



Published in final edited form as:

*JAMA Otolaryngol Head Neck Surg.* 2014 August 1; 140(8): 712–719. doi:10.1001/jamaoto.2014.1045.

## Response shift in quality of life after endoscopic sinus surgery for chronic rhinosinusitis

Adam S. DeConde, MD<sup>1</sup>, Todd E. Bodner, Ph.D<sup>2</sup>, Jess C. Mace, MPH, CCRP<sup>1</sup>, and Timothy L. Smith, MD MPH<sup>1</sup>

<sup>1</sup>Division of Rhinology and Sinus Surgery, Oregon Sinus Center, Department of Otolaryngology – Head and Neck Surgery; Oregon Health & Science University, Portland, Oregon, USA

<sup>2</sup>Department of Psychology, Portland State University, Portland, Oregon, USA

### Abstract

**Importance**—Patient-reported measures are designed to detect a true change in outcome, but they are also subject to change from biases inherent to self-reporting: changing internal standards, changing priorities, and changing interpretations of a given instrument. These biases are collectively known as ‘response shifts’ and can obscure true change after medical interventions.

**Objective**—To determine the presence of response shifts in patients with chronic rhinosinusitis (CRS) after endoscopic sinus surgery.

**Design, Setting, and Participants**—Multisite, prospective, observational cohort study conducted at academic tertiary care centers between February 2011 and May 2013. Study participants comprised a population-based sample of 514 adults (age ≥ 18 years) with CRS, who elected surgical intervention for continuing medically refractory symptoms.

**Intervention**—Endoscopic sinus surgery.

**Main Outcome and Measures**—Preoperative and postoperative data from the 22-item Sinonasal Outcome Test (SNOT-22) survey instrument was characterized using exploratory factor analysis. Subsequent longitudinal structural equation models were estimated to test structure, potential response shifts, and true change in the SNOT-22.

**Results**—A total of 339 participants (66.0%) provided survey evaluations at baseline and 6-month follow-up. Factor analysis of the SNOT-22 revealed 5 correlated, yet distinguishable, underlying factors. Endoscopic sinus surgery had a differential impact across these factors, with the largest effect size in rhinologic symptoms (mean[SD] SNOT-22 score before and after surgery, 13.18[5.11] and 7.37[5.48], respectively;  $d = -1.13$  [ $P < .001$ ] and extranasal rhinologic symptoms (8.31[3.46] and 4.83[3.68], respectively;  $d = -1.00$  [ $p < 0.05$ ]) ( $d$  is an effect size measure defined as

---

**Corresponding Author:** Timothy L. Smith, MD, MPH Oregon Health & Science University Department of Otolaryngology – Head and Neck Surgery Division of Rhinology and Sinus/Skull Base Surgery, Oregon Sinus Center 3181 SW Sam Jackson Park Road, PV-01 Portland, Oregon 97239 FAX: 503-494-4631 smithtim@ohsu.edu.

**Conflict of Interest Disclosures:** None reported

**Authors Contributions:** Adam S. DeConde: Manuscript preparation, study design, intellectual content

Todd E. Bodner: Manuscript preparation, study design, and data analysis, data interpretation and intellectual content

Jess C. Mace: Study coordination, data collection, data analysis, and manuscript preparation

Timothy L. Smith: Study design, administrative support, study supervision, and manuscript preparation.

the difference in means divided by the presurgery SD). Endoscopic sinus surgery had a smaller, yet significant, effect size on the remaining 3 factors: ear/facial symptoms (7.32[4.6] and 3.90[4.1], respectively;  $d = -0.74$ ;  $P < 0.001$ ), psychological dysfunction (11.90[7.21] and 6.50[6.69], respectively;  $d = -0.75$ ;  $P < 0.05$ ), and sleep dysfunction (10.12[5.59] and 5.88[5.37], respectively;  $d = -0.76$ ;  $P < 0.001$ ). Participants were found to undergo recalibration, reprioritization, and reconceptualization of symptoms after intervention; however, the magnitude of these response shifts was small and not clinically significant.

**Conclusions and Relevance**—The SNOT-22 measures 5 distinct factors, not a single construct. Reporting of individual subscale scores may improve sensitivity of this instrument in future studies. Participants undergoing endoscopic sinus surgery experience only clinically insignificant response shifts, validating assessment of change through use of presurgery and postsurgery SNOT-22 responses.

Medical outcomes research is predicated upon interval changes of self-reported quality of life (QOL) after interventions. These patient-reported measures are designed to detect a true change in outcome, but they are also subject to change from the biases inherent to self-reporting: changing internal standards (recalibration), changing priorities (reprioritization) and changing interpretations (reconceptualization) of a given instrument. These 3 unmeasured dynamic internal biases can result in a change in the meaning of the QOL instrument, and this change is termed a *response shift*.<sup>1</sup>

Response shifts are particularly important in health-related QOL studies using repeated measures, where efficacy is determined as the change from a pretreatment baseline after an intervention. The response shift has been identified in a wide range of medical conditions and can both positively and negatively affect the detection of treatment effects.<sup>2-4</sup> Types and magnitudes of response shift are unique to each intervention and disease process. To date, no one has investigated to what degree interval measurements of QOL after endoscopic sinus surgery (ESS) for chronic rhinosinusitis (CRS) reflect a true change in QOL or if they merely reflect a change in the instrument used to make that measurement. For example, theoretically, a patient with longstanding nasal obstruction may adapt to this as their “normal state” and report ‘no problem’ for this question preoperatively, but postoperatively find an unexpected improvement that would still be reported as ‘no problem.’ This hypothetical example illustrates a recalibration response shift that would mask a true change in quality of life. The goal of this analysis was to investigate the direction and magnitude of this response shift in a cohort of patients who underwent ESS for medically refractory CRS through secondary statistical analysis with a previously described and applied technique using confirmatory factor analysis and structural equation modeling (SEM).<sup>5,6</sup>

## METHODS

### Patient Population and Data Collection

Before enrollment, written informed consent was obtained for all participants. The institutional review board at each of the 4 sites monitored and approved all investigational protocols. The institutional review board at Oregon Health & Science University provided comprehensive oversight and review for the entire study as the coordinating center. Adult

patients (age 18 years) with CRS were enrolled into an ongoing prospective, observational cohort investigation utilizing 4 academic, tertiary, rhinology practices (Oregon Health & Science University, Portland; Medical University of South Carolina, Charleston; Stanford University, Palo Alto, California; and University of Calgary, Calgary, Alberta, Canada). Preliminary findings from this cohort study have been previously reported.<sup>7-9</sup> Inclusion criteria consisted of a current diagnosis of symptomatic refractory CRS as defined by the 2007 Adult Sinusitis Guidelines;<sup>10</sup> prior treatment with oral, broad-spectrum, or culture-directed antibiotics (2 weeks); and either topical nasal corticosteroid sprays (3 weeks) or a 5-day trial of systemic steroid therapy. Patients deemed surgical candidates that elected ESS were enrolled and required to complete the Sinonasal Outcome Test (SNOT-22) at both baseline and a 6-month follow-up visit. The SNOT-22 is a 22-item, validated, treatment outcome measure applicable to chronic sinonasal conditions (score range, 0–110).<sup>11</sup> Lower total scores on the SNOT-22 suggest better QOL and symptom severity.

## Analytic Strategy

**Preliminary Analyses and Exploratory Factor Analysis**—Preliminary analyses tested for differences across the 4 surgery locations. Because no significant differences in scale and item scores across locations were found, all reported analyses ignored location. Prior to testing for response shifts, a reasonably well-fitting factor model was needed. This analysis seeks to identify the unique factors, or “constructs,” (ie, aspects of health-related QOL measured that each individual question measures) that the SNOT-22 measures by examining correlating groups of questions. Although exploratory factor analyses have been conducted on the SNOT-20,<sup>12,13</sup> to our knowledge, exploratory factor analysis methods have not been used on the SNOT-22. Thus, analysis began with exploring and testing the SNOT-22 factor structure before surgery, prior to building measurement models as previously described.<sup>12,13</sup>

On defining the underlying factors of the SNOT-22, a series of longitudinal structural equation models were then estimated to evaluate for any changes of this factor structure to clarify response shifts and true change in the SNOT-22. Given the skewed nature of the item response distributions, robust estimation procedures were employed in these models. The use of robust estimation procedures complicates model comparisons using the  $\chi^2(\lambda^2)$  difference test; thus, the recommended procedure based on scaled likelihoods was used for model comparisons.<sup>14</sup> Across models, we followed the 4 steps below for detecting response shifts outlined by Oort and colleagues and described in the following subsections.<sup>5,6</sup> Deviations from the recommended procedures are discussed. Statistical analyses were conducted using SPSS version 22.0 (IBM Corporation), and SEM was conducted using Mplus version 4.2 (Muthén & Muthén).

**Step 1: Establishing a Measurement Model**—Following the recommendation by Oort and colleagues,<sup>5,6</sup> an initial model for the SNOT-22 measurement structure was tested without any across-time parameter constraints. The model (Model L0) specification was based on the results of the exploratory factor analysis model where factor loadings were determined by the primary loading in the presurgery exploratory factor analysis model results. This structure was extended longitudinally to the 6-month postsurgery measurement

occasion. As is typical in longitudinal structural equation models, factors were permitted to correlate over time (eg, the rhinologic symptoms factor before and after surgery) and item residual variances were permitted to correlate over time (eg, the residual variances for item 1 before and after surgery).

**Step 2: Overall Test of Response Shift**—Similar to Oort and colleagues,<sup>5,6</sup> successive models place or release constraints on the factor loadings, means, variances, and correlations over time as well as place constraints on the item intercepts and item residual variances. In this step, to provide an overall test of response shifts, invariance constraints across the 2 time periods were applied to the item intercepts, factor loadings, and the residual variances.

**Step 3: Detection of Types of Response Shifts**—In this step, the invariance constraints were lifted one at a time to test their impact on model fit. Lifted across-time invariance constraints that improve model fit were retained in the final model. Secondly, modification indices were inspected for the postsurgery part of the model to identify potential changes in the measurement structure; such additional factor loadings were retained if their presence increased model fit and made theoretical sense.

**Step 4: Assessment of True Change**—In Step 4, attention turned to changes in factor means and covariances across time. Significant changes in factor means over time indicate true change after measurement error, and changes in the measurement structure over time were accounted for in the prior steps.

## RESULTS

### Patient Enrollment and Clinical Characteristics

Between February 2011 and May 2013, 514 participants who met inclusion criteria and gave informed consent were enrolled into this on-going cohort, among whom 339 (66.0%) had provided both a baseline and 6-month follow-up SNOT-22 survey for analysis. Of the 339 participants (overall mean[SD] age, 51.0[15.0] years), 151 (44.5%) were male, 178 (52.5%) reported a history of sinus surgery, 123 (36.3%) had polyps, 118 (34.8%) had asthma, 124 (36.6%) tested positive for allergies, 58 (17.1%) were depressed, and 29 (8.6%) had aspirin sensitivity.

### Preliminary Analyses and Exploratory Factor Analysis

Descriptive statistics for the SNOT-22 items were calculated before and after surgery (Table 1). For all items, descriptively, item means and SDs decreased from before to after surgery and item positive skew increased from before to after surgery (ie, a mix of positive and negative skew before surgery is uniformly positive after surgery).

Eigenanalysis indicated 5 factors with eigenvalues greater than 1.0. Table 2 provides the estimated factor loadings for each of the SNOT-22 items as well as the factor correlations after Promax rotation using the presurgery data. Two features of this solution warrant discussion. First, the rhinologic symptom items based on earlier research was partitioned into 2 dimensions, the latter we have called extranasal rhinologic symptoms. Second, 4 of the items do not load uniquely onto a single dimension (sneezing, thick nasal discharge,

waking up tired, and fatigue). These 4 items with “cross-loadings” are kept in mind because model modifications are entertained based on the confirmatory factor analysis results.

Scale scores were created and tested for change from before to after surgery. These scale score included the overall SNOT-22 scale score (ie, the sum of responses to the 22 items) and 5 subscale scores corresponding to the subdimensions identified in the exploratory factor analysis. Table 3 presents the results of paired *t* tests comparing mean scale scores from before to after surgery. All scale scores tests indicated significant reductions in symptoms and dysfunctions; inspection of effect sizes indicated that all reductions would be characterized as large in a standardized metric.

### Response Shift Testing

**Step 1: Establishing a Measurement Model**—A summary of the fit of the various models to the data appears in Table 4. Model L0 fit reasonably well ( $\chi^2_{835}=2088.28$ ,  $P<0.001$ , root mean square error of approximation [RMSEA]=0.067, standardized root mean residual [SRMR]=0.064). Given the 4 identified salient cross-loadings in the exploratory factor analysis, we next fit a model (model L1) that permitted these 4 cross-loadings (mentioned in the previous subsections and appearing in Table 4) both before and after surgery. Model L1 also fits reasonably well, ( $\chi^2_{827} = 1844.61$ ,  $P<0.001$ , RMSEA=0.060, SRMR=0.060), with an RMSEA value approaching 0.05, which is indicative of a “close fit.” Furthermore, model L1 fit significantly better than model L0 ( $\chi^2_8=87.61$ ,  $P<0.001$ ). These results confirm the SNOT-22’s correlated 5-factor structure; model L1 serves as the base model for tests of response shifts in the subsequent step.

**Step 2 : Overall Test of Response Shift**—Invariance constraints across the 2 periods were applied to the item intercepts, factor loadings, and the residual variances. A significant reduction in model fit between this model and the model L1 from step 1 indicates the presence of some type of response shift. However, given the nature of the SNOT-22 data, we deviate slightly from the recommendation by Oort and colleagues.<sup>6</sup> Inspection of the estimated residual variances before and after surgery in model L1 indicated that 21 of the 22 residual variances were smaller after surgery than before, with a mean reduction of 30%. Although this pattern could be indicative of a nonuniform recalibration response shift per Oort and colleagues,<sup>6</sup> we believe this reduction may be due in part to the positively skewed nature of item responses after surgery caused by floor effects on the response scale (ie, many more scores of 0 indicating “no problem”). From this perspective, any reduction in symptoms after surgery would reduce the mean item response as well as the variance of the item response as observed (Table 1). Thus, we do not apply invariance constraints on the residual variances and therefore will not test for nonuniform recalibration response shifts.

Table 4 presents the fit of a longitudinal structural equation model with invariance constraints across time on the item intercepts and factor loadings (ie, model L2). The fit of model L2 might be considered adequate ( $\chi^2_{870}=2023.67$ ,  $P<0.001$ , RMSEA=0.063, SRMR=0.068). However, this model fits significantly worse than model L1 ( $\chi^2_{43}=87.74$ ,  $P<0.001$ ). This evidence suggests that some of these invariance constraints are reducing

model fit and suggest the presence of response shifts; the change in the RMSEA and SRMR fit indices suggest, however, that magnitude of these response shifts may be small.

**Step 3: Detection of Types of Response Shifts**—Because the first part of Step 3 is tedious, releasing 48 individual invariance constraints separately, we present the end result of this process as model L3. Model L3 released the across-time factor loading invariance constraints on items 8, 9, and 10 for the ear/facial symptoms factor and on item 22 for the rhinologic symptoms factor and the across-time item intercept invariance constraints for items 1, 9, 10, and 22. Model L3 fit significantly better than model L2 ( $\chi^2_8=79.03$ ,  $P<0.001$ ). Despite a large number of across-time invariance constraints, model L3 did not fit significantly worse than model L1 ( $\chi^2_{35}=23.22$ ,  $P=0.94$ ). Thus, any response shifts appear to be localized to these particular items. Inspection of the model modification indices for model L3 suggested the addition of only 1 additional factor loading after surgery, ie, item 18 loading onto extranasal rhinologic symptoms. Model L4, which specified this additional factor loading, fit the data reasonably well ( $\chi^2_{861}=1871.35$ ,  $P<0.001$ , RMSEA=0.059, SRMR=0.062), and significantly better than model L3 ( $\chi^2_1=7.97$ ,  $p=0.005$ ). From Table 4, the model Bayesian information criteria (BICs), which balance model fit and parsimony, support the superiority of model L4. Tables 5 and 6 provide select model parameters from model L4.

**Reconceptualization:** We next turn to an interpretation of the 9 identified response shifts in model L4. Any shifts in the factor loading *patterns* from before to after surgery indicate a reconceptualization in the underlying factors. The addition of item 18 (ie, frustrated / restless / irritable) onto the extranasal rhinologic symptoms factor after surgery suggests that responses to this item are affected by one's standing on this factor, unlike before surgery, perhaps indicating differing levels of postsurgery frustration for this factor. A comparison of this factor loading before ( $\lambda=0.00$ ) and after ( $\lambda=0.20$ ) surgery in Table 5, however, suggests that the level of reconceptualization is small in magnitude.

**Reprioritization:** Shifts in the *magnitude* of factor loadings over time indicate a reprioritization of the importance of that item as it relates to the underlying factor. Items 10 (facial pain/pressure;  $\lambda=0.87$  and  $\lambda=1.01$  before and after surgery, respectively) and 22 (blockage/congestion of nose;  $\lambda=0.87$  and  $\lambda=1.26$  before and after surgery, respectively) demonstrated increases in the size of the factor loadings over time onto the ear/facial symptoms and rhinologic symptoms factors, respectively, suggesting that these items more strongly indicate their underlying factors after surgery. In contrast, items 8 (dizziness;  $\lambda=1.00$  and  $\lambda=0.74$  before and after surgery, respectively) and 9 (ear pain;  $\lambda=1.12$  and  $\lambda=0.89$  before and after surgery, respectively) demonstrated decreases in the size of the factor loadings over time onto the ear/facial symptoms factor, suggesting that these items less strongly indicate their underlying factors after surgery.

**Recalibration:** Shifts in item intercepts across time indicate a (uniform) recalibration of item responses relative to the underlying factors. Items 1 (need to blow nose;  $\tau=2.72$  and  $\tau=2.80$  before and after surgery, respectively) and 9 (ear pain;  $\tau=1.28$  and  $\tau_{\text{post}}=1.39$  before and after surgery, respectively) demonstrated increases in the item intercepts after surgery,

indicating that patients rate these symptoms as more problematic, relative to before surgery and on average, than implied by their standing on the underlying factors. In contrast, items 10 (facial pain/pressure;  $\tau=2.53$  and  $\tau=2.05$  before and after surgery, respectively) and 22 (blockage/congestion of nose;  $\tau=3.50$  and  $\tau=3.09$  before and after surgery, respectively) demonstrated decreases in item intercepts after surgery, indicating that patients rate these symptoms as less problematic, relative to before surgery and on average, than implied by their standing on the underlying factors. Given item response options ranging from 0 to 5, these item intercepts shifts may be considered small.

**Step 4: Assessment of True Change**—Model L5 specifies invariance constraints on the factor means from before to after surgery. Model L5 fits significantly worse than model L4 ( $\chi^2_5=66.60$ ,  $p<0.001$ ). Follow-up tests indicate that all 5 factor means differ from before to after surgery. Table 6 presents the factor means before and after surgery from model L4. Given the scaling of the factor variances to 1.00, the postsurgery means are equivalent to a standardized mean difference in the factor means. The standardized mean differences are similar to those reported using observed scale scores in Table 2. Inspection of the factor correlations before and after surgery in Table 6 suggests that the correlations among factors are stronger after surgery than before. Model L6 specifies across-time invariance constraints on the within-time factor correlations. Model L6 fits significantly worse than model L4 ( $\chi^2_{10}=42.38$ ,  $P<0.001$ ), confirming this descriptive comparison. Thus, the factors underlying the SNOT-22 appear to represent a more unitary set of symptoms and dysfunctions after surgery.

## DISCUSSION

Accurate and sensitive measures of how interventions affect QOL are critically important for our subspecialty. The rationing of national healthcare resources is inevitable, and QOL measures are already used by the National Health Service in the United Kingdom and the recent Patient Protection and Affordable Care Act in the United States invests in comparative clinical outcomes research. Accurately capturing the impact of an intervention on our patients will be essential in guiding individual and societal decisions on the value of any given intervention. Establishing to what extent a response shift plays a role in a given intervention may preserve the value of an intervention or accurately guide us to another, more effective treatment.

Response shifts have the potential to misinform comparative clinical outcomes research. For example, in edentulous patients, response shift completely masks improvement in QOL<sup>2</sup> after denture rehabilitation. Patients undergoing cholecystectomy have greater improvements in gastrointestinal QOL when response shift is considered.<sup>3</sup> A psychosocial intervention for cancer survivors appeared to worsen QOL based on change in pretreatment baseline, but evaluation of the response shift in fact demonstrated a positive effect that was not identified by a recalibration of the QOL towards that of healthy controls.<sup>4</sup>

A range of methods for detecting and quantifying response shifts has been described. In general these methods take 1 of 3 forms: (1) additional administration of a test questionnaire to retrospectively evaluate baseline (eg, a then-test), (2) additional evaluation of the target

outcome (eg, interviews, direct assessments of values or preferences), and (3) post hoc statistical analysis.<sup>4</sup> Retrospective analysis of baselines (ie, then-tests) are limited by recall bias, and may be confounded by alternative explanations such as implicit theories of change, that is, patients suffered through an intervention and therefore are invested in its success and recall an artificially worsened baseline.<sup>15</sup> Additional evaluation of the target outcome is labor intensive and not feasible on data sets already collected. Statistical methods of detecting a response shift can be applied post hoc to data, requires no additional measurements, and only requires a minimum of 2 longitudinal time points (eg, baseline and post-treatment scores).

The research in the present article makes several contributions to the literature on outcomes of rhinologic surgery. To our knowledge, this is the first article testing the factor structure of the SNOT-22 in a confirmatory manner. The results of this analysis indicate 5 correlated yet distinguishable factors underlying the SNOT-22, providing a new level of discrimination to this instrument. The longitudinal structural equation models testing for response shifts did indeed find evidence of response shifts; however, the magnitude of these shifts may be considered small and unimportant for clinical practice. Perhaps most important is the finding of the invariance of most model parameters from before to after surgery. This result provides statistical and measurement evidence validating the comparison of SNOT-22 item responses or scale scores before and after surgery to quantify changes in symptoms and dysfunctions. Had larger degrees of invariance been found, the factors underlying the SNOT-22 before and after surgery would have had differing meanings and interpretations making any assessment of change questionable.

Detection of 5 distinguishable factors of the SNOT-22 offers a new resolution to this instrument and has potential to better characterize the impacts of interventions and comorbidities on CRS in future studies. Prior factor analysis of the SNOT-20 revealed 4 separate constructs: rhinologic symptoms, ear/facial symptoms, sleep function, and psychological function,<sup>12,13</sup> but the present study reveals a fifth construct that uniquely captures “cough” and “post-nasal discharge”. Total SNOT-22 scores are often used to investigate the impact of interventions across populations, but aggregate scores lack the resolution of reporting of the individual domains identified in the present study. For example, aggregate scores cannot detect symmetrically divergent changes in separate domains of health. Similarly, aggregate QOL scores are an abstract concept, whereas patients and clinicians are faced with specific symptoms that they are attempting to improve. Knowledge of what domains are captured by the SNOT-22 and how these domains are changed by ESS will aid in patient-oriented clinical decision-making. Further investigation using these domains could help explain the clinical endpoints achieved by patients with comorbid depression, fibromyalgia and migraine.<sup>16–18</sup> Patients with these comorbidities experience comparable overall gains to the general population but have diminished baselines and postoperative QOL. The present factor analysis provides tools to further investigate the prior observation that comorbid depression, fibromyalgia, and migraine impact SNOT-22 pretreatment and posttreatment QOL measurements.

There are several important limitations to the present study. By using SEM to detect a response shift, these results can only be applied at the population level. Conceivably



individuals may undergo equal and opposite response shifts that would not be detected at this population level. Similarly, unless a significant portion of a study population experiences a response shift it may not appear in a model as the response shift is averaged across the group.<sup>5,19</sup> This limitation could be addressed through future studies employing another method to detect response shifts, such as a then-test, allowing for cross-validation of these results or through a subgroup analysis of participants with different types of CRS or other comorbidities. Another concern is that our sample size was not adequate to detect response shifts; however, some conventional guidelines are available to help make this determination. Kline<sup>20</sup> summarizes research on SEM practices and notes that a typical sample size is around 200 participants.<sup>20</sup> Thus, our sample size of 338 would be considered larger than average against this benchmark. Outside of SEM, population surveys typically consist of around 1000 participants to represent populations of 100 million people (eg, the population of registered voters who intend to vote in a US presidential election) with great success. Because the population of patients experiencing rhinological symptoms warranting rhinological surgery is far less than this number, we find some solace in the size of our sample. Finally, we report BIC values in our Table 4. As noted in the article, BICs permit model comparisons that balance the fit of the model with the complexity of the model (ie, all else being equal, complex models tend to fit better). Unfortunately, more complex models also tend to replicate, generalize, and cross-validate less well. Our favored model (model 4) has the lowest BIC value. Importantly, another common fit statistic, the expected cross validation index (ECVI), preserves the ordering of model favorability based on the BIC values. Thus, our model with the best BIC value also is the best model with respect to ECVI and therefore expected degree of cross validation and generalizability. Thus, although a larger sample size is always appreciated, we are cautiously optimistic about the generalizability of our results.

## CONCLUSION

The results of this analysis identified 5 correlated but distinguishable factors underlying the SNOT-22, which carries important implications for future QOL outcomes research in CRS. The longitudinal structural equation models testing for response shifts reveals response shifts; however, the magnitude of these shifts may be considered small and unimportant for clinical practice. This result provides statistical and measurement evidence validating the comparison of SNOT-22 item responses or scale scores before and after surgery to quantify changes in symptoms and dysfunctions.

## Acknowledgments

**Adam S. DeConde, MD** and **Todd E. Bodner, PhD** had full access to all the data in the study and take responsibility for the integrity of the data and accuracy of the data analysis. **Adam S. DeConde, MD** was responsible for manuscript preparation and study design. **Todd E. Bodner, PhD** contributed to manuscript preparation, study design, and data analysis, data interpretation and intellectual content. **Jess C. Mace, MPH, CCRP** contributed to study coordination, data collection, data analysis, and manuscript preparation. **Timothy L. Smith, MD, MPH**, provided study design, administrative support, study supervision, and manuscript preparation.

This study was supported 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/PD: TL Smith). This funding organization did not contribute to the design or conduct of this study; collection, management, analysis, or interpretation of the data; preparation, review, approval or decision to submit this manuscript for publication. This

study is registered in a public trials registry (<http://www.clinicaltrials.gov>) ID# NCT01332136. Timothy L. Smith, MD, MPH and Jess C. Mace, MPH, CCRP receive partial support from the NIDCD. Timothy L. Smith, MD, MPH is also a consultant for IntersectENT, Inc. (Menlo Park, CA) which is not affiliated with this investigation. Todd E. Bodner, PhD is supported by grants from the National Institute of Child Health & Human Development, the National Heart, Lung, and Blood Institute / Kaiser Permanente, the National Institute for Occupational Safety and Health, and the U.S. Department of Defense, none of which are associated with funding or publishing this study.

**Financial Disclosures:** This study was partially supported by a grant from the National Institute on Deafness and other Communication Disorders (NIDCD), one of the National Institutes of Health, Bethesda, Maryland, USA. (RO1 DC005805; **PI/PD:** TL Smith).

**Funding/support:** Timothy L. Smith, MD, MPH and Jess C. Mace, MPH, CCRP are supported by a grant from NIDCD. Timothy L. Smith, MD, MPH is also a consultant for IntersectENT, Inc. (Menlo Park, CA) which is not affiliated with this investigation. Todd E. Bodner, PhD is supported by grants from the National Institute of Child Health & Human Development, the National Heart, Lung, and Blood Institute/Kaiser Permanente, the National Institute for Occupational Safety and Health, and the U.S. Department of Defense, none of which are associated with funding for this study.

**Trial Registration:** (<http://www.clinicaltrials.gov>) ID# NCT01332136.

## REFERENCES

1. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med*. 1999; 48(11):1507–1515. [PubMed: 10400253]
2. Ring L, Höfer S, Heuston F, Harris D, O'Boyle CA. Response shift masks the treatment impact on patient reported outcomes (PROs): the example of individual quality of life in edentulous patients. *Health Qual Life Outcomes*. 2005; 3:55. [PubMed: 16146573]
3. Shi HY, Lee KT, Lee HH, Uen YH, Chiu CC. Response shift effect on gastrointestinal quality of life index after laparoscopic cholecystectomy. *Qual Life Res*. 2011; 20(3):335–341. [PubMed: 20936506]
4. Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med*. 1999; 48(11):1531–1548. [PubMed: 10400255]
5. Oort FJ. Using structural equation modeling to detect response shifts and true change. *Qual Life Res*. 2005; 14(3):587–598. [PubMed: 16022054]
6. Oort FJ, Visser MR, Sprangers MA. An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. *Qual Life Res*. 2005; 14(3):599–609. [PubMed: 16022055]
7. Alt JA, Mace JC, Buniel MCF, Soler ZM, Smith TL. Predictors of olfactory dysfunction in rhinosinusitis using the brief smell identification test. *Laryngoscope*. 2014; 124(7):E259–266. [PubMed: 24402746]
8. Alt JA, Smith TL, Mace JC, Soler ZM. Sleep quality and disease severity in patients with chronic rhinosinusitis. *Laryngoscope*. 2013; 123(10):2364–2370. [PubMed: 23918740]
9. Soler ZM, Rudmik L, Hwang PH, Mace JC, Schlosser RJ, Smith TL. Patient-centered decision making in the treatment of chronic rhinosinusitis. *Laryngoscope*. 2013; 123(10):2341–2346. [PubMed: 23856802]
10. Rosenfeld RM, Andes D, Bhattacharyya N, et al. Clinical practice guideline: Adult sinusitis. *Otolaryngol Head Neck Surg*. 2007; 137(3 suppl):S1–S31. [PubMed: 17761281]
11. Hopkins C, Gillett S, Slack R, Lund VJ, Browne JP. Psychometric validity of the 22-item Sinonasal Outcome Test. *Clin Otolaryngol*. 2009; 34(5):447–454. [PubMed: 19793277]
12. 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(5):736–741. [PubMed: 17478207]
13. Pynnonen MA, Kim HM, Terrell JE. Validation of the Sino-Nasal Outcome Test 20 (SNOT-20) domains in nonsurgical patients. *Am J Rhinol Allergy*. 2009; 23(1):40–45. [PubMed: 19379611]
14. Satorra A, Bentler PM. Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*. 2010; 75:243–248. [PubMed: 20640194]

15. Norman G. Hi! How are you? Response shift, implicit theories and differing epistemologies. *Qual Life Res.* 2003; 12(3):239–249. [PubMed: 12769136]
16. Mace JC, Michael YL, Carlson NE, Litvack JR, Smith TL. Effects of depression on quality of life improvement after endoscopic sinus surgery. *Laryngoscope.* 2008; 118(3):528–534. [PubMed: 18043488]
17. Soler ZM, Mace J, Smith TL. Fibromyalgia and chronic rhinosinusitis: outcomes after endoscopic sinus surgery. *Am J Rhinol.* 2008; 22(4):427–432. [PubMed: 18702911]
18. DeConde AS, Mace JC, Smith TL. The impact of comorbid migraine on quality of life outcomes after endoscopic sinus surgery. *Laryngoscope.* 2014; 124(8):1750–1755. [PubMed: 24431279]
19. Ahmed S, Mayo NE, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Change in quality of life of people with stroke over time: true change or response shift? *Qual Life Res.* 2005; 14(3): 611–627. [PubMed: 16022056]
20. Kline, RB. *Principles and practices of structural equation modeling.* 3rd ed.. Guilford Press; New York City, New York: 2010.

**Table 1**

Descriptive Statistics for SNOT-22 Items Before and After Surgery

SNOT-22 Item	Mean		SD		Skew	
	Before	After	Before	After	Before	After
1. Need to blow nose	2.72	1.55	1.36	1.29	-0.40	0.39
2. Sneezing	1.76	1.02	1.29	1.15	0.33	0.87
3. Runny nose	2.45	1.32	1.39	1.26	-0.10	0.68
4. Cough	2.10	1.24	1.52	1.37	0.12	0.88
5. Post nasal discharge	3.19	1.96	1.41	1.42	-0.67	0.24
6. Thick nasal discharge	3.02	1.63	1.48	1.59	-0.56	0.53
7. Ear fullness	2.24	1.28	1.55	1.39	-0.02	0.81
8. Dizziness	1.29	0.72	1.40	1.13	0.70	1.63
9. Ear pain	1.27	0.67	1.41	1.10	0.90	1.73
10. Facial pain/ pressure	2.52	1.23	1.55	1.42	-0.29	0.85
11. Difficulty falling asleep	2.05	1.15	1.68	1.49	0.16	1.11
12. Waking up at night	2.53	1.45	1.56	1.43	-0.16	0.67
13. Lack of a good night's sleep	2.74	1.58	1.55	1.51	-0.30	0.64
14. Waking up tired	2.81	1.70	1.50	1.49	-0.35	0.54
15. Fatigue	2.76	1.68	1.51	1.50	-0.34	0.50
16. Reduced productivity	2.36	1.28	1.55	1.44	-0.08	0.90
17. Reduced concentration	2.24	1.19	1.50	1.40	0.04	0.94
18. Frustrated / restless / irritable	2.23	1.15	1.48	1.34	0.01	0.96
19. Sad	1.27	0.68	1.41	1.11	0.83	1.81
20. Embarrassed	1.03	0.51	1.36	1.05	1.19	2.28
21. Sense of smell / taste	2.76	1.73	1.77	1.73	-0.25	0.61
22. Blockage / congestion of nose	3.50	1.74	1.34	1.47	-0.93	0.51

N = 339. SNOT-22, Sinonasal Outcome Test. SD, standard deviation. Response scale: 0 = No Problem, 1 = Very Mild Problem, 2 = Mild or Slight Problem, 3 = Moderate Problem, 4 = Severe Problem, 5 = Problem as Bad as It Can Be.

**Table 2**

Exploratory Factor Analysis Promax-Rotated Standardized Factor Loadings and Factor Correlations of SNOT-22 Items at Baseline

	Rhinologic Symptoms	Extra-nasal Rhinologic Symptoms	Ear/Facial Symptoms	Psychological Dysfunction	Sleep Dysfunction
<u>SNOT-22 Items</u>					
1. Need to blow nose	<b>0.86</b>	0.07	-0.12	0.02	-0.02
2. Sneezing *	<b>0.46</b>	-0.05	<b>0.37</b>	-0.22	0.01
3. Runny nose	<b>0.77</b>	0.04	0.04	-0.14	0.00
4. Cough	0.09	<b>0.45</b>	0.04	-0.09	0.01
5. Post nasal discharge	-0.01	<b>0.88</b>	0.07	0.02	0.00
6. Thick nasal discharge *	<b>0.37</b>	<b>0.43</b>	-0.14	0.12	-0.03
7. Ear fullness	-0.02	0.17	<b>0.71</b>	0.02	-0.03
8. Dizziness	0.04	-0.05		0.23	0.01
9. Ear pain	-0.10	-0.03	<b>0.92</b>	0.00	0.01
10. Facial pain/ pressure	-0.04	0.06	<b>0.30</b>	0.25	0.14
11. Difficulty falling asleep	0.06	-0.01	0.08	0.09	<b>0.64</b>
12. Waking up at night	0.00	0.04	-0.01	-0.03	<b>0.86</b>
13. Lack of a good night's sleep	-0.01	-0.02	-0.02	0.04	<b>0.95</b>
14. Waking up tired *	0.05	-0.06	0.00	<b>0.38</b>	<b>0.60</b>
15. Fatigue *	-0.05	0.06	-0.02	<b>0.60</b>	<b>0.38</b>
16. Reduced productivity	-0.07	0.07	-0.08	<b>0.88</b>	0.09
17. Reduced concentration	-0.05	0.01	0.02	<b>0.81</b>	0.09
18. Frustrated / restless / irritable	0.03	-0.02	0.04	<b>0.75</b>	0.09
19. Sad	0.09	-0.06	0.07	<b>0.74</b>	-0.12
20. Embarrassed	0.14	0.00	0.16	<b>0.48</b>	-0.14
21. Sense of smell / taste	<b>0.50</b>	-0.15	-0.01	0.05	0.01
22. Blockage / congestion of nose	<b>0.57</b>	0.02	-0.09	0.15	0.07
<u>SNOT-22 Factor Correlations</u>					
Rhinologic Symptoms	1.00	...	...	...	...
Ext. Rhinologic Symptoms	.48	1.00	...	...	...
Ear/Facial Symptoms	.45	.29	1.00	...	...
Psychological Dysfunction	.30	.27	.39	1.00	...
Sleep Dysfunction	.43	.30	.57	.60	1.00

\* Item does not load clearly on a single dimension. SNOT-22, Sinonasal Outcome Test. Boldface factor loadings indicate salient loading at or exceeding 0.30; ellipses represent duplicate discrete pairings.

**Table 3**

Mean SNOT-22 Scale and Subscale Scores Before and After Surgery

SNOT-22 Scale Scores	Pre-Surgery		Post-Surgery		<i>t</i> (338)	<i>d</i>
	Mean	SD	Mean	SD		
Overall Scale Score	50.83	19.61	28.48	21.00	-17.51*	-1.14
Rhinologic Symptoms	13.18	5.11	7.37	5.48	-16.67*	-1.13
Extra-nasal Rhinologic Symptoms	8.31	3.46	4.83	3.68	-15.49*	-1.00
Ear/Facial Symptoms	7.32	4.63	3.90	4.07	-13.77*	-0.74
Psychological Dysfunction	11.90	7.21	6.50	6.69	-13.35*	-0.75
Sleep Dysfunction	10.12	5.59	5.88	5.37	-13.38*	-0.76

\*  $p < 0.050$ ; *d* is an effect size measure defined as the difference in means divided by the presurgery standard deviation; subscale scores based on primary item loadings based on exploratory factor analysis results. SNOT-22, Sinonasal Outcome Test. SD, standard deviation, *t*, matched-paired *t*-test statistic.

**Table 4**

Fit of Longitudinal Structural Equation Models of the SNOT-22 under Robust Maximum Likelihood Estimation

<b>Model:</b>	$\chi^2$	df	RMSEA	SRMR	BIC
<b>L0:</b> 5 Factors, no cross-loadings, no constraints from preSurgery EFA	2088.28	835	0.067	0.064	43,272
<b>L1:</b> L0 with cross-loadings identified in presurgery EFA	1844.613*	827	0.060	0.060	43,942
<b>L2:</b> L1 with across-time item intercept and factor loading invariance constraints	2023.67*	870	0.063	0.068	42,988
<b>L3:</b> L2 with 4 factor loadings and 4 item intercepts invariance constraints released	1893.64*	862	0.059	0.062	42,892
<b>L4:</b> L3 permitting a factor loading of item 18 on extranasal rhinologic symptoms	1871.35*	861	0.059	0.062	42,875
<b>L5:</b> L4 with factor mean invariance constraints	2076.27*	866	0.064	0.130	43,065
<b>L6:</b> L4 with within-time factor covariance invariance constraints across time	1928.34*	871	0.060	0.071	42,881

\*  $P < 0.050$ .  $N = 339$ . SNOT-22, Sinonasal Outcome Test, df, degrees of freedom. RMSEA, root mean square error of approximation. SRMR, standardized root mean residual. BIC, Bayesian information criteria. EFA, exploratory factor analysis.

**Table 5**  
Model L4 Item Intercepts and Unstandardized Factor Loadings Before and After Surgery

SNOT-22 Items	Factors											
	Item Intercepts		Rhinologic Symptoms		Extra-nasal Rhinologic Symptoms		Ear/Facial Symptoms		Psychological Dysfunction		Sleep Dysfunction	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
1. Need to blow nose	2.72	2.80	1.11	1.11	...	...	...	...	...	...	...	...
2. Sneezing	1.80	1.80	0.55	0.55	...	0.23	0.23	0.23	...	...	...	...
3. Runny nose	2.45	2.45	1.01	1.01	...	...	...	...	...	...	...	...
4. Cough	2.06	2.06	...	...	0.81	0.81	...	...	...	...	...	...
5. Post nasal discharge	3.21	3.21	...	...	1.28	1.28	...	...	...	...	...	...
6. Thick nasal discharge	3.02	3.02	0.59	0.59	0.74	0.74	...	...	...	...	...	...
7. Ear fullness	2.23	2.23	...	...	...	...	1.16	1.16	...	...	...	...
8. Dizziness	1.31	1.31	...	...	...	...	1.00	0.74	...	...	...	...
9. Ear pain	1.28	1.39	...	...	...	...	1.12	0.89	...	...	...	...
10. Facial pain/ pressure	2.53	2.05	...	...	...	...	0.87	1.01	...	...	...	...
11. Difficulty falling asleep	2.06	2.06	...	...	...	...	...	...	...	...	1.16	1.16
12. Waking up at night	2.51	2.51	...	...	...	...	...	...	...	...	1.31	1.31
13. Lack of a good night's sleep	2.75	2.75	...	...	...	...	...	...	...	...	1.49	1.49
14. Waking up tired	2.82	2.82	...	...	...	...	...	...	0.52	0.52	0.90	0.90
15. Fatigue	2.78	2.78	...	...	...	...	...	...	0.94	0.94	0.47	0.47
16. Reduced productivity	2.37	2.37	...	...	...	...	...	...	1.37	1.37	...	...
17. Reduced concentration	2.24	2.24	...	...	...	...	...	...	1.32	1.32	...	...
18. Frustrated / restless / irritable	2.23	2.23	...	...	...	0.20	...	...	1.12	1.12	...	...
19. Sad	1.31	1.31	...	...	...	...	...	...	0.82	0.82	...	...
20. Embarrassed	1.00	1.00	...	...	...	...	...	...	0.60	0.60	...	...
21. Sense of smell / taste	2.73	2.73	0.88	0.88	...	...	...	...	...	...	...	...
22. Blockage / congestion of nose	3.50	3.09	0.89	1.21	...	...	...	...	...	...	...	...

Notes: Differences between before and after surgery in boldface are significant parameters. Factor variances equals 1. SNOT-22, Sinonasal Outcome Test. Pre, before surgery. Post, after surgery. Ellipses indicate factor loadings of 0.



**Table 6**

**Model L4 Factor Means and Correlations Before and After Surgery**

Factors	Mean	Correlations										
		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	
1. Rhinologic Symptoms (Pre)	0.00	1.00	...	...	...	...	...	...	...	...	...	...
2. Extra-nasal Rhinologic Symptoms (Pre)	0.00	.51	1.00	...	...	...	...	...	...	...	...	...
3. Ear/Facial Symptoms (Pre)	0.00	.39	.38	1.00	...	...	...	...	...	...	...	...
4. Psychological Dysfunction (Pre)	0.00	.39	.34	.63	1.00	...	...	...	...	...	...	...
5. Sleep Dysfunction (Pre)	0.00	.27	.29	.43	.62	1.00	...	...	...	...	...	...
6. Rhinologic Symptoms (Post)	-1.12	<b>.22</b>	.02	.16	.14	.10	1.00	...	...	...	...	...
7. Extra-nasal Rhinologic Symptoms (Post)	-0.79	.15	<b>.37</b>	.18	.15	.12	.78	1.00	...	...	...	...
8. Ear/Facial Symptoms (Post)	-0.81	.02	.03	<b>.51</b>	.24	.13	.67	.57	1.00	...	...	...
9. Psychological Dysfunction (Post)	-0.79	.06	.01	.24	<b>.43</b>	.22	.60	.51	.64	1.00	...	...
10. Sleep Dysfunction (Post)	-0.79	.03	.03	.20	.34	<b>.42</b>	.53	.51	.53	.77	1.00	...

Notes: Factor variance set to 1.00 in model. Factor correlations greater than .11 (in absolute value) are significant at P < 0.05. Correlations of the same factor before and after surgery in boldface; across-time factor correlations shaded. Ellipses represent duplicate discrete pairings.