

Original Article

Responsiveness and Convergent Validity of a New Patient-Reported Outcome Measure for Chronic Rhinosinusitis (CRS-PRO)

Saied Ghadersohi, MD^a, Caroline P.E. Price, BA^a, Jennifer L. Beaumont, MS^b, Robert C. Kern, MD^{a,c}, David B. Conley, MD^a, Kevin C. Welch, MD^a, Alexis M. Calice, PA-C^a, Elizabeth Stanton, PA-C^a, Marisa K. VanderMeeden, PA-C^a, Sally E. Jensen, PhD^d, Anju T. Peters, MD^c, Leslie C. Grammer III, MD^c, Whitney W. Stevens, PhD, MS^{a,c}, Robert P. Schleimer, PhD^{a,c}, and Bruce K. Tan, MD, MS^{a,c} *Chicago, Ill; and Los Angeles, Calif*

What is already known about this topic? The 22-item Sino-Nasal Outcome Test is the best currently available patient-reported outcome measure for chronic rhinosinusitis (CRS). However, its development process may preclude Food and Drug Administration acceptance as an end point for use in clinical trials.

What does this article add to our knowledge? The 12-item CRS-PRO is a responsive and valid disease-specific patient-reported outcome measure of CRS. It is more concise than current patient-reported outcome measures (ie, 22-item Sino-Nasal Outcome Test) and was developed using extensive input from patients with CRS, including patients with CRS with nasal polyps.

How does this study impact current management guidelines? The validated CRS-PRO was developed with extensive documented input from patients with CRS. It was developed in concordance with Food and Drug Administration guidelines on patient-reported outcome measures acceptable for use as end points in clinical trials.

BACKGROUND: The CRS-PRO is a new patient-reported outcome measure (PROM) for chronic rhinosinusitis (CRS) that was developed using extensive patient input per Food and Drug Administration guidance on PROMs acceptable for use as end points in clinical trials.

OBJECTIVE: To assess the responsiveness and convergent validity of the CRS-PRO following standard-of-care medical therapy.

METHODS: This was a prospective study of 51 patients (21 with nasal polyps and 30 without) with newly diagnosed

CRS or having an acute CRS exacerbation who were initiated on appropriate medical therapy. At the baseline visit each patient completed the CRS-PRO questionnaire, the 22-item Sino-Nasal Outcome Test, the EuroQol 5-dimensional questionnaire, and 4 Patient-Reported Outcome Measure Information System short forms along with objective testing including endoscopic and radiographic scores, smell discrimination, and nasal inspiratory flow testing. This same battery of questionnaires and testing was administered at a follow-up visit 4 to 8 weeks later.

^aDepartment of Otolaryngology – Head and Neck Surgery, Northwestern University, Feinberg School of Medicine, Chicago, Ill

^bTerasaki Research Institute, Los Angeles, Calif

^cDivision of Allergy-Immunology, Northwestern University, Feinberg School of Medicine, Chicago, Ill

^dDepartment of Medical Social Sciences, Northwestern University, Feinberg School of Medicine, Chicago, Ill

This study was supported by the National Institutes of Health (grant nos. U01 AI106683-Chronic Rhinosinusitis Integrative Studies Program, K23DC12067, and R01DC016645) and the Ernest S. Bazley Foundation.

Conflicts of interest: R. C. Kern is a consultant of Lyra Therapeutics, Sanofi/Regeneron, GSK, and Genentech. A. T. Peters reports personal fees from Sanofi/Regeneron, grants and personal fees from AstraZeneca and Optinose, outside the submitted work. L. C. Grammer III has received grants from the National Institutes of Health and Bazley Foundation to support her research; has received grants for her institution from the National Institute of Health Food Allergy Network, AstraZeneca, and Sanofi/Regeneron; serves as a consultant for Astellas Pharmaceuticals; and has received royalties and payment for lectures from the American Academy of Allergy, Asthma and Immunology (AAAAI), Mount Sinai, Lippincott, UpToDate, Elsevier, and Wolters Kluwer. W. W. Stevens has served on an advisory board for GlaxoSmithKline. R. P. Schleimer is a consultant for

Intersect ENT, GlaxoSmithKline, Merck, Sanofi, AstraZeneca/Medimmune, Genentech, Otsuka, Actobio Therapeutics, Lyra Therapeutics, Astellas Pharm Inc, Genzyme/Sanofi Corp, and Celgene Corp; is also a consultant with stock/stock options that are not valued yet with Allakos, Aurasense, BioMarck, Exicure, and Aqualung Therapeutics Corp; has a patent with Allakos related to Siglec-8 and Siglec-8 ligand related products that have not been developed yet; and has received grants from the National Institutes of Health to support this and other research. B. K. Tan has received grants from the National Institutes of Health to support this research and has served on advisory boards for Sanofi/Genzyme and Optinose. All other authors declare that they have no relevant conflicts of interest. Received for publication June 4, 2019; revised March 30, 2020; accepted for publication April 15, 2020.

Available online ■■

Corresponding author: Bruce K. Tan, MD, MS, Department of Otolaryngology-Head and Neck Surgery, Feinberg School of Medicine, Northwestern University, 676 N Saint Clair Ste 1325, Chicago, IL 60611. E-mail: btan@nm.org.

2213-2198

© 2020 American Academy of Allergy, Asthma & Immunology

<https://doi.org/10.1016/j.jaip.2020.04.031>

Abbreviations used

BSIT- brief smell identification test
 CRS- chronic rhinosinusitis
 CRSsNP- chronic rhinosinusitis without nasal polyps
 CRSwNP- chronic rhinosinusitis with nasal polyps
 CT- computed tomography
 EQ-5D questionnaire- EuroQol 5-dimensional questionnaire
 FDA- Food and Drug Administration
 NPIF- nasal peak inspiratory flow
 PRO- patient-reported outcome
 PROM- patient-reported outcome measure
 PROMIS- Patient-Reported Outcome Measure Information System
 QOL- quality of life
 SNOT-22- 22-item Sino-Nasal Outcome Test

RESULTS: We verified that shortening the 21-item CRS-PRO to 12 items as previously described maintains its psychometric properties. The 12-item CRS-PRO was responsive with a large effect size (Cohen's *d*, 0.94) comparable to the longer 22-item Sino-Nasal Outcome Test (Cohen's *d*, 0.93). The instrument was slightly more responsive to medically treated patients with CRS without nasal polyps compared with patients with CRS with nasal polyps (Cohen's *d*, 1.1 vs 0.89, respectively). The change in 12-item CRS-PRO total score has moderate correlation with change in Lund-Mackay computed tomography scores.

CONCLUSIONS: The CRS-PRO is a 12-item rigorously developed, responsive, and valid PROM that was developed using extensive input from patients with current definitions of CRS, including its 2 major phenotypes. © 2020 American Academy of Allergy, Asthma & Immunology (J Allergy Clin Immunol Pract 2020;■:■-■)

Key words: Chronic rhinosinusitis; Patient-reported outcome measure; Responsiveness; Validity; CRS-PRO; Nasal polyps

INTRODUCTION

Chronic rhinosinusitis (CRS) is an inflammatory disease of the paranasal sinuses whose symptoms affect between 10.9% and 11.9% of respondents in symptom-based surveys in Europe and United States.^{1,2} Patients with CRS experience impaired quality of life (QOL) and previous studies demonstrate improvements among patients with CRS after endoscopic sinus surgery in both disease-specific symptoms and general health domains such as bodily pain, social functioning, fatigue, pain interference, sleep disturbance, and anxiety.^{3,4}

Patient-reported outcome (PRO) instruments can measure disease severity and symptom burden in chronic diseases and provide the patients' own interpretation of their health status without modification by a clinician. For PROs to be acceptable end points in clinical trials, instruments need to detect clinically significant spontaneous or treatment-related changes over a period of time. The sensitivity of a PRO to detect these changes, or the responsiveness of the PRO, is an important attribute when selecting PROs for measuring treatment-related effects from the intervention. General health measures such as the EuroQol 5-dimensional (EQ-5D) questionnaire, short-form 36 health survey, and Patient-Reported Outcome Measure Information System (PROMIS) 29 have domains that are sensitive to the large clinical changes following surgical treatment for CRS but

do not measure impairment from CRS-specific parameters and subsequently are less responsive to smaller changes such as those following medical therapy for CRS.^{4,5}

A recent systematic review by Rudmik et al⁶ assessed the quality of available CRS-specific PROMs and found that the 22-item Sino-Nasal Outcomes Test (SNOT-22) was the best currently available CRS-specific PRO instrument based on development and psychometric quality analysis.⁶ However, many CRS-specific instruments, including the SNOT-22, were developed from patient-generated content before current definitions of CRS. Potentially for this reason, some symptom items are not found at high prevalence of average severity among patients with CRS or are minimally responsive to treatment.⁴ In addition, many CRS-specific PROMs were developed or modified with limited input from patients and therefore include compound symptom descriptors (eg, "nasal obstruction/congestion"), items lacking situational context (eg, "embarrassed"), or potentially highly correlated items (eg, "lack of good night's sleep" and "waking up tired"). Some instruments use long recall time frames (eg, 8 weeks), which make it difficult to measure relatively rapid symptom changes following treatment.^{3,7}

Previously we introduced the CRS-PRO instrument, which was initially developed with 21 draft items, as a disease-specific PROM for CRS. It was designed with significant patient input and validation to meet Food and Drug Administration (FDA) guidance on PROs acceptable for use as end points in clinical trials. In this previous assessment of reliability in a cohort of patients with stable symptoms, we found that the 21 items could be reduced to 12 items by eliminating multiple items from similar conceptual constructs, highly correlated items, items with low baseline severity, and items that reduced internal consistency without compromising reliability. One of the goals of this rigorously developed patient-centered tool was to accurately measure the relatively rapid symptom fluctuations in CRS during exacerbations and resolution following treatment. In the initial study, we discussed the development and reliability assessment of the instrument in patients with stable CRS. This study will assess the responsiveness of the CRS-PRO instrument after administration of standard-of-care medical therapy. In addition, we will assess the clinical interpretability and convergent validity of the CRS-PRO instrument by comparing with the SNOT-22, general QOL measures, and objective clinical measures used in CRS.

METHODS

Subjects

Approval for this study was obtained from the Northwestern University Institutional Review Board, and informed consent was obtained from each enrolled subject. We prospectively enrolled 54 patients between November 2016 and July 2017 from the Northwestern Medical Group Department of Otolaryngology-Head and Neck Surgery, a tertiary care referral center. All patients were at least 18 years old, were either newly diagnosed with CRS (by European position paper on rhinosinusitis and nasal polyps criteria including documented radiographic or endoscopic evidence of CRS) or carried a previous diagnosis of CRS but were experiencing acute exacerbation of symptoms.⁸ All patients were prescribed standard-of-care medical therapy on the basis of type of CRS and severity of disease (Table 1). Exclusion criteria included limited English comprehension, patients who were or were trying to get pregnant, and patients who were initiated on medical therapy (excluding topical therapy) within the 2 weeks before enrollment.

TABLE I. Patient baseline demographic information including disease status and comorbidities

Characteristic	All patients	Patients with CRSsNP	Patients with CRSwNP
N (%)	51 (100)	30 (59)	21 (41)
Age (y), mean $\pm$ SD (range)	44.3 $\pm$ 14.6 (25-79)	44.6 $\pm$ 16.6 (25-79)	44 $\pm$ 11.7 (27-69)
Race			
White	41 (80)	22 (73)	19 (90)
Black	5 (10)	5 (17)	0 (0)
Asian	3 (6)	2 (7)	1 (5)
Other	2 (4)	1 (3)	1 (5)
Sex			
Female	26 (51)	19 (63)	7 (33)
Male	25 (49)	11 (37)	14 (67)
CRS Exacerbation	24 (47)	10 (33)	14 (67)*
New CRS diagnosis	27 (53)	20 (67)	7 (33)*
Previous ESS	21 (41)	8 (27)	13 (62)*
Asthma	21 (41)	8 (27)	13 (62)*
AERD	1 (2)	0 (0)	1 (5)
Allergies	29 (57)	16 (53)	13 (62)
Smoking status			
Current	1 (2)	0 (0)	1 (5)
Former	14 (27)	6 (20)	8 (38)
Never	36 (71)	24 (80)	12 (57)
Medical therapy administered			
Oral antibiotics	41 (80)	24 (80)	17 (81)
Oral steroids	34 (67)	17 (57)	17 (81)
Both antibiotics and oral steroids	26 (51)	13 (43)	13 (62)
Intranasal steroids	51 (100)	30 (100)	21 (100)

AERD, Aspirin-exacerbated respiratory disease; ESS, endoscopic sinus surgery. Values are n (%) unless otherwise indicated.

* $P < .05$.

The patients were seen on 2 separate visits before and after standard-of-care medical therapy. At the baseline visit, we collected pertinent patient information including age, sex, race, CRS subtype (with or without polyps), history of asthma, allergies, aspirin sensitivity, smoking status, and endoscopic sinus surgery. The patients completed the 21-item CRS-PRO questionnaire, along with the SNOT-22, 4 PROMIS short forms (Satisfaction of Participation in Discretionary Social Activities, Fatigue, Pain Interference, and Sleep Disturbance), and finally the 5-level EQ-5D including the visual analog scale.^{5,9,10} Each PROMIS questionnaire has a standardized scoring system that generates *T* scores (mean, 50 ± 10 in the general population). To assess the relationship between symptoms and objective measures, during this same clinic visit, patients underwent a comprehensive battery of testing including low-dose in-office cone beam computed tomography (CT) scan, comprehensive nasal endoscopy assessing for nasal edema, discharge, or polyps, assessments of nasal peak inspiratory flow (NPIF) rate, and brief smell identification test (BSIT) at each visit.^{11,12} CT scans were scored on the basis of Lund-Mackay scoring system, assessing the severity of CT findings for CRS.¹² Endoscopic scores were rated on each side in 3 domains, presence of polyps (grade 0: absent, grade 1: polyps limited to the middle meatus, grade 2: polyps extending beyond the middle meatus into the nasal cavity, grade 3: polyps extending to inferior aspect of inferior turbinate, grade 4: polyps extending to floor of the nasal cavity).¹³ Other endoscopic parameters collected included the presence of edema (0-absent, 1-mild, 2-severe) and presence of discharge (0-absent, 1-thin/clear drainage, 2-thick/purulent drainage).¹⁴ Each nasal cavity was scored

separately and the patient's score was reported as the sum of both sides.

The patients then followed up 4 to 8 weeks later completing the same battery of questionnaires and objective testing, including a transition rating comparing pre- and posttreatment health-related QOL problems on a 5-point scale (1 = "much better," 2 = "a little better," 3 = "about the same," 4 = "a little worse," 5 = "much worse"). Study data were collected and managed using REDCap electronic data capture tools hosted at Northwestern University Feinberg School of Medicine.¹⁵ Any missing data points (0.3% of item answers) were either imputed on the basis of average for that measure or duplicated if there were thematically similar questions (eg, "nose blocked" and "difficulty breathing through nose" item in the CRS-PRO instrument) between the different instruments.

Responsiveness

We analyzed the responsiveness (ability to detect clinical change) of the CRS-PRO instrument in all patients and by CRS type comparing total scores between the baseline and follow-up visits using a paired *t* test. We also calculated the effect size (magnitude of change in symptom burden/SD of baseline measure) and compared this with the effect size obtained with the SNOT-22 and PROMIS questionnaires using the Cohen's *d*. An effect size of more than 0.2 is considered small, more than 0.5 moderate, and more than 0.8 is considered large.¹⁶

Clinical interpretability

To provide a clinical interpretation of scores, we used an anchor-based approach, with groups defined by a transitional rating

TABLE II. Comparison of baseline and follow-up objective scores for all patients, patients with CRSsNP, and patients with CRSwNP

	Baseline	Follow-up
All patients (N = 51)		
CT score	11.2 ± 5.0 (9.7 to 12.6)	8.0 ± 5.6 (6.4 to 9.6)*
Endoscopy scoring		
Polyp (n = 21)	2.4 ± 1.7 (1.7 to 3.2)	1.8 ± 1.7 (1.0 to 2.5)*
Edema	2.2 ± 1.4 (1.9 to 2.6)	1.4 ± 1.3 (0.99 to 1.7)*
Discharge	2.0 ± 1.5 (1.6 to 2.4)	0.78 ± 1.0 (0.50 to 1.1)*
BSIT score	7.68 ± 3.4 (6.7 to 8.6)	8.9 ± 2.9 (8.1 to 9.7)*
NPIF rate	135.6 ± 49.9 (121.4 to 149.8)	136.8 ± 40.7 (124.3 to 147.3)
Patients with CRSsNP (N = 30)		
CT score	9.7 ± 4.8 (7.9 to 11.4)	6.2 ± 5.2 (4.2 to 8.2)*
Endoscopy scoring		
Polyp	NA	NA
Edema	1.9 ± 1.4 (1.4 to 2.4)	0.8 ± 0.9 (0.45 to 1.1)*
Discharge	2 ± 1.4 (1.5 to 2.5)	0.63 ± 0.96 (0.27 to 0.99)*
BSIT score	9.1 ± 2.6 (8.2 to 10.1)	9.8 ± 2.1 (9.0 to 10.6)
NPIF rate	125.9 ± 43.3 (109.4 to 142.4)	128 ± 32.1 (116.0 to 140.0)
Patients with CRSwNP (N = 21)		
CT score	13.3 ± 4.6 (11.2 to 15.4)	10.5 ± 5.3 (8.1 to 12.9)*
Endoscopy scoring		
Polyp	2.4 ± 1.7 (1.7 to 3.2)	1.8 ± 1.7 (1.0 to 2.5)*
Edema	2.7 ± 1.2 (2.2 to 3.3)	2.1 ± 1.3 (1.5 to 2.7)
Discharge	2 ± 1.5 (1.3 to 2.7)	1 ± 1.1 (0.50 to 1.5)*
BSIT score	5.7 ± 3.3 (4.1 to 7.2)	7.8 ± 3.4 (6.2 to 9.3)*
NPIF rate	149 ± 55.9 (123.6 to 174.5)	149.5 ± 48.5 (127.4 to 171.6)

Values are mean ± SD (95% CI).

CT score refers to the Lund-Mackay CT scoring system, which is scored with increasing severity on a 0-24 scale. Endoscopy scoring was carried out using the sum of the polyp score of each individual side (0-4) and the sum of the edema and discharge scores of each side (0-2). BSIT is scored from 0 to 12 with ascending olfactory identification.

* $P < .01$.

scale comparing pre- and posttreatment health (1 = “much better,” 2 = “a little better,” 3 = “about the same,” 4 = “a little worse,” 5 = “much worse”). Mean changes in CRS-PRO and SNOT-22 scores were calculated within these groups to assess clinical interpretability.^{10,17}

Convergent validity

To assess convergent validity, we calculated the Pearson correlations between the changes in the CRS-PRO score and changes in the (1) SNOT-22, (2) PROMIS short forms, and (3) objective measures including Lund-Mackay CT score, endoscopic findings, BSIT score, and NPIF rate. All statistical analyses were performed using Stata 14.1 (Statacorp, College Station, Texas). Demographic and clinical characteristics were reported as frequencies and percentages for categorical variables, and means, SDs, and range for continuous variables. Significance was determined at $P < .05$.

RESULTS

Demographic characteristics

A total of 54 patients were enrolled in the responsiveness study; however, 3 patients declined further participation or were lost to follow-up. Patients without follow-up data were excluded from further analysis. Table I summarizes the baseline demographic characteristics of all patients and categorized by chronic rhinosinusitis without nasal polyps (CRSsNP) and chronic rhinosinusitis with nasal polyps (CRSwNP). Twenty-one of the patients (41.2%) had CRSwNP, whereas 30 (58.8%) had

CRSsNP. There were expected differences in demographic characteristics of patients with CRSwNP, noting a higher percentage of patients having a CRS exacerbation rather than new diagnosis, higher rates of previous endoscopic sinus surgery, and comorbid asthma. Patients enrolled in the study received standard-of-care medical therapy at the discretion of their treating physicians on the basis of presenting disease features, previous culture results, and patient-specific contraindications. Overall, 41 (80.4%) patients received up to a 3-week course of oral antibiotics. Typical antibiotics chosen included doxycycline (41.2%), amoxicillin with clavulanate (21.6%), clarithromycin (11.8%), or trimethoprim-sulfamethoxazole (6%). In addition, 34 (66.7%) patients received up to a 3-week course of oral steroids. Twenty-six (51%) patients received both antibiotics and oral steroids. Finally, 51 (100%) patients were started or continued on topical nasal steroids administered as a nasal spray or nasal rinses in pre- and postsurgical patients, respectively (Table I).

Baseline CT showed an average Lund-Mackay CT score of 11.2 ± 5.0 , which improved to 8.0 ± 5.6 after medical therapy ($P < .00005$). Endoscopic scores also significantly improved after medical therapy, average total edema score from 2.2 ± 1.4 to 1.4 ± 1.3 ($P < .00005$), average total drainage score from 2.0 ± 1.5 to 0.78 ± 1.0 ($P < .00005$), and average total polyp score from 2.4 ± 1.7 to 1.8 ± 1.7 , among the patients who had polyps ($P = .0095$) (Table II). As expected by comparison, patients with CRSwNP were more likely than those with CRSsNP to have

TABLE III. Symptom burden for each item on the 21-item draft CRS-PRO at baseline to follow-up and the effect size of each item

CRS-PRO 21-items	Baseline, mean $\pm$ SD	Follow-up, mean $\pm$ SD	Effect size, Cohen's <i>d</i> (95% CI)
Physical symptoms			
1) I had difficulty breathing through my nose	2.6 $\pm$ 1.2	1.5 $\pm$ 1.4*	0.84 (0.43 to 1.2)
XXX 2) <i>My nose was blocked</i>	2.6 $\pm$ 1.2	1.6 $\pm$ 1.3*	0.83 (0.42 to 1.2)
3) I felt pressure in my face	2.2 $\pm$ 1.4	1.6 $\pm$ 1.4†	0.36 (−0.028 to 0.75)
4) My face hurt	1.6 $\pm$ 1.5	1.1 $\pm$ 1.3†	0.32 (−0.068 to 0.71)
5) I had to blow my nose	2.8 $\pm$ 1.2	1.8 $\pm$ 1.2*	0.84 (0.43 to 1.2)
6) I have been coughing	1.6 $\pm$ 1.3	1.2 $\pm$ 1.3†	0.27 (−0.11 to 0.66)
XXX 7) <i>I had trouble clearing mucus from my throat</i>	2.1 $\pm$ 1.4	1.4 $\pm$ 1.3*	0.51 (0.11 to 0.90)
8) I had mucus in my throat	2.4 $\pm$ 1.2	1.5 $\pm$ 1.3*	0.70 (0.30 to 1.1)
9) I had mucus in my nose	3.3 $\pm$ 0.84	2.1 $\pm$ 1.1*	1.1 (0.71 to 1.5)
XXX 10) <i>I had mucus dripping from my nose</i>	2.4 $\pm$ 1.0	1.5 $\pm$ 1.1*	0.88 (0.47 to 1.3)
Sensory impairment			
11) I had problems with my sense of smell	2.1 $\pm$ 1.6	1.3 $\pm$ 1.4*	0.57 (0.17 to 0.96)
XXX 12) <i>I was able to enjoy the taste of food</i>	1.9 $\pm$ 1.3	1.7 $\pm$ 1.3	0.16 (−0.23 to 0.55)
Psychosocial effects			
13) My symptoms kept me awake at night	1.5 $\pm$ 1.2	0.8 $\pm$ 1.1*	0.65 (0.25 to 1.0)
14) I felt fatigued	2.1 $\pm$ 1.2	1.3 $\pm$ 1.3*	0.60 (0.21 to 1.0)
XXX 15) <i>My sleep was refreshing</i>	2.6 $\pm$ 1.0	2.2 $\pm$ 1.2†	0.34 (−0.05 to 0.73)
XXX 16) <i>I felt anxious about the uncertainty of my chronic rhinosinusitis</i>	1.5 $\pm$ 1.1	1.0 $\pm$ 0.96*	0.43 (0.04 to 0.82)
17) I worried that my condition will get worse	1.8 $\pm$ 1.2	1.2 $\pm$ 1.1*	0.56 (0.16 to 0.96)
XXX 18) <i>I felt overwhelmed by my condition</i>	1.1 $\pm$ 1.1	0.73 $\pm$ 1.2†	0.36 (−0.031 to 0.75)
19) I was frustrated by my condition	2.2 $\pm$ 1.3	1.4 $\pm$ 1.3*	0.61 (0.22 to 1.0)
Impact			
XXX 20) <i>Because of my illness, some people avoided me</i>	0.29 $\pm$ 0.58	0.24 $\pm$ 0.51	0.11 (−0.28 to 0.50)
XXX 21) <i>Because of my illness, I felt embarrassed in social situations</i>	0.90 $\pm$ 1.1	0.53 $\pm$ 0.78†	0.39 (0.0015 to 0.79)

Symptoms were scored 0–4, with 4 representing the most severe (note that items 12 and 15 are reverse scored). The italicized items also denoted by XXX are ones removed in the abbreviated 12-item CRS-PRO.

* $P < .01$.

† $P < .05$.

worse Lund-Mackay CT scores ($P = .0094$), endoscopic edema score ($P = .032$), and BSIT scores ($P = .0001$).

Baseline scores

Individual item baseline and follow-up scores were calculated for patients in this cohort (Table III). We sought to verify our previous analysis suggesting that the 12-item CRS-PRO could sufficiently reflect patient experience with CRS. On calculating interitem correlations, we once again found that items 1 and 2, items 7 and 8, and items 16 to 19 had high interitem correlations of more than 0.75, suggesting they were measuring highly overlapping experiences. This verified that items 2, 7, 16, and 18 could be omitted as discussed in our previous reliability study. In addition, item 9 “I had mucus in my nose” was noted to have better responsiveness than its dyad question item 10 “I had mucus dripping from my nose.” Similarly, when comparing the “sleep quality” items 13 and 15, item 15 had higher responsiveness. The 2 “impact” items (items 20 and 21) continued to

show low baseline severity, with the average patient reporting “never” or “rarely” being affected by these items.

Responsiveness

The ability to detect clinical change was assessed by comparing mean scores at the baseline and follow-up visit and by assessing effect size. The 12-item CRS-PRO was noted to have a statistically significant decrease in the baseline score from 26.2 ± 8.5 to 17.0 ± 10.8 at the follow-up visit ($P < .00005$). Effect size for the 12-item CRS-PRO was large at 0.95 (95% CI, 0.54–1.4). In comparison, the 21-item CRS-PRO decreased from 41.6 ± 13.5 to 27.9 ± 16.2 ($P < .00005$) with an effect size of 0.92 (95% CI, 0.51–1.3). This was similar to the SNOT-22, for which the baseline score decreased from 44 ± 18.7 to 25.6 ± 20.7 at the follow-up visit ($P < .00005$), with an effect size of 0.93 (95% CI, 0.52–1.3). The SNOT-22 subdomain effect sizes were also assessed, with the largest effect size of 1.26 (95% CI, 0.83–1.7) for the rhinological subdomain and medium effect sizes for the other SNOT-22 subdomains (Table IV).

TABLE IV. Responsiveness assessment: Comparison of baseline and follow-up scores and effect sizes of total instrument and subdomain scores of the studied PROMs

Measure	Baseline, mean $\pm$ SD (95% CI)	Follow-up, mean $\pm$ SD (95% CI)	Effect size (95% CI), Cohen's <i>d</i> (95% CI)
CRS-PRO 21-item	41.6 $\pm$ 13.5 (37.8 to 45.5)	27.9 $\pm$ 16.2 (23.4 to 32.4)*	0.922 (0.51 to 1.3)
CRS-PRO 12-item	26.2 $\pm$ 8.5 (23.8 to 28.6)	17.0 $\pm$ 10.8 (13.9 to 20.0)*	0.949 (0.54 to 1.4)
SNOT-22	44.0 $\pm$ 18.7 (38.7 to 49.2)	25.6 $\pm$ 20.7 (19.8 to 31.4)*	0.932 (0.52 to 1.3)
SNOT Rhinological	19.1 $\pm$ 7.0 (17.2 to 21.1)	10.2 $\pm$ 7.3 (8.1 to 12.2)*	1.26 (0.83 to 1.7)
SNOT Ear/pain	5.5 $\pm$ 3.9 (4.4 to 6.7)	3.7 $\pm$ 3.7 (2.7 to 4.8)†	0.469 (0.075 to 0.862)
SNOT Sleep	5.9 $\pm$ 4.3 (4.7 to 7.1)	3.4 $\pm$ 4.0 (2.3 to 4.5)*	0.597 (0.199 to 0.993)
SNOT Psychological	9.1 $\pm$ 6.7 (7.2 to 11.0)	5.6 $\pm$ 6.5 (3.8 to 7.4)*	0.541 (0.145 to 0.935)
PROMIS Fatigue	56.9 $\pm$ 9.7 (54.2 to 59.6)	49.5 $\pm$ 11.2 (46.4 to 52.7)*	0.705 (0.30 to 1.1)
PROMIS Sleep	52.9 $\pm$ 4.7 (51.6 to 54.2)	51.2 $\pm$ 4.0 (50.1 to 52.3)‡	0.385 (−0.010 to 0.78)
PROMIS Social	48.1 $\pm$ 9.3 (45.6 to 50.8)	52.9 $\pm$ 9.8 (50.2 to 55.7)*	−0.495 (−0.89 to −0.10)
PROMIS Pain	51.2 $\pm$ 8.9 (48.6 to 53.7)	47.4 $\pm$ 8.5 (45.0 to 49.8)†	0.431 (0.04 to 0.82)
5-Level EQ-5D questionnaire (n = 49)	74.5 $\pm$ 16.8 (69.7 to 79.3)	81 $\pm$ 13.5 (76.8 to 84.5)*	−0.429 (−0.82 to −0.03)

The 21-item CRS-PRO score range is 0-84, 12-item CRS-PRO score range is 0-48, SNOT-22 score range is 0-110, SNOT Rhinological score range is 0-40, SNOT Ear/face score range is 0-20, SNOT Sleep score range is 0-20, SNOT Psychological score range is 0-30, and EQ-5D questionnaire score range is 0-100. PROMIS instruments are scored on the basis of standardized *T* scores.

**P* < .0001.

†*P* < .01.

‡*P* < .05.

The mean changes in scores and effect sizes were also calculated separately for patients with and without nasal polyposis (see Tables E1 and E2 in this article's Online Repository at www.jaci-inpractice.org). Overall, the 12-item CRS-PRO was noted to decrease significantly from baseline in the CRSsNP group (23.2 ± 6.9 to 14.4 ± 8.6 ; *P* < .00005) and in the CRSwNP group (30.5 ± 8.9 to 20.7 ± 12.5 ; *P* = .001), during the follow-up visit. The effect size was greater for CRSsNP (1.12) than for CRSwNP (0.90). Similarly, the effect size for the SNOT-22 was greater for CRSsNP (1.13; 95% CI, 0.58-1.7) than for CRSwNP (0.78; 95% CI, 0.14-1.4).

Clinical interpretation

Using the anchor-based method with the transitional rating system, the mean changes in the CRS-PRO 12-item score for each of the clinical status rating groups are presented in Table V. For both the CRS-PRO and the SNOT-22, the mean change in the “about the same” group is negligible, whereas the mean changes are increasingly greater in magnitude for the groups whose overall health was “a little better,” “moderately better,” and “very much better.” Fortunately, only 4 patients reported responses as “worse” after medical therapy (Table V).

Floor and ceiling effects

Before and after medical therapy, 0% of patients scored the highest or lowest possible score on the 12-item CRS-PRO instrument, indicating that the full range of the construct was

assessed. In comparison, on the SNOT-22, no patients scored the highest or lowest score before medical therapy; however, 2 (3.9%) patients did score the lowest possible score (0 on SNOT-22) after medical therapy.

Convergent validity

We assessed the convergent validity of the CRS-PRO instrument by correlating the change in CRS-PRO mean score after medical therapy to the change in objective measure scores (Table VI). The change in the 12-item CRS-PRO score had a moderate correlation to changes in objective measures including Lund-Mackay CT score of 0.49 (*P* < .01) and endoscopic polyp score of 0.52 (*P* < .05). However, it did not correlate well with changes in the BSIT scores, NPIF rate, and endoscopic scores of edema or discharge. Change in SNOT-22 total scores also had a similar pattern of correlations, with moderate correlation to changes in Lund-Mackay CT score (0.52; *P* < .01), BSIT score (−0.32; *P* < .05), endoscopic polyp scores (0.56; *P* < .05) and poor correlations to changes in the NPIF rate and endoscopic edema and discharge scores. In addition, we assessed the validity of the instruments in the CRSsNP and the CRSwNP groups. There were overall higher correlations between change in instrument scores and objective scores (Lund-Mackay CT score, endoscopic polyp) in the CRSwNP group than in the CRSsNP group (data not shown).

TABLE V. Assessment of clinical interpretability—Mean change in CRS-PRO and SNOT-22 scores based on an anchor value reflecting change in overall patient health after medical therapy

Change in overall health	N	CRS-PRO 12-item	CRS-PRO 21-item	SNOT-22
Very much better	15	−16.5 $\pm$ 9.4 (−31 to −1)	−23.5 $\pm$ 16.2 (−46 to 5)	−31.8 $\pm$ 15.9 (−60 to −7)
Moderately better	13	−12.8 $\pm$ 9.1 (−33 to 7)	−19.8 $\pm$ 12.1 (−45 to 6)	−25.9 $\pm$ 19.8 (−38 to −8)
A little better	9	−7.4 $\pm$ 5.9 (−16 to 3)	−12.4 $\pm$ 9.4 (−27 to 3)	−20.4 $\pm$ 10.6 (−38 to −8)
About the same	10	0.5 $\pm$ 5.4 (−5 to 10)	1.3 $\pm$ 8.3 (−8 to 15)	3.2 $\pm$ 17.1 (−17 to 38)
Worse	4	1.8 $\pm$ 8.3 (−8 to 10)	2.3 $\pm$ 16.0 (−18 to 19)	7.0 $\pm$ 15.0 (−10 to 24)

Values represent mean $\pm$ SD (range).

TABLE VI. Correlation coefficients of the change in instrument and subdomain scores to change in objective measure scores

Measure	Lund-Mackay CT score	BSIT score	NPIF rate	Endoscopic scores		
				Polyp (N = 21)	Discharge	Edema
CRS-PRO 21-item	0.47*	−0.31†	0.011	0.48†	0.22	0.18
CRS-PRO 12-item	0.49*	−0.27	0.0029	0.52†	0.26	0.22
SNOT-22	0.52*	−0.32†	−0.068	0.56†	0.30†	0.24
SNOT Rhinological	0.63*	−0.41*	0.0029	0.62*	0.29†	0.35†
SNOT Ear/face	0.21	−0.18	−0.022	0.34	0.19	0.09
SNOT Sleep	0.21	−0.18	−0.13	0.34	0.26	0.17
SNOT Psychological	0.40*	−0.25	−0.12	0.46†	0.2	0.096

* $P < .01$.† $P < .05$.

We also compared the change in CRS-PRO mean scores with the change in PROMIS short-form and EQ-5D questionnaire mean scores. The change in the 12-item CRS-PRO total instrument score had a low correlation to the EQ-5D questionnaire (-0.32 ; $P < .05$) and the PROMIS pain (0.35 ; $P < .01$) scores. The PROMIS fatigue (0.56 ; $P < .01$) and PROMIS Satisfaction of Participation in Discretionary Social Activities (-0.51 ; $P < .05$) scores had a moderate correlation to the change in the 12-item CRS-PRO. The change in the SNOT-22 score had a moderate correlation to the PROMIS fatigue score (0.57 ; $P < .01$) and had a low correlation to the PROMIS sleep (0.30 ; $P < .05$), PROMIS pain (0.31 ; $P < .05$), and PROMIS satisfaction (-0.37 ; $P < .01$) scores. The SNOT-22 score did not correlate to the EQ-5D questionnaire visual analogue scale score (Table VII).

DISCUSSION

The 12-item CRS-PRO is a recently developed disease-specific patient-reported outcome measure for CRS specifically designed to measure the symptom fluctuations during exacerbations and after treatment. This study establishes the psychometric validity of the instrument and demonstrates that the CRS-PRO has excellent responsiveness to changes after medical management. The CRS-PRO has a large effect size of 0.95, which was the same as the effect size noted for the longer SNOT-22. The CRS-PRO was noted to have an excellent responsiveness in both patients with CRSsNP and patients with CRSwNP when analyzed separately. The CRS-PRO is able to capture the full spectrum of CRS clinical status and is associated with change in overall health. We also showed strong convergent validity with the

SNOT-22 and moderate convergent validity when comparing the change in CRS-PRO scores to the change in objective measures such as Lund-Mackay CT score (correlation, 0.49; $P < .01$) and endoscopic polyp scores among patients with CRSwNP. In addition, the CRS-PRO instrument had moderate correlation to general PROMs including the EQ-5D questionnaire and specific PROMIS short forms.

With a greater focus on patient-centered care, there is a need for PROMs that accurately inform clinicians of the patient experience and can be used in clinical trials. The FDA has developed guidelines for PROMs that are to be used in clinical trials for labeling purposes. At the current time, there are no PROMs that have been accepted by the FDA for use as an end point for CRSsNP or CRSwNP. With that shortcoming in mind, our interdisciplinary team of health psychologists, otolaryngologists, and allergists used psychometric theory to develop the CRS-PRO in concordance with the guidelines prescribed by the FDA for PROM development. We previously demonstrated that the instrument had significant patient input for development, is reliable, is internally consistent, and has good concurrent validity. Here, we demonstrate the excellent responsiveness of the 12-item CRS-PRO that is equivalent to the longer SNOT-22. As highlighted in our previous study, although the SNOT-22 has been well validated for studying CRS, it is unclear whether its evolution from the Rhinosinusitis Outcome Measure (RSOM) -31 and development before current European position paper on rhinosinusitis and nasal polyps definition for CRS that did not require objective signs for diagnosis can be acceptable as an end point for studying CRS and its 2 clinical phenotypes.^{4,8}

We believe that the use of the 1997 CRS diagnostic criteria may have led to the inclusion of items such as “dizziness” and “ear pain”

TABLE VII. Correlation of the change in instrument and subdomain score to change in general PROM scores (subjective scores)

Measure	SNOT-22	EQ-5D Questionnaire	PROMIS Fatigue	PROMIS Sleep	PROMIS Pain	PROMIS Satisfaction
CRS-PRO 21-item	0.75*	−0.39*	0.53*	0.19	0.34†	−0.44*
CRS-PRO 12-item	0.77*	−0.32†	0.56*	0.17	0.35*	−0.51*
SNOT-22	1	−0.25	0.57*	0.30†	0.31†	−0.37*
SNOT Rhinological	0.81*	−0.23	0.26	0.03	0.018	−0.15
SNOT Ear/face	0.72*	−0.18	0.49*	0.23	0.45*	−0.26
SNOT Sleep	0.82*	−0.21	0.53*	0.52*	0.33†	−0.36*
SNOT Psychological	0.87*	−0.18	0.60*	0.36†	0.36*	−0.43*

* $P < .01$.† $P < .05$.

that are more common in other disorders commonly masquerading as CRS such as allergic rhinitis or migraine disorders.¹⁸ In a supplemental analysis (see Table E3 in this article's Online Repository at www.jaci-inpractice.org; Table III), we can see that neither of these items had a statistically significant improvement after medical therapy, with effect sizes less than 0.10. In addition, the importance of patient input in the construction of the CRS-PRO and any PROM is highlighted by item 17 "I worried that my condition would get worse" and item 19 "I was frustrated by my condition." These 2 items with more specific wording performed significantly better from a responsiveness perspective, with effect sizes of more than 0.55. This is compared with the psychiatric items "sad" and "embarrassed" in the SNOT-22 that did not have a statistically significant improvement after medical therapy and therefore had poor responsiveness.

The CRS-PRO can be used for both phenotypes of CRS, and this study shows a high level of responsiveness for both groups. We demonstrate that both the CRS-PRO and SNOT-22 appear to be more responsive in patients with CRSsNP than in patients with CRSwNP but both groups have a "large" effect size. In addition, it appears that the patients with CRSwNP had a higher overall disease burden than the patients with CRSsNP as evidenced by higher baseline scores (CRS-PRO and SNOT-22). It is known that patients with CRSwNP are disproportionately affected by certain symptoms (eg, smell loss and lower airway symptoms). However, qualitative analysis of the focus group transcripts done during the initial development of the CRS-PRO found that CRSsNP and CRSwNP groups shared a similar conceptual framework and symptoms although there were subtle differences in the most important symptoms.

The 12-item CRS-PRO is scored as a total score of all items, with a maximum of 48. In an analysis that is not shown, we note that the 12-item CRS-PRO total score had a greater effect size than its conceptual subdomains, meaning that the total score was a more responsive measure capturing all the facets of CRS symptomatology. In contrast, the *post hoc* developed SNOT rhinological subdomain is noted to be more responsive than the SNOT-22 total score. This would be expected because the total SNOT-22 score contains multiple underlying constructs that may not be as important in CRS.^{19,20}

Legacy measures also use long time frames of symptom assessment, preventing accurate capture of transition states and possibly increasing recall bias. Most previously developed instruments were tested in a surgical patient cohort assessing preoperative and postoperative symptoms and outcomes over 2- to 12-week spans.^{10,21,22} However, most fluctuations in CRS patient symptoms including acute exacerbations, or short course of prednisone therapy, will worsen or improve over days to 1 to 2 weeks. The CRS-PRO instrument assesses items over the last 7 days, thereby allowing for a more timely and accurate capture of the kinetics of acute symptom changes.

Several previous studies have not been able to find a strong correlation between PROMs and objective findings such as a CT scan.²³⁻²⁵ One reason that is often cited is the variability in subjective and QOL outcomes that are evaluated at a single time point. We therefore decided to compare change in CRS-PRO and SNOT-22 total scores with change in objective measure scores, thereby using 2 time points in a single patient to remove inherent patient variability. In a study by Pallanch et al,²⁶ it was shown that change in volumetric CT score better correlated with change in symptom severity and QOL measures than Lund-

Mackay CT score. However, even with this improvement, the nasal symptoms only correlated on the order of 0.20 to 0.52 at best with the volumetric CT analysis.²⁶ In our series, the change in Lund-Mackay CT score correlated moderately with the change in both SNOT-22 and the 12-item CRS-PRO scores (0.52 and 0.49, respectively; $P < .01$). However, these correlations could potentially be improved with similar volumetric CT analysis.²⁶ Change in CRS-PRO scores only weakly correlated with change in endoscopic polyp scores. In comparison, change in the SNOT-22 score weakly correlated with change in the endoscopic discharge and polyp scores.

This study does have limitations including the use of a modest-sized sample of patients in a tertiary care center. Although the sample size was chosen to have a greater than 80% power to detect a relatively small effect size of 0.4, the relatively small cohort of patients precluded stable subanalyses of instrument subdomains and determination of minimal clinically important differences. Second, medical therapy administered in our patient group was variable and not standardized, but prescribed therapy was consistent with guideline-recommended treatments chosen using clinical findings, along with physician recommendations taking account of patient preference. In addition, we did not demonstrate divergent validity of the CRS-PRO because it was not necessarily based on guidelines for PROM validation.²⁷ Finally, it is difficult to differentiate measurement responsiveness from actual treatment efficacy. Nevertheless, this study demonstrates that the CRS-PRO instrument is responsive to clinical changes. Future goals will include ongoing further validation of the instrument in a broader representation of patients with CRS who are treated with both medical and surgical therapies, particularly in randomized controlled trials from which estimates of placebo effects and associations with other clinical metrics can be ascertained.

CONCLUSIONS

The CRS-PRO is a responsive PROM able to detect the relatively rapid clinical changes in patient symptoms and correlated with changes in Lund-Mackay CT scores. The CRS-PRO has similar psychometric measures and was developed with extensive input from patients with contemporary definitions of CRSsNP and CRSwNP in concordance with FDA guidelines for PROMs acceptable for use as end points in clinical trials.

Acknowledgments

REDCap is supported at the Feinberg School of Medicine by the Northwestern University Clinical and Translational Science Institute. Research reported in this publication was supported, in part, by the National Institutes of Health National Center for Advancing Translational Sciences (grant no. UL1TR001422). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Hirschi AG, Stewart WF, Sundaresan AS, Young AJ, Kennedy TL, Scott Greene J, et al. Nasal and sinus symptoms and chronic rhinosinusitis in a population-based sample. *Allergy* 2017;72:274-81.
2. Tomassen P, Newson RB, Hoffmans R, Lotvall J, Cardell LO, Gunnbjornsdottir M, et al. Reliability of EP30S symptom criteria and nasal endoscopy in the assessment of chronic rhinosinusitis—a GA(2) LEN study. *Allergy* 2011;66:556-61.

3. Gliklich RE, Metson R. The health impact of chronic sinusitis in patients seeking otolaryngologic care. *Otolaryngol Head Neck Surg* 1995;113:104-9.
4. Lidder AK, Detwiller KY, Price CP, Kern RC, Conley DB, Shintani-Smith S, et al. Evaluating metrics of responsiveness using patient-reported outcome measures in chronic rhinosinusitis. *Int Forum Allergy Rhinol* 2017;7:128-34.
5. Remenschneider AK, D'Amico L, Gray ST, Holbrook EH, Gliklich RE, Metson R. The EQ-5D: a new tool for studying clinical outcomes in chronic rhinosinusitis. *Laryngoscope* 2015;125:7-15.
6. Rudmik L, Hopkins C, Peters A, Smith TL, Schlosser RJ, Soler ZM. Patient-reported outcome measures for adult chronic rhinosinusitis: a systematic review and quality assessment. *J Allergy Clin Immunol* 2015;136:1532-1540.e1-1540.e2.
7. Benninger MS, Senior BA. The development of the Rhinosinusitis Disability Index. *Arch Otolaryngol Head Neck Surg* 1997;123:1175-9.
8. Fokkens WJ, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, et al. EPOS 2012: European position paper on rhinosinusitis and nasal polyps 2012. A summary for otorhinolaryngologists. *Rhinology* 2012;50:1-12.
9. Cook KF, Jensen SE, Schalet BD, Beaumont JL, Amtmann D, Czajkowski S, et al. PROMIS measures of pain, fatigue, negative affect, physical function, and social function demonstrated clinical validity across a range of chronic conditions. *J Clin Epidemiol* 2016;73:89-102.
10. Hopkins C, Gillett S, Slack R, Lund VJ, Browne JP. Psychometric validity of the 22-item Sinonasal Outcome Test. *Clin Otolaryngol* 2009;34:447-54.
11. Alt JA, Mace JC, Buniel MC, Soler ZM, Smith TL. Predictors of olfactory dysfunction in rhinosinusitis using the brief smell identification test. *Laryngoscope* 2014;124:E259-66.
12. Lund VJ, Mackay IS. Staging in rhinosinusitis. *Rhinology* 1993;31:183-4.
13. Gevaert P, Calus L, Van Zele T, Blomme K, De Ruyck N, Bauters W, et al. Omalizumab is effective in allergic and nonallergic patients with nasal polyps and asthma. *J Allergy Clin Immunol* 2013;131:110-116.e1.
14. Lund VJ, Kennedy DW. Quantification for staging sinusitis. The Staging and Therapy Group. *Ann Otol Rhinol Laryngol Suppl* 1995;167:17-21.
15. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377-81.
16. Health measurement scales: a practical guide to their development and use (5th edition). Aust N Z J Public Health 2016;40:294-5.
17. Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski CJ, Rothrock N, et al. Patient-reported Outcomes in Performance Measurement. Research Triangle Park, NC: RTI Press; 2015.
18. Lanza DC, Kennedy DW. Adult rhinosinusitis defined. *Otolaryngol Head Neck Surg* 1997;117:S1-7.
19. Browne JP, Hopkins C, Slack R, Cano SJ. The Sino-Nasal Outcome Test (SNOT): can we make it more clinically meaningful? *Otolaryngol Head Neck Surg* 2007;136:736-41.
20. Sedaghat AR, Gray ST, Caradonna SD, Caradonna DS. Clustering of chronic rhinosinusitis symptomatology reveals novel associations with objective clinical and demographic characteristics. *Am J Rhinol Allergy* 2015;29:100-5.
21. Atlas SJ, Metson RB, Singer DE, Wu YA, Gliklich RE. Validity of a new health-related quality of life instrument for patients with chronic sinusitis. *Laryngoscope* 2005;115:846-54.
22. Gliklich RE, Metson R. Techniques for outcomes research in chronic sinusitis. *Laryngoscope* 2015;125:2238-41.
23. Ryan WR, Ramachandra T, Hwang PH. Correlations between symptoms, nasal endoscopy, and in-office computed tomography in post-surgical chronic rhinosinusitis patients. *Laryngoscope* 2011;121:674-8.
24. Wabnitz DA, Nair S, Wormald PJ. Correlation between preoperative symptom scores, quality-of-life questionnaires, and staging with computed tomography in patients with chronic rhinosinusitis. *Am J Rhinol* 2005;19:91-6.
25. Zheng Y, Zhao Y, Lv D, Liu Y, Qiao X, An P, et al. Correlation between computed tomography staging and quality of life instruments in patients with chronic rhinosinusitis. *Am J Rhinol Allergy* 2010;24:e41-5.
26. Pallanch JF, Yu L, Delone D, Robb R, Holmes DR III, Camp J, et al. Three-dimensional volumetric computed tomographic scoring as an objective outcome measure for chronic rhinosinusitis: clinical correlations and comparison to Lund-Mackay scoring. *Int Forum Allergy Rhinol* 2013;3: 963-72.
27. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147-57.

ONLINE REPOSITORY

TABLE E1. Responsiveness assessment: Comparison of baseline and follow-up scores and effect sizes of total instrument and subdomain scores of the PROMs of patients with CRSsNP (N = 30)

Measure	Baseline, mean $\pm$ SD (95% CI)	Follow-up, mean $\pm$ SD (95% CI)	Effect size (95% CI)
CRS-PRO 21-item	36.8 $\pm$ 11.1 (32.6 to 40.9)	24.3 $\pm$ 13.2 (19.3 to 29.2)*	1.02 (0.48 to 1.6)
CRS-PRO 12-item	23.2 $\pm$ 6.9 (20.6 to 25.8)	14.4 $\pm$ 8.6 (11.2 to 17.6)*	1.12 (0.57 to 1.7)
SNOT-22	40.5 $\pm$ 17.2 (34.0 to 46.9)	21.5 $\pm$ 16.4 (15.3 to 27.6)*	1.13 (0.58 to 1.7)
SNOT Rhinological	17.3 $\pm$ 6.8 (14.7 to 19.8)	9.2 $\pm$ 6.7 (6.7 to 11.7)*	1.20 (0.65 to 1.7)
SNOT Ear/pain	5.0 $\pm$ 3.8 (3.6 to 6.4)	3.1 $\pm$ 3.2 (1.9 to 4.3)†	0.539 (0.021 to 1.05)
SNOT Sleep	5.8 $\pm$ 4.5 (4.1 to 7.5)	2.8 $\pm$ 3.8 (CI 1.4 to 4.2)*	0.717 (0.192 to 1.24)
SNOT Psychological	8.2 $\pm$ 6.8 (5.6 to 10.7)	4.0 $\pm$ 5.2 (2.1 to 6.0)†	0.687 (0.163 to 1.21)

The 21-item CRS-PRO score range is 0-84, 12-item CRS-PRO score range is 0-48, SNOT-22 score range is 0-110, SNOT Rhinological score range is 0-40, SNOT Ear/face score range is 0-20, SNOT Sleep score range is 0-20, SNOT Psychological score range is 0-30.

* $P < .0001$.

† $P < .01$.

TABLE E2. Responsiveness assessment: Comparison of baseline and follow-up scores and effect sizes of total instrument and subdomain scores of the PROMs of patients with CRSwNP (N = 21)

Measure	Baseline, mean $\pm$ SD (95% CI)	Follow-up, mean $\pm$ SD (95% CI)	Effect size, Cohen's d (95% CI)
CRS-PRO 21-item	48.6 $\pm$ 13.8 (42.3 to 54.9)	33.1 $\pm$ 18.7 (24.6 to 41.6)*	0.945 (0.30 to 1.6)
CRS-PRO 12-item	30.5 $\pm$ 8.9 (26.4 to 34.5)	20.7 $\pm$ 12.5 (15.0 to 26.4)*	0.903 (0.26 to 1.5)
SNOT-22	49 $\pm$ 20 (39.9 to 58.1)	31.5 $\pm$ 24.9 (20.1 to 42.8)†	0.776 (0.14 to 1.4)
SNOT Rhinological	21.8 $\pm$ 6.5 (18.8 to 24.8)	11.6 $\pm$ 8.0 (7.9 to 15.2)*	1.4 (0.72 to 2.1)
SNOT Ear/pain	6.3 $\pm$ 4.1 (4.5 to 8.2)	4.7 $\pm$ 4.3 (2.7 to 6.6)	0.397 (−0.22 to 1.0)
SNOT Sleep	6.0 $\pm$ 4.1 (4.1 to 7.9)	4.2 $\pm$ 4.2 (2.3 to 6.2)	0.423 (−0.19 to 1.0)
SNOT Psychological	10.5 $\pm$ 6.6 (7.5 to 13.5)	7.8 $\pm$ 7.5 (4.3 to 11.2)	0.390 (−0.22 to 1.0)

The 21-item CRS-PRO score range is 0-84, 12-item CRS-PRO score range is 0-48, SNOT-22 score range is 0-110, SNOT Rhinological score range is 0-40, SNOT Ear/face score range is 0-20, SNOT Sleep score range is 0-20, SNOT Psychological score range is 0-30.

* $P < .0001$.

† $P < .01$.

TABLE E3. Mean ($\pm$ SD) symptom burden for each item on the SNOT-22 at baseline to follow-up and the effect size of each item

Symptoms	Baseline	Follow-up	Effect size (95% CI)
Rhinological symptoms			
Need to blow nose	3.1 $\pm$ 1.1	1.6 $\pm$ 1.4*	1.2 (0.74 to 1.6)
Sneezing	1.9 $\pm$ 1.2	1.1 $\pm$ 0.95*	0.84 (0.43 to 1.2)
Runny nose	2.6 $\pm$ 1.3	1.3 $\pm$ 1.3*	1.1 (0.64 to 1.5)
Cough	1.7 $\pm$ 1.5	1.1 $\pm$ 1.3†	0.39 (0.006 to 0.79)
Postnasal drip	2.8 $\pm$ 1.5	1.8 $\pm$ 1.4*	0.72 (0.32 to 1.1)
Thick nasal discharge	2.9 $\pm$ 1.6	1.3 $\pm$ 1.5*	0.99 (0.57 to 1.4)
Loss of taste/smell	2.2 $\pm$ 1.9	1.3 $\pm$ 1.5‡	0.48 (0.09 to 0.87)
Blockage/congestion of nose	3.6 $\pm$ 1.3	1.8 $\pm$ 1.5*	1.3 (0.90 to 1.8)
Ear/face symptoms			
Ear fullness	1.9 $\pm$ 1.4	1.1 $\pm$ 1.3‡	0.57 (0.17 to 0.96)
Dizziness	0.84 $\pm$ 1.2	0.72 $\pm$ 1.1	0.10 (−0.29 to 0.49)
Ear pain	0.7 $\pm$ 1.0	0.59 $\pm$ 0.94	0.08 (−0.31 to 0.47)
Facial pain/pressure	2.2 $\pm$ 1.7	1.3 $\pm$ 1.4‡	0.54 (0.14 to 0.93)
Sleep function			
Difficulty falling asleep	1.8 $\pm$ 1.5	1.0 $\pm$ 1.3‡	0.56 (0.18 to 0.96)
Waking up at night	1.8 $\pm$ 1.6	1.1 $\pm$ 1.4‡	0.50 (0.10 to 0.89)
Lack of a good nights sleep	2.2 $\pm$ 1.5	1.3 $\pm$ 1.6‡	0.59 (0.19 to 0.98)
Waking up tired	2.6 $\pm$ 1.4	1.6 $\pm$ 1.5*	0.68 (0.28 to 1.1)
Psychological symptoms			
Fatigue	2.5 $\pm$ 1.5	1.5 $\pm$ 1.5*	0.66 (0.26 to 1.1)
Reduced productivity	1.8 $\pm$ 1.6	1.2 $\pm$ 1.5†	0.40 (0.009 to 0.79)
Reduced concentration	1.8 $\pm$ 1.5	1.1 $\pm$ 1.4‡	0.49 (0.10 to 0.89)
Frustrated/restless/irritable	1.6 $\pm$ 1.5	0.88 $\pm$ 1.2‡	0.53 (0.13 to 0.92)
Sad	0.86 $\pm$ 1.3	0.53 $\pm$ 0.90	0.30 (−0.09 to 0.69)
Embarrassed	0.51 $\pm$ 1.0	0.29 $\pm$ 0.61	0.26 (−0.13 to 0.65)

Symptoms were scored 0-5, with 5 representing the most severe.

* $P < .0001$.† $P < .05$.‡ $P < .01$.