (24.5)	17.98 (28.9)	15.83 (28.28)	25.67 (33.7)	38.15 (40.01)	<.001
Leukotriene antagonists	14.57 (31.12)	16.36 (33.34)	16.25 (33.08)	8.65 (23.62)	21.72 (36.24)	15.17 (32.23)	.009
Oral steroids	11.89 (19.71)	13.26 (23.74)	11.85 (19.17)	9.58 (16.77)	10.89 (16.16)	19.75 (27.06)	.011
Saline irrigations	43.47 (37.94)	36.17 (37.92)	50.21 (37.95)	41.96 (37.75)	34.84 (36.67)	47.38 (36.63)	.003
Steroid irrigations/drops	13.6 (28.98)	13.42 (29.43)	16.09 (31.5)	11.79 (26.6)	9.16 (23.5)	17.56 (33.3)	.213
Steroid nasal spray	39.55 (38.39)	47.04 (40.65)	42.04 (39.05)	34.01 (36.73)	37.53 (37.43)	41.95 (38.07)	.06

Variables used to define clusters are presented as either count data (percentages) or means  $\pm$  SDs.

AFRS, Allergic fungal rhinosinusitis (as defined by Bent-Kuhn criteria); ASA, acetylsalicylic acid intolerance (ie, aspirin-exacerbated respiratory disease); B-SIT, Brief Smell Identification Test (range, 0-12); COPD, chronic obstructive pulmonary disease; CT, computed tomography (range, 0-24); Endoscopy, Lund-Kennedy scoring (range, 0-20); OSA, obstructive sleep apnea; PHQ, Patient Health Questionnaire 2 (range, 0-6); Productivity loss, missed days of work of the last 90 days (range, 0-90); PSQI, Pittsburgh Sleep Quality Index (range, 0-21); RSDI, Rhinosinusitis Disability Index (range, 0-120); RSDI [physical] range = 0-44, RSDI [emotional] range = 0-40, RSDI [functional] range = 0-36; SF-12, 12-item Medical Outcomes Study Short Form Health Survey (range, 0-100); Oral steroid dependency, requiring oral steroids daily; SNOT-22, 22-Item Sino-Nasal Outcome Test (range, 0-110); SNOT-22 [rhinologic] range = 0-30, SNOT-22 [extranasal-rhinologic] range = 0-15, SNOT-22 [ear/facial] range = 0-25, SNOT-22 [psychologic] range = 0-35; SNOT-22 [sleep] range = 0-25.





**FIG 2.** SNOT-22 and RSDI overall outcomes between clusters treated with surgery or continuing medication only. \* $P < .05$ .

management. However, those undergoing surgery had significantly better total SNOT-22 scores at 6, 12, and 18 months, as well as significantly better total RSDI scores at 6 and 18 months. SNOT-22 differences were driven mainly by improvements in rhinologic and ear/facial domains in the surgical group. RSDI differences were most notable in the physical subscore, which was better in the surgical group at every follow-up time point. Although the surgical group had a greater absolute improvement, the odds of at least achieving an MCID did not reach statistical significance for either the SNOT-22 (OR, 6.2; 99.5% CI, 0.3-152.7) or RSDI (OR, 4.0; 99.5% CI, 0.4-36.0) instruments.

### Cluster 5

Just less than 9% of the cohort was classified into cluster 5 ( $n = 61$ ), with 50 undergoing surgery and 11 continuing with only medical treatments. This group had the highest prevalence of female subjects (72%), tended to be of average age (50 years), and had slightly worse than average CRS severity measures. Both medical and surgical groups appeared to have initial improvement at 6 months, with progressive worsening at 12 and 18 months. There were no differences between the surgical and medical groups for the SNOT-22 total score or any individual domain. Similarly, RSDI total scores were not different between the treatment groups, although physical subscores were better in the surgical group at 6 and 18 months. The odds of achieving at least 1 MCID were not different between groups for either the SNOT-22 (OR, 4.9; 99.5% CI, 0.2-119.8) or RSDI (OR, 3.9; 99.5% CI, 0.2-99.4) instruments.

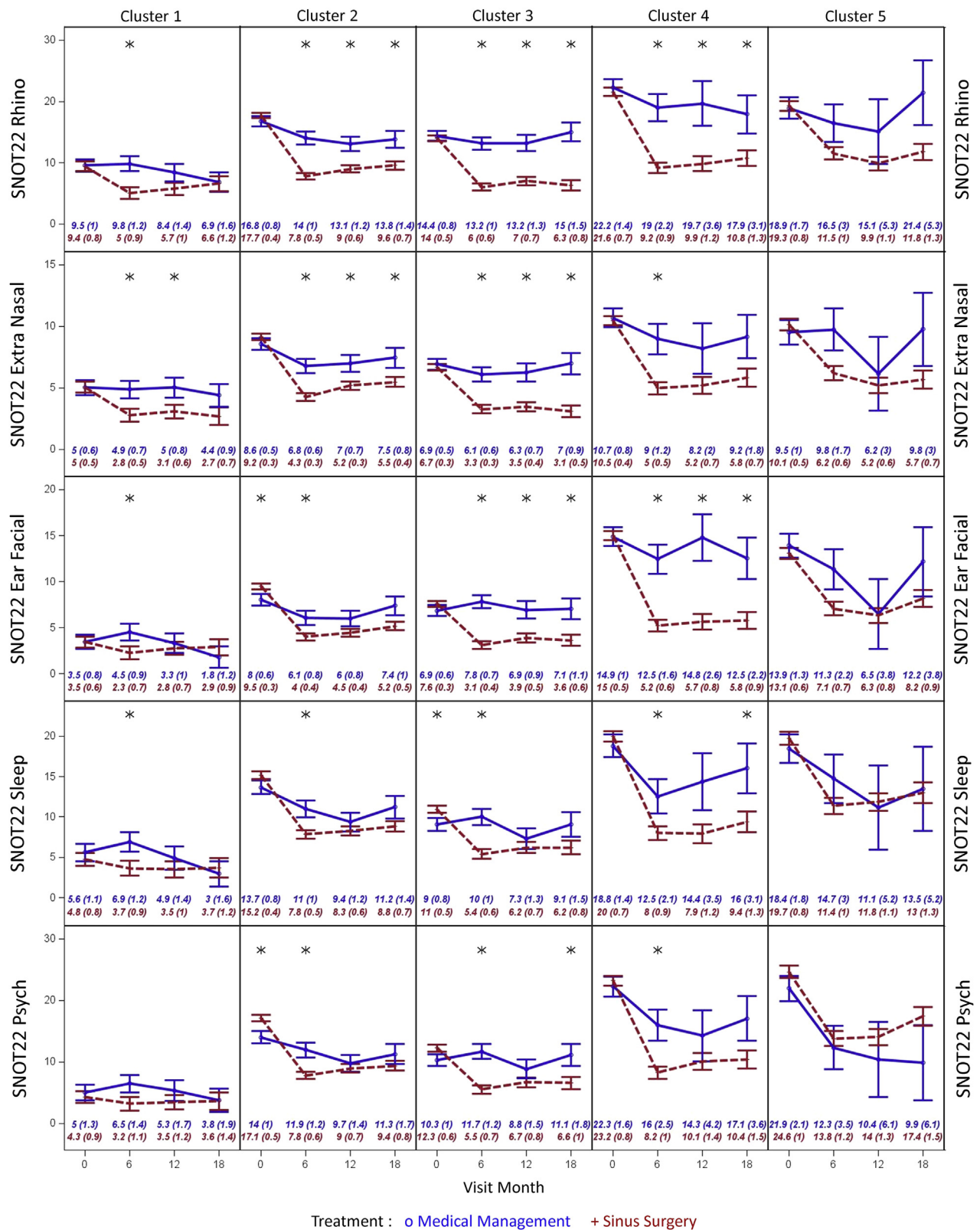
### Follow-up

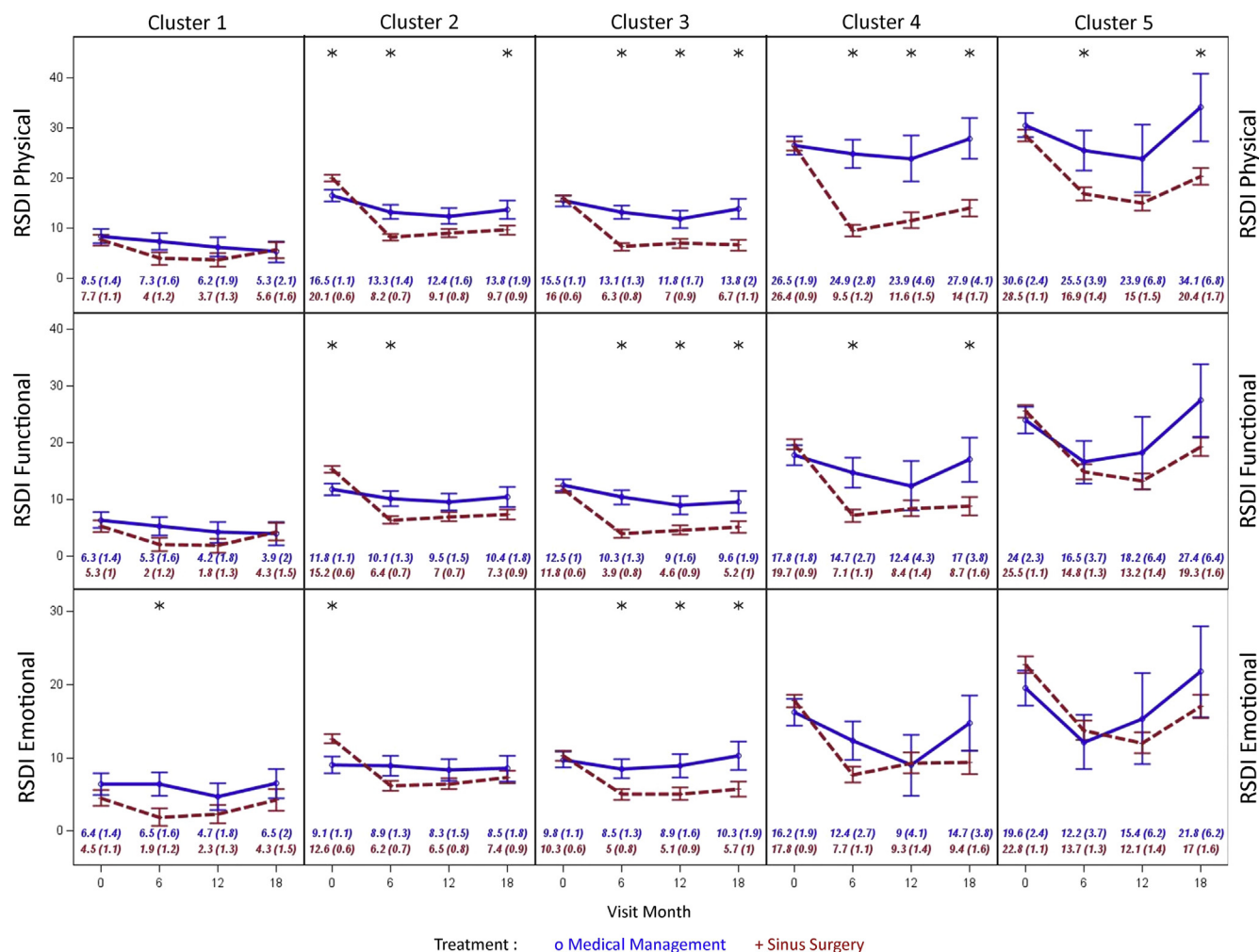
Overall, we had follow-up on 59% of patients at a minimum of 1 time point. Follow-up rate by cluster and time point is shown in

**Table III.** There were no statistically significant differences in follow-up rates between clusters. However, the mixed-effects model we used is designed to account for loss to follow-up. To confirm that loss to follow-up did not affect our findings, we also analyzed total SNOT-22 scores, comparing medical and surgical outcomes at each time point for each cluster only for patients with at least 1 follow-up, and this did not change our findings reported above.

### DISCUSSION

Although sinus surgery results in significant QOL improvements over time, most studies report mean values for an entire cohort. Prior studies demonstrate that approximately 75% to 80% of patients achieve an MCID after ESS,<sup>11</sup> but there are few traditional clinical factors that reliably predict surgical success or magnitude of QOL improvement for an individual patient. In this study a large cohort of patients with CRS was divided into clusters, and outcomes were compared between those undergoing ESS and those continuing solely with medical management. Distinct differences in outcomes were seen between clusters, with clusters 1 and 5 failing to show added benefit with surgery, whereas robust advantages were seen with surgical treatment in clusters 2, 3, and 4, with the greatest odds in cluster 3. These differences in outcomes were confirmed by using 2 different sinus-specific QOL instruments: the SNOT-22 and RSDI. These 2 QOL instruments have previously been shown to have a high degree of correlation,<sup>12</sup> but given some variation in questions and differing constructs, it was possible that outcomes could differ. As demonstrated in **Fig 2**, SNOT-22 and RSDI results correlated with rare exceptions at select time points in clusters 2 and 4, supporting the concept of cluster prognostication for CRS QOL with 2 separate instruments. Interestingly, another benefit noted in this study was that the most symptomatic patients (clusters 4 and 5)





**FIG 4.** RSDI outcomes by subscale score between clusters treated with surgery or continuing medication only. \* $P < .05$ .

**TABLE II.** ORs of achieving improvement of at least 1 MCID after sinus surgery compared with medical management

Cluster	SNOT-22 (99.5% CI)	RSDI (99.5% CI)
1	2.1 (0.6-7.0)	3.0 (0.7-13.4)
2	<b>4.7 (1.5-15.4)</b>	<b>5.5 (1.2-19.8)</b>
3	<b>16.9 (3.7-77.5)</b>	<b>7.8 (2.2-28.6)</b>
4	6.2 (0.3-152.7)	4.0 (0.4-36.0)
5	4.9 (0.2-119.8)	3.9 (0.2-99.4)

Odds ratios are defined as the odds of achieving at least 1 MCID after surgery compared with the odds of achieving 1 MCID with only continued medical therapy. ORs in boldface are statistically significant.

had similar baseline SNOT-22 scores, and traditional classifications might have grouped these patients together. However, discriminant algorithms are able to separate those more likely to benefit from surgery (cluster 4) from those who will probably not receive as much benefit from surgery (cluster 5).

A simple discriminant analysis with a clinical algorithm (Fig 1) based on our prior report<sup>5</sup> was used in this study to place each patient into a cluster at baseline. Use of this clinical algorithm, as opposed to repeating an entirely new statistical clustering, is crucial for the clinical applicability of these findings to future

**TABLE III.** Follow-up rates by cluster for each time point

Cluster	Time point		
	6 mo	12 mo	18 mo
1	65%	54%	36%
2	62%	48%	36%
3	52%	37%	25%
4	47%	24%	21%
5	52%	42%	31%
Overall	56.8%	41.6%	30.5%

patients. In our previous study this clinical algorithm placed patients into the correct cluster 89.4% of the time.<sup>5</sup> Importantly, the clusters generated in this expanded cohort closely mirror those seen in our original study, with respect to overall proportions and qualitative clinical characteristics. The only differences were quantitative; namely, some variables became significant across groups because of the larger sample size and increased power (ie, B-SIT). This provides further validation of our prior discriminant analysis, suggesting that the simple clinical algorithm proposed, with just 3 variables, accurately replicates the original

statistical clustering based on more than 100 variables. It is important to note that in addition to differences in the 3 variables used in the discriminant analysis (total SNOT-22 score, age, and missed productivity), there are many other baseline differences among clusters, as seen in Table I. Although statistically significant, it remains to be determined whether these differences are clinically relevant. These 3 variables are probably best thought of as proxies for a constellation of many other clinical characteristics, which collectively describe an individual cluster and theoretically influence outcomes.

There are several reasons to think these results might prove generalizable across different patient populations. The first reason is that the original clustering methods were generated with a preliminary cohort of patients, whereas the outcome assessment reported herein uses a much larger cohort providing some degree of replication, although it is not fully independent. Additionally, patients were enrolled from 4 centers across North America, with multiple participating surgeons, diverse geography, and varied patient populations. Thus differences are less likely to be driven by a single center, individual surgeon, or specific patient population. The sample size is also large relative to most other prospective studies examining treatment outcomes in patients with CRS.

Ultimately, widespread generalizability will need to be assessed by using additional patient cohorts. As shown in Fig 1, our simple algorithm could be used potentially in clinical practice to aid in individualized decision making. However, at the present time, findings should be limited to the patient population studied, namely a heterogeneous group of patients with CRS who present to tertiary centers who are surgical candidates after appropriate medical therapy fails. In our study this was defined as 2 or more weeks of broad-spectrum or culture-directed antibiotics, 5 or more days of oral steroids, and 1 or more month of topical steroids. These findings should not be extrapolated to patients who present *de novo* with untreated CRS or to those treated with techniques other than traditional ESS. Rather, this analysis is designed to serve as a foundation for future, more widespread, multi-institutional studies.

A limitation of this analysis is that patients self-selected treatment rather than being randomly allocated into groups. The lack of randomization means that unmeasured confounding can never be fully ruled out and could account for some differences seen between groups. Although one can never fully rule out this possibility, the clustering algorithm results in near homogeneity within clusters for all input variables. Thus, within each cluster, there were no notable differences between medical and surgical treatment groups for objective disease severity measures, medical comorbidities, or baseline QOL measures. Additionally, the outcome analysis was controlled for enrollment site and length of follow-up, further reducing the risks of selection and follow-up bias, respectively. Although a blind randomized clinical trial would be ideal, the feasibility of such a trial would be low given patient reluctance to enroll in randomized surgical trials, ethical concerns of sham surgery, and costs involved in long-term assessments. Other study designs would also be possible, including complete standardization of medical therapy to include dosing and selection of antibiotic, oral steroid, and topical steroid; standardization of surgical approach; technique and equipment; and strict inclusion criteria for subtypes of CRS to enroll. Although these designs would be ideal, unfortunately, there is limited evidence to support specific standardization of many of

these approaches, and individual practices vary widely from physician to physician.

When conducting prospective outcomes research, there is always concern regarding patients lost to follow-up and their effect on study findings. The overall follow-up of 59% at any time point in our study compares favorably with other cluster analyses. Prior asthma studies examining therapeutic outcomes after clustering did not explicitly state the follow-up rate, but from extrapolation of the data, it appears to be between 49% and 72% depending on the outcome variable and time point.<sup>13</sup> Other studies containing surgical cohorts range from a low of 48% at 6 months in a single-institution series<sup>14</sup> to a high of 63% at 12 months in a multi-institutional study previously published by our group.<sup>2</sup> Rather than collapsing all follow-up to 1 time point, we wished to examine the long-term durability of our findings, realizing that at 18 months, our follow-up rate would continue to decrease. Despite this, we still found that many of the differences identified at 6 months persisted to 18 months. The mixed-effects modeling used in this study is the statistical approach ideal for longitudinal analysis with incomplete follow-up,<sup>15</sup> but to confirm our findings, we analyzed results using only patients with follow-up at any time point, and our conclusions were identical.

The question remains as to how these findings should affect clinical decision making. All patients in this study were appropriate surgical candidates based on the fact that they had CRS based on accepted diagnostic criteria and remained symptomatic despite appropriate medical therapy. Findings from this study suggest that those patients falling into clusters 2, 3, and 4 have a significantly increased odds of durable improvement with surgery as opposed to continuing solely with medication, whereas those patients fitting into clusters 1 and 5 might not.

Although these data suggest certain patients might not derive additional benefit from surgery, use of these findings to withhold surgery for patients in clusters 1 and 5 seems extreme based on these preliminary findings alone. Rather, this information might prove useful in patient counseling as to possible outcomes. For example, cluster 5 patients improve with surgery, but the magnitude of improvement is similar to those receiving continued medical management. Thus other factors, such as medication side effects, might lead a cluster 5 patient to opt for surgery. Similarly, a cluster 3 patient with significant medical comorbidities or an inability to miss work/school might opt for continued medical management despite greater odds of improvement with surgery. It is also important to point out that sample sizes varied between clusters, with cluster 5 being the smallest, and that clusters with smaller numbers might be more affected by patients lost to follow-up over the 18-month study duration, although there was no difference in follow-up rates across clusters or between treatment interventions within clusters. Thus it remains possible that a type II error exists within this group because of a lack of power. Future studies with larger sample sizes, particularly with regard to medical treatment, would help refine estimates for this particular cluster.

Sinus-specific QOL is an important metric for CRS outcome studies, but it should be remembered that these measures are not the only outcome of interest and might not be the sole driving force behind pursuing surgery in all patients. Some patients could elect surgery to alleviate a specific symptom, such as olfactory loss, which is not a major contributor to QOL, as measured by using the SNOT-22 or RSDI instruments. Other examples include



orbital/skull base erosion related to disease burden or a desire to transition from systemic to topical medication administration. With these considerations in mind, strict adherence to this clustering algorithm to guide surgical decision making should not supersede the specific needs of each patient. Future studies might explore additional outcome measures, such as medication use, personal productivity, and medical costs.

It remains quite likely that inclusion of additional cluster-defining measures would further refine our cluster analysis. Clinicians inherently feel that classifications, such as revision surgery status, the presence of polyps, or a number of other variables, affect the presentation and outcomes of patients with CRS. Unfortunately, traditional classifications and CRS severity metrics, such as Lund-Kennedy endoscopy scores and Lund-Mackay CT scores have not been overly useful in predicting surgical success.<sup>11</sup> We analyzed our data controlling for revision surgery status. Total SNOT-22 results were identical across all clusters and at all time points to data presented in this article. This demonstrates that the presented clustering algorithm adequately accounts for this variability without the need to directly classify patients based on this multitude of traditional clinical variables. Development of more precise clinical metrics, such as volumetric CT analysis or modified endoscopy scales, and refined phenotypic subclassifications might be useful in defining clusters or in developing discriminant algorithms. In addition, although we included more than 100 clinical variables in our original cluster analysis, there might be unknown key clinical variables that are missing or as yet unknown.

Also notably absent from our original study was inclusion of histologic or biomolecular markers. A wide array of possible measures could be included in the future, such as eosinophil counts, cytokine profiles, genotypes, or measures of the local microbiome among others. Inclusion of these measures is likely necessary for clusters to begin to mirror underlying endotypes (ie, subgroups differentiated based on underlying pathophysiologic mechanisms). Certainly, identifying discrete endotypes will be crucial to develop and assess future targeted therapeutics. Therefore the current study can be considered proof of concept that clustering methods can be used to predict outcomes but should serve as a starting rather than an ending point for future research.

As mentioned above, there are numerous implications from our study. Our discriminant analysis relied on 3 variables: total SNOT-22 scores, days of missed productivity, and age. Although age is readily available, SNOT-22 scores and missed productivity might not be collected during routine clinical practice. It is not expected that these 3 variables will immediately be incorporated into all practices but rather serve as a starting point for further development of clustering algorithms and prognostic studies using these and additional clinical variables. Furthermore, although practitioners can begin to use this information to better counsel patients, they must do so within the limitations of our study acknowledged above, namely that these findings might not apply to all patients with CRS. It is hoped that this analysis is replicated by other centers around the world and applied to other

patient populations and that further refinement in clinical and biomarker variables is pursued. Another important area for further study is understanding why certain clusters respond better or worse to a given therapeutic modality (eg, why cluster 3 appears to have better surgical outcomes).

In conclusion, hierarchic clustering of patients with CRS was reproducible, and use of a novel discriminant analysis provides some prognostic utility in determining patients most likely to benefit from surgical therapy. It can serve as an aid in counseling patients with CRS regarding treatment choices and likely outcomes but is not a substitute for physician recommendations that take into account individual patient differences or preferences. Future refinements in clinical metrics and inclusion of biomarkers will likely further improve our prognostic ability and aid us in individualizing treatment algorithms.

**Clinical implications: Cluster analysis and the resulting simple discriminant algorithms might improve therapeutic recommendations for patients with CRS.**

## REFERENCES

1. Smith TL, Kern RC, Palmer JN, Schlosser RJ, Chandra RK, Chiu AG, et al. Medical therapy vs surgery for chronic rhinosinusitis: a prospective, multi-institutional study. *Int Forum Allergy Rhinol* 2011;1:235-41.
2. Smith TL, Kern R, Palmer JN, Schlosser R, Chandra RK, Chiu AG, et al. Medical therapy vs surgery for chronic rhinosinusitis: a prospective, multi-institutional study with 1-year follow-up. *Int Forum Allergy Rhinol* 2013;3:4-9.
3. Smith KA, Smith TL, Mace JC, Rudmik L. Endoscopic sinus surgery compared to continued medical therapy for patients with refractory chronic rhinosinusitis. *Int Forum Allergy Rhinol* 2014;4:823-7.
4. Fokkens WJ, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, et al. European position paper on rhinosinusitis and nasal polyps 2012. *Rhinol Suppl* 2012;(23): 1-298, 3 p preceding table of contents.
5. Soler ZM, Hyer JM, Ramakrishnan V, Smith TL, Mace J, Rudmik L, et al. Identification of chronic rhinosinusitis phenotypes using cluster analysis. *Int Forum Allergy Rhinol* 2015;5:399-407.
6. Bent JP 3rd, Kuhn FA. Diagnosis of allergic fungal sinusitis. *Otolaryngol Head Neck Surg* 1994;111:580-8.
7. DeConde AS, Mace JC, Bodner T, Hwang PH, Rudmik L, Soler ZM, et al. SNOT-22 quality of life domains differentially predict treatment modality selection in chronic rhinosinusitis. *Int Forum Allergy Rhinol* 2014;4:972-9.
8. Soler ZM, Smith TL. Quality of life outcomes after functional endoscopic sinus surgery. *Otolaryngol Clin North Am* 2010;43:605-12, x.
9. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989;28:193-213.
10. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003;41:1284-92.
11. Smith TL, Litvack JR, Hwang PH, Loehrl TA, Mace JC, Fong KJ, et al. Determinants of outcomes of sinus surgery: a multi-institutional prospective cohort study. *Otolaryngol Head Neck Surg* 2010;142:55-63.
12. Quintanilla-Dieck L, Litvack JR, Mace JC, Smith TL. Comparison of disease-specific quality-of-life instruments in the assessment of chronic rhinosinusitis. *Int Forum Allergy Rhinol* 2012;2:437-43.
13. Schatz M, Hsu JW, Zeiger RS, Chen W, Dorenbaum A, Chipps BE, et al. Phenotypes determined by cluster analysis in severe or difficult-to-treat asthma. *J Allergy Clin Immunol* 2014;133:1549-56.
14. Ramakrishnan VR, Hauser LJ, Feazel LM, Ir D, Robertson CE, Frank DN. Sinus microbiota varies among chronic rhinosinusitis phenotypes and predicts surgical outcome. *J Allergy Clin Immunol* 2015;136:334-42.e1.
15. Fitzmaurice G, Laird N, Ware J. *Applied longitudinal analysis*. New York: John Wiley and Sons; 2004.