Validation of the Sinonasal Outcome Test-16 (SNOT-16) in patients with acute bacterial sinusitis

Andrew Lloyd DPhil [1], Kerstin Uhl-Hochgräber PhD [2]; Jay Piccirillo, MD, FACS [3]

1. Oxford Outcomes Ltd, Oxford, United Kingdom
2. Bayer Healthcare Pharmaceuticals, Berlin, Germany
3. Department of Otolaryngology, Washington University School of Medicine, St. Louis, MO, USA

INTRODUCTION

Sinusitis encompasses a spectrum of acute and chronic, neutrophilic and eosinophilic, non-allergic and allergic inflammatory processes. Sinusitis affects approximately 16% of the adult population, and is responsible for nearly $5.8 billion in health care costs annually (Slavin et al., 1995). Acute sinusitis symptoms last as long as 4 weeks and is considered a bacterial sinusitis (or rhinosinusitis) when the inflammation of the paranasal sinus mucosa is caused by bacterial overgrowth in a closed cavity (Poole, 1999). The most prominent symptoms of acute sinusitis include headache, nasal congestion, facial (and dental) pain, purulent rhinorrhea, post-nasal drainage and cough.

The assessment of the effectiveness of treatment for patients with sinusitis generally, has been hindered by the lack of valid and reliable patient reported outcome measures (Morley and Sharp, 2006). Some self complete measures do exist including the Rhinosinusitis Outcome Measure (RSOM-31) (Piccirillo et al., 1995), the Sinonasal Outcome Test -20 (SNOT-20) (Piccirillo et al., 2002), and the SNOT-16 (Anderson et al., 1999).

Here we report secondary analysis of trial data to explore the psychometric properties of the SNOT-16. The trial was designed to evaluate the efficacy and safety of moxifloxacin once-daily versus placebo in the treatment of acute bacterial sinusitis in patients with mild, moderate and severe disease.

METHODS

Patients (n=374) with acute bacterial sinusitis received either placebo, or 400mg moxifloxacin once daily (1:2 ratio) over a period of 5 days. Patients were then reviewed at test-of-cure (TOC) and follow-up.

Study Population & Assessments

All psychometric analyses were conducted on a combined sample of placebo and active treatment patients. Data from the following patient reported outcomes (PROs) were included in the analyses.

The Activity Impairment Assessment: The AIA was developed based on an existing work-productivity measure, the Stanford Presenteeism Scale (Koopman et al., 2002). It was designed to evaluate the impact of health problems on individual performance and productivity and other activities including social activities (Wild et al 2005).

Rand SF-36 Item Health Survey 1.0: The Rand SF-36 is similar to the Medical Outcomes Study Short Form 36 (Hays et al., 1993), and includes the same dimensions. The Rand SF-36 was selected as the anchor measure for this study because it is possible to use a 24 hour recall period.

SNOT-16: This 16-item measure assesses rhinosinusitis symptoms. The measure has been validated in a population of chronic rhinosinusitis patients (Anderson et al. 1999) The symptoms and associated problems listed in the SNOT-16 are consistent with those listed in the treatment guidelines developed by the Sinus and Allergy Health Partnership (SAHP) for acute bacterial sinusitis.

In addition a Global Rating of Change question was included.

Statistical Analysis

Data were assessed using standard psychometric tests and criteria to evaluate the item performance, reliability, validity and minimal important difference (MID) of the scales. All analyses (unless otherwise stated) were conducted on trial data imputed using last observation carried forward. Analyses were conducted to explore the following issues

- Data distribution: The distribution and missingness patterns of the data were examined at baseline to identify ceiling and floor effects and whether the scales are normally distributed.
- Internal Consistency Reliability: Calculated overall using Cronbach's alpha, with a value of 0.70 set as the benchmark for declaring the scale as internally consistent (Nunnally, 1978).
- Construct Validity: Convergent validity will be supported if correlation coefficients between the SNOT-16, AIA and Rand SF-36 fall between 0.40 and 0.70.
- Responsiveness/ sensitivity: The responsiveness of the SNOT-16 to change in health were assessed by comparing baseline data with TOC data. Paired sample t-tests were conducted, and the effect size and standardised response means statistics were calculated.
- Minimal Important Difference (MID): MID was estimated using distribution and anchor based methods. Anchor based MID estimates were derived from the smallest detectable change on the Global Rating of Change (e.g. "a little better") is referred to as the anchor method. Distribution based methods offer an alternative to anchor-based methods and rely on expressing an effect in terms of the underlying distribution of the results. The Standard Error of Measurement (SEM) and ½ a standard deviation are widely used methods for estimating MID (Guyatt et al., 2002; Revicki et al, 2007). All three estimates were considered in order to settle on a single value, with most weight given to the anchor based methods (Revicki et al., 2007).

RESULTS

Participant characteristics

Most participants (65.5%) were women and middle aged (mean age =40.2; std dev=13.5). A high proportion were White Caucasian (70.9%); followed by Hispanic (17.9%); and African American (8.3%).

Distributional characteristics

There is no evidence of floor or ceiling effects in the SNOT-16 data collected at baseline. There was very little (~4%) missing data.

Internal Consistency Reliability

Internal consistency for the SNOT-16 assessed at baseline was high (Cronbach's $\alpha = 0.874$).

Construct Validity

The SNOT-16 was significantly correlated with dimensions on the SF-36. Correlations greater than 0.40 were considered important a priori. The SNOT-16 was associated (r>0.40) with dimensions of role function (physical); vitality; social functioning and bodily pain, and as predicted a priori in the SAP had a low association with general health. Additionally the SNOT-16 showed a high correlation with the total AIA score (r=0.673).

Correlation between Rand SF-36 and SNOT-16 at baseline as a test of construct validity.

Baseline SF-36	SNOT-16
Physical function	-.272**
Role function physical	**-.443****
Role function emotional	-.320**
Energy/ vitality	**-.448****
Mental Health	-.311**
Social functioning	**-.486****
Bodily Pain	**-.507****
General health	*-.259***

Figures in bold represent important associations (r>0.40); italics represents predicted low associations (r<0.40)

Responsiveness/ Sensitivity

Analyses revealed that the SNOT-16 was sensitive to change in the patients' health status over time. The SNOT-16 showed a large shift from baseline to Test of Cure. The corresponding effect size and standardised response means were high indicating the high sensitivity of the instrument.

Sensitivity of the SNOT-16 showing mean (std dev) scores at baseline and test of cure (TOC)

	Baseline	TOC
SNOT-16 Mean	29.53 (10.51)	8.97 (9.11)
Effect size	*1.68*	
Standardised response mean	*1.36*	

Minimally Important Difference (MID)

The three MID estimates for the SNOT-16 were widely spaced, with the anchor estimate much higher than the two distributional based methods. The anchor based estimate was considered to be the most appropriate estimate of MID to use.

Estimates of minimally important difference for SNOT-16

	Total score	Average score
SNOT-16		
Anchor method	13.56	0.85
Standard error of measurement	3.26	0.21
½ standard deviation	4.60	0.29

N.b. Average score is often calculated for SNOT-16 so both values are presented here

DISCUSSION

The analyses of the trial data support the reliability, validity, and sensitivity of the SNOT-16 in patients with acute bacterial sinusitis and also establish an MID for this measure.

- The SNOT-16 has good internal consistency.
- With respect to construct validity, the SNOT-16 correlated significantly with several of the Rand SF-36 subscales, but not General Health.
- Given that acute sinusitis was being assessed, the findings suggest that despite a 24 hour recall period for all instruments, the acute nature of sinusitis is valued differently to overall general health.
- The analyses that are reported are consistent with the recommendations made by the US Food and Drug Administration regarding how PROs should be evaluated (FDA, 2006).

The analyses have some limitations which should be considered when interpreting the results.

- The analyses were conducted on trial data which was not designed or collected for the purposes of validating these instruments.
- Data from trial populations may be restricted in its generalisability because of strict study entry criteria.
- There was quite a large amount of missing data by TOC phase of the study.
- In estimating the MID for the SNOT-16 we originally intended to use, as an anchor, the degree of change reported by people who described themselves as having changed "a little better" i.e. the smallest measurable change. However in the analyses only 7 people were in this group and so this group was considered too small to serve as a basis for estimating MID. Therefore it was decided to adopt the value associated with "somewhat better".

REFERENCES

Anderson E, et al. *Otolaryng Head Neck Surg* 1999;121: 702-7.

Food and Drug Administration. Guidance for industry. Patient-Reported Outcome Measures: use in medical product development to support labelling claims. Draft Guidance. US Department of Health and Human Services, February 2006.

Guyatt GH et al *Mayo Clinic Proceedings*; 2002, 77: 371-383.

Hays R, *Health Economics* 1993;2:217-27.

Koopman C et al. *J Occupat Environ Med* 2002;44:14-20.

Morley A et al *Clin Otolaryng* 2006;31:103-9.

Nunnally et al. *Psychometric Theory*. 3rd Edition. New York: McGraw-Hill Series in Psychology,1978.

Piccirillo J et al. *Am J Rhinol* 1995;9:297-306.

Piccirillo J et al. *Otolaryng Head Neck Surg* 2002;126: 41-7.

Poole M. *Am J Med* 1999;106(5A):38S-47S.

Revicki DA et al. *Value Health*, 2007; 10(Supplement 2):S116-S124.

Slavin RG et al. *J Allergy Clin Immunol* 2005;116(6 SUPPL.).

Wild D, et al. *Health Qual Life Outcomes* 2005;3:42.

Oxford Outcomes