Psychometric Evaluation of the Sinonasal Outcome Test-16 and Activity Impairment Assessment in Acute Bacterial Sinusitis

Nuz Quadri, Andrew Lloyd, Karen N. Keating, Beenish Nafees, Jay Piccirillo and Diane Wild

Otolaryngology -- Head and Neck Surgery 2013 149: 161 originally published online 3 April 2013
DOI: 10.1177/0194599813482872

The online version of this article can be found at:
http://oto.sagepub.com/content/149/1/161

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
AMERICAN ACADEMY OF OTOLARYNGOLOGY--HEAD AND NECK SURGERY

American Academy of Otolaryngology- Head and Neck Surgery

Additional services and information for Otolaryngology -- Head and Neck Surgery can be found at:

Email Alerts: http://oto.sagepub.com/cgi/alerts
Subscriptions: http://oto.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Jul 1, 2013
OnlineFirst Version of Record - Apr 3, 2013
What is This?
Objective. To validate the Sinonasal Outcome Test-16 and Activity Impairment Assessment in patients with acute bacterial sinusitis.

Study Design. Data were used from a phase III clinical trial designed to evaluate the efficacy and safety of moxifloxacin 400 mg once daily for 5 consecutive days in the treatment of acute bacterial sinusitis. The psychometric properties and factor structure of the 2 measures were assessed.

Setting. Participants were given the measures to self-complete using either a telephone voice response system or a paper-and-pencil format.

Subjects and Methods. Three hundred seventy-four patients with acute bacterial sinusitis were used in the analysis. Patients received either a placebo or 400 mg moxifloxacin once daily. Patients were then reviewed at test of cure and follow-up. All analyses were conducted on a combined sample of placebo and active treatment patients.

Results. The Sinonasal Outcome Test-16 was associated with minimal missing data at baseline but a higher proportion by test of cure. There was no evidence of floor or ceiling effects and no significant skew. The Activity Impairment Assessment also had low missing data at baseline and no obvious floor or ceiling effects, but the data were not normally distributed. Both measures had good internal consistency. Convergent and divergent validity as well as sensitivity and the minimally important difference are also reported.

Conclusion. The measures both have good psychometric properties and are suitable for use with patients with acute bacterial sinusitis. Both instruments are sensitive. The minimal important difference estimates for the Sinonasal Outcome Test-16 are quite high but are similar to estimates reported previously.

Keywords
acute bacterial sinusitis, Sinonasal Outcome Test-16, Activity Impairment Assessment, psychometric analyses

Received August 21, 2012; revised February 6, 2013; accepted February 26, 2013.

Acute bacterial sinusitis (ABS) is a bacterial infection of one or more of the paranasal sinuses that usually complicates the common cold and other viral infections of the upper respiratory tract. Sinusitis affects approximately 16% of the adult population and is responsible for nearly $5.8 billion in health care costs annually. Sinusitis is classified on the basis of duration of symptoms and anatomic location. It is defined as an “inflammation of one or more of the paranasal sinuses, is characterized as acute when lasting less than four weeks, subacute when lasting four to eight weeks, and chronic when lasting longer than eight weeks.” Chronic sinusitis is the second most prevalent chronic disorder in the United States among persons aged 18 to 44 years, affecting approximately 31.8 million people in the United States annually. Acute sinusitis is considered a bacterial sinusitis (or rhinosinusitis) when the inflammation of the paranasal sinus mucosa is caused by bacterial overgrowth in a closed cavity. The most prominent symptoms of acute sinusitis include headache, nasal congestion, facial (and dental) pain, purulent rhinorrhea, postnasal drainage, and cough. The most common form of treatment for ABS is antibiotics, with a patient responding within 7 to 14 days.
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 (PRO) measures. Recently, Morley and Sharp5 provided a review of sinonasal outcome scoring systems in an attempt to identify the most appropriate tool in evaluating the effectiveness of treatments for chronic rhinosinusitis. In their review, they identify 15 instruments, including the Rhinosinusitis Outcome Measure-31 (RSOM-31),6 the Sinonasal Outcome Test-20 (SNOT-20),7 and the Sinonasal Outcome Test-16 (SNOT-16).8 The SNOT-20 was based on the RSOM-31, following the elimination of 11 items. The Sinonasal Outcome Test-22 (SNOT-22) was developed subsequently to include 2 further items that many clinicians thought should be included.9 The SNOT-16 was derived from the SNOT-20 but has been less widely used than either the SNOT-20 or SNOT-22; however, it has been reported to demonstrate good psychometric properties when assessed on patients with rhinosinusitis.10 The SNOT instruments were primarily designed to measure the effectiveness of treatment.

Garbutt et al11 showed that the SNOT-16 is a valid outcome measure for patients with clinically diagnosed acute rhinosinusitis. In this study, 166 adult patients were recruited from primary care practices in St Louis, Missouri. Diagnosis was based on clinical features, but sinus x-ray confirmation was not performed. Patients were part of a randomized controlled trial to evaluate antibiotic treatment for acute rhinosinusitis.

The Activity Impairment Assessment (AIA) was developed based on an existing work-productivity measure, the Stanford Presenteeism Scale-6.12 It was designed to evaluate the impact of health problems on individual performance and productivity, as well as other activities, including social activities. It also takes into account patients who may have responsibilities other than working full-time, such as going to school/university or looking after children. It has previously been validated in a trial in lower urinary tract infections.13 Wild et al13 found the AIA to have high levels of internal consistency (Cronbach’s $\alpha = 0.93$), convergent validity (all $r_s > .70$), and divergent validity ($r_s = .078$).

The study described in this article reports secondary analysis of clinical trial data to explore the psychometric properties of the SNOT-16 and AIA following their inclusion in a phase III, prospective, multicenter, randomized, double-blind, placebo-controlled trial. This trial was designed to evaluate the efficacy and safety of moxifloxacin 400 mg once daily for 5 days vs placebo in the treatment of ABS.

**Methods**

Data were collected from 74 clinical sites across the United States. Institutional review board approval was provided by 4 independent ethics committees/institutional review boards: Sterling Institutional Review Board, Western Institutional Review Board, Robbins Health Alliance, and PharmaTrials, Inc.

**Sample**

Three hundred seventy-four patients with ABS were recruited into the study; this sample size calculation was based on the primary variable of clinical response among patients with organisms. Based on assumptions of clinical cure rates of 80% and 50% for the moxifloxacin and placebo arms, respectively, and to obtain power of 90%, 117 patients were to be recruited into the study, assuming 30% of patients would have organisms. Therefore, the total estimated sample size required was 390. The organim rate was monitored vigilantly and was better than anticipated, and therefore the 30% needed was reached much earlier, and only 374 patients were recruited into the study.

All subjects had cultures and Gram stain of the sinus aspirate specimen obtained by using sinus puncture performed at the time of enrollment. Microbiologically valid subjects were defined as those subjects whose initial quantitative culture, obtained by sinus puncture, was positive for at least one of the following organisms: Streptococcus pneumoniae, Haemophilus influenzae, Moraxella catarrhalis, Streptococcus pyogenes, or Staphylococcus aureus. Any growth in culture of S pneumoniae, H influenzae, M catarrhalis, or S pyogenes was considered positive, whereas S aureus was considered causative only if $\geq 104$ colony-forming units were present.

Patients received either a placebo or 400 mg moxifloxacin once daily (1:2 ratio) over 5 days and were reviewed at test of cure (TOC) and follow-up.

Data from 374 patients were included in the analyses, which were conducted on a combined sample of placebo and active treatment patients. Patients were given outcome measures to self-complete using either an interactive voice response system (IVRS) or a paper-and-pencil format.

**Measures**

**Activity Impairment Assessment.** The AIA is a 5-item measure assessing activity impairment on a 5-point scale from $0 =$ none of the time to $4 =$ all of the time over the past 24 hours. The AIA was administered via IVRS and was completed at baseline (prior to first dose) and every 24 hours during treatment, at the TOC visit, and at premature discontinuation. A total score is generated from summing the item scores, where a higher score indicates a greater degree of impairment.

**Rand SF-36 Item Health Survey 1.0.** The Rand SF-36 is similar to the Medical Outcomes Study Short Form 36 (MOS SF-36).14 The MOS SF-36 has been previously used in the validation of the SNOT-16 for chronic sinusitis, with 7 of the 8 domains reaching statistically significant levels of agreement. The Rand SF-36 has a simpler scoring method and can be used with a 24-hour recall period. This recall period is consistent with the AIA and SNOT-16 and provides a more consistent approach to the validation of the 2 instruments.

The Rand SF-36 comprises 8 dimensions: physical functioning, role functioning/physical, role functioning/emotional, energy/fatigue, emotional well-being, social functioning, pain, and general health. In this study, a paper version of the Rand SF-36 was completed at baseline (prior to first dose), on day 3...
at the therapy visit, and at premature discontinuation/treatment failure in the investigator’s office.

**Sinonasal Outcome Test-16.** The SNOT-16 is a 16-item measure assessing rhinosinusitis symptoms on a 4-point scale from 0 = no problem to 3 = severe problem over the past 24 hours. The measure has been validated in a population of patients with chronic and acute rhinosinusitis. The SNOT-16 scores can be reported as the average score or the sum of all completed items (range, 0-48); both were reported in this study. The SNOT-16 was completed via IVRS at baseline (prior to first dose), every 24 hours during treatment, at the TOC visit, at premature discontinuation, and at the follow-up visit. The SNOT-16 has demonstrated internal consistency (Cronbach’s $\alpha = 0.89$), discriminant validity (against a cohort of patients with no symptoms of rhinosinusitis, $t = 3.87$, $P < .001$), and construct validity. A significant correlation was reported between the SNOT-16 and patients’ reported overall health and overall bother. In addition, the SNOT-16 was significantly correlated with 7 of the 8 domains of the SF-36. 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 for ABS.

**Global Rating of Change Questions.** Patients were asked to rate the severity of their sinus infection symptoms on a 4-point scale ranging from 0 = no symptoms at all to 3 = severe at baseline (prior to first dose), on day 3 at the therapy visit, at premature discontinuation, and at the TOC visit. At all visits after baseline, a further question was asked about the change in sinus infection symptoms since pretherapy. Patients who indicated their symptoms had improved were asked to indicate how much, using a 6-point scale ranging from 1 = a little to 6 = a very great deal. The Global Rating of Change (GRC) questions were administered in paper format in the investigator’s office.

**Statistical Analysis**

Analyses were conducted to evaluate the item performance, reliability, validity, and minimal important differences (MID) of the scales. All analyses (unless otherwise stated) were conducted on trial data imputed using last observation carried forward (LOCF).

The distribution and missingness patterns of the data were examined at baseline. Item distributions for the SNOT-16 and AIA were examined to identify ceiling and floor effects and whether the scales are normally distributed.

The internal consistency reliability of the measures was calculated using Cronbach’s $\alpha$, with a value of 0.70 set as the benchmark for declaring the scale as internally consistent. Cronbach’s $\alpha$ was also estimated with each item removed in turn. Convergent validity was supported if correlation coefficients between the SNOT-16, AIA, and Rand SF-36 fell between 0.40 and 0.70. Scores below 0.40 suggested the data are not possible to interpret in terms of what they were measuring; scores above 0.70 may have suggested that the data are too similar and may question the unique value of the AIA and SNOT-16 beyond the use of the Rand SF-36.

The responsiveness of the SNOT-16 and AIA to change in health was assessed by comparing baseline data with TOC data. Paired sample $t$ tests were conducted, and the effect size and standardized response means statistics were calculated.

The MID can be estimated using different methods; the change in SNOT-16 and AIA scores, which corresponds with the smallest detectable change on the GRC (eg, “a little better”), was 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 half a standard deviation are both widely used and accepted methods for estimating MID. All 3 estimates were considered to settle on a single value, with most weight given to the anchor-based methods, as it will associate the MID with change in a clinical indicator (GRC).

**Results**

**Participant Characteristics**

Two hundred fifty-one patients comprised the treatment arm, and 123 patients comprised the placebo arm. In both arms, 65% of participants were women and middle aged (mean [SD], age 40.1 [13.8] for the treatment arm and 40.3 [13.0] for the placebo arm). A high proportion of the participants in both arms was white (>65.9%), followed by Hispanic (17.5%-22.8%) and African American (7.3%-8.8%). The same proportion of participants in both arms was employed full-time (62%), employed part-time (12%), and looking after the house and/or children full-time (13%).

**SNOT-16 Distributional Characteristics**

The individual response frequencies and missing data for items from the SNOT-16 at baseline are shown in Table 1.

**AIA Distributional Characteristics**

The individual response frequencies and missing data for items from the AIA at baseline are shown in Table 1.

**Internal Consistency Reliability**

**Tables 3 and 4** report the Cronbach’s $\alpha$ coefficients for both instruments at baseline and also report this result with each item deleted one at a time. Internal consistency for both measures was high (Cronbach’s $\alpha$: 0.874 for SNOT-16 and 0.940 for AIA). Cronbach’s $\alpha$ of items generally went down from the overall statistics (Cronbach’s $\alpha$: 0.874 for SNOT-16 and 0.940 for AIA) following item deletion, indicating that each item is consistent with the other items. This analysis did not identify evidence for poorly performing items.

**Construct Validity**

The SNOT-16 and AIA were both significantly correlated with dimensions on the SF-36 (Table 5). Correlations

Downloaded from oto.sagepub.com at SOCIEDADE BRASILEIRA DE CIRUR on July 2, 2013
greater than 0.40 were considered important a priori. In addition, the SNOT-16 showed a high correlation with the total AIA score ($r = 0.673$).

**Responsiveness/Sensitivity**

The analyses found that both instruments were sensitive to change in the patients’ health status over time (Table 6). Sensitivity was assessed in terms of significance testing (paired samples $t$ test), effect size, and standardized response mean. Both measures showed large shifts from baseline to TOC. The corresponding effect size and standardized response mean were also high for both instruments, supporting claims regarding their sensitivity in this population.

**MID**

The MID for both instruments was also assessed using the trial data (Table 7). The 3 estimates for the SNOT-16 were widely spaced, with the anchor estimate much higher than the 2 distributional-based methods. The 3 estimates for the AIA were much closer. For both instruments, the anchor-based estimate was adopted as the most appropriate estimate of MID to use as it is the most recommended method, and it estimated the smallest change in a clinical indicator (GRC).

**Discussion**

These analyses serve to support the reliability, validity, and sensitivity of the SNOT-16 and AIA in patients with ABS. The SNOT-16 does effectively satisfy most of the psychometric criteria showing good internal consistency and construct validity, and it is sensitive to changes in symptom burden. Similarly, the AIA also has evidence to support its reliability and validity in this indication. The MID values for these measures apply to group-level data, and thus the most conservative estimate would be the largest values.
However, as the largest value for the SNOT was unusually high in relation to the other values (see Table 7), the next largest value was taken: 4.60 and 2.84 for the SNOT-16 and AIA, respectively.
Table 6. Sensitivity of the Sinonasal Outcome Test-16 (SNOT-16) and Activity Impairment Assessment (AIA) showing mean (SD) scores at baseline and test of cure (TOC).

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>TOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNOT-16</td>
<td>29.53 (10.51)</td>
<td>8.97 (9.11)</td>
</tr>
<tr>
<td>Effect size</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>Standardized response mean</td>
<td>1.36</td>
<td></td>
</tr>
<tr>
<td>AIA</td>
<td>8.80 (5.48)</td>
<td>2.21 (3.70)</td>
</tr>
<tr>
<td>Effect size</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>Standardized response mean</td>
<td>0.82</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Estimates of minimally important difference for the Sinonasal Outcome Test-16 (SNOT-16) and Activity Impairment Assessment (AIA).

<table>
<thead>
<tr>
<th></th>
<th>Total Score</th>
<th>Average Score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNOT-16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anchor method</td>
<td>13.56</td>
<td>0.85</td>
</tr>
<tr>
<td>Standard error of measurement</td>
<td>3.26</td>
<td>0.21</td>
</tr>
<tr>
<td>Half standard deviation</td>
<td>4.60</td>
<td>0.29</td>
</tr>
<tr>
<td>AIA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anchor method</td>
<td>2.63</td>
<td></td>
</tr>
<tr>
<td>Standard error of measurement</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td>Half standard deviation</td>
<td>2.84</td>
<td></td>
</tr>
</tbody>
</table>

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

With respect to construct validity, the SNOT-16 correlated highly with the AIA total score and with the Rand SF-36 subscales. The General Health subscale of the SF-36 did not correlate with the SNOT-16 total scores, nor did it correlate highly with other subscales of the Rand SF-36 itself, and perhaps suggests that participants were judging their general health differently to the severity of sinusitis problems as identified in the SNOT-16. These correlations demonstrate that on the whole, the lower the overall health status of the respondent, the greater his or her problems with the symptoms and functioning identified by the SNOT-16. Given that acute sinusitis was being assessed, this would suggest that despite a 24-hour recall period for all instruments, the acute nature of sinusitis is valued slightly differently to overall general health.

The final Food and Drug Administration guidance on the use of PRO measures to support product label claims includes recommendations that existing measures continue to be assessed over time to confirm their psychometric properties. In addition, as measures are used outside of their original indication, it becomes very important to verify their psychometric properties. This current work is in line with this recommendation. The SNOT-16 in the current context is being used in an indication that is very close to the original indication it was developed for. Interestingly, the AIA in this study was used in a very different indication to previous validation work but the internal consistency is reported as high. Therefore, the study demonstrates that the SNOT-16 is reliable and valid in this sample. Perhaps more interestingly, it also supports the validity and reliability of the AIA in this population, despite the measure having been validated in a very different patient population previously.

The study has some limitations that should be considered when interpreting the results. The analyses were conducted using clinical trial data, which were not designed or collected for the purposes of validating these instruments. Trial populations generally have quite strict entry criteria to minimize variation in the data from comorbidities or other factors.

However, since a large sample was recruited into the study and the incidence rate of ABS was reflected, we consider that the sample meets the criteria for a validation study of this type. The sample appears to be skewed toward more women of middle age, but the incidence rate of ABS is in fact higher in this demographic, as women are almost twice as likely to be diagnosed, and adults between the ages of 45 and 74 years are most commonly affected. The use of trial data as opposed to a stand-alone study means that it was not possible to estimate test-retest reliability. The reason for this is that the study sample could not be considered to be stable between any 2 assessments. Although this effectively means that the measures have not been fully validated, the test-retest reliability of the SNOT-16 has been shown previously. In addition, a stand-alone test-retest study could be conducted using a short interval between administrations in an effort to ensure stability in the condition.

There were quite substantial amounts of missing data by the TOC phase of the study. This is a common problem in many trials and may be an even greater problem in such a short study with daily evaluations. To address this, we used the LOCF method to impute missing values; this method is actually quite conservative for this patient population, in which patients are likely to be improving over time as compared with an oncology trial, for example, where the LOCF method may actually underestimate a decline in health-related quality of life. Large amounts of missing data will inevitably make us question how representative the study data are. However, personal communications with the instrument developer (Dr Jay Piccirillo) and references to previous publications have indicated that the results reported here are similar to previous analyses of the SNOT-16.

In estimating the MID for both measures, it was originally intended as an anchor to use the degree of change reported by people who described themselves as having changed “a little better” (ie, the smallest measurable change). However, in the analyses, only 7 people were in this group, and their estimate of MID was actually higher for both instruments than the estimate for the category “somewhat better.” Therefore, given that the minimally important difference was being estimated, the group decided to adopt the value associated with “somewhat better.”
Conclusion
This analysis supports the reliability, construct validity, and sensitivity of the SNOT-16 and AIA measures in ABS, although further research is warranted to determine the test-retest reliability of the SNOT-16 and AIA. The analysis also established estimates for the MID for both measures in this patient group. The results were consistent with analyses reported in other indications for these 2 measures.

Acknowledgments
We would like to acknowledge the assistance of Christina Donatti, PhD, in working on an earlier draft of the analysis plan. We would also like to acknowledge and thank those who participated in this study.

Author Contributions
Nuz Quadri, interpretation of the data, revising the first draft of the manuscript leading to final approval of all authors; Andrew Lloyd, initial conception and design of the research study, analysis of the data, revising the draft manuscript, approving the final manuscript; Karen N. Keating, conception and design of the research study, initial data collection stage, assisting with drafting the manuscript and final approval; Beenish Nafees, analysis of the data, preparing a first draft of the manuscript, approving the final manuscript; Jay Piccirillo, initial conception of the research study, interpretation of the data collected using the SNOT-16 questionnaire, assisting with drafting the manuscript and final approval; Diane Wild, initial conception and design of the research study, revising the draft manuscript, approving the final manuscript.

Disclosures
Competing interests: Nuz Quadri, Diane Wild, Andrew Lloyd, and Beenish Nafees are consultants at Oxford Outcomes (an ICON plc company) and were paid a fixed fee to carry out research and prepare the manuscript. Karen N. Keating is employed by Bayer, the sponsor of the study. Jay Piccirillo, developer of the SNOT-16, received royalty payments for its commercial use.

Sponsorships: Bayer Healthcare Pharmaceuticals. The sponsor was involved in preparing the design of the study, by approaching Oxford Outcomes to conduct this work, and being involved in discussions with Oxford Outcomes regarding how to conduct the study and analyze the data. The sponsor conducted the initial research, which resulted in the data collection, and was also involved in drafting the manuscript before providing approval of the final manuscript.

Funding source: None.

References