

Clinimetric properties of the Glasgow Health Status Inventory, Glasgow Benefit Inventory, Peak Nasal Inspiratory Flow, and 4-Phase Rhinomanometry in adults with nasal obstruction*

M.M.H.T. van Egmond¹, N. van Heerbeek¹, E.L.M. ter Haar², M.M. Rovers^{2,3}

Rhinology 55: 126-134, 2017

<https://doi.org/10.4193/Rhino16.296>

¹ Department of Otorhinolaryngology, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, the Netherlands

² Department of Operating Rooms, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, the Netherlands

³ Department of Health Evidence, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, the Netherlands

***Received for publication:**

October 14, 2016

Accepted: December 11, 2016

Abstract

Background: The validity of many measurement instruments frequently used in rhinology is unknown. This study describes clinimetric properties of well-known subjective and objective outcomes, i.e., the Glasgow Health Status Inventory, Glasgow Benefit Inventory, Peak Nasal Inspiratory Flow, and 4-Phase Rhinomanometry, in adults with nasal obstruction.

Methodology: Construct validity and responsiveness were determined in 111 patients. Inter-rater and intra-rater reliability were analysed in 30 patients. We assessed content validity by interviewing patients and ENT-surgeons; construct validity by comparing hypothesised associations to calculated correlations between the outcomes; inter-rater reliability by having two researchers perform objective measurements in the same patients; intra-rater reliability by having one rater administer all instruments twice within a two-week interval; and responsiveness by comparing patients' scores at baseline and three months after septoplasty or non-surgical management.

Results: All instruments demonstrated adequate content validity, inter-, and intra-rater reliability. Analyses of construct validity yielded low Pearson's correlations between the subjective and objective outcomes. Comparing septoplasty to non-surgical management, only the Glasgow Health Status Inventory scores were different between the two groups (mean difference 10.4, 95% CI 6.9 - 13.9).

Conclusion: All measurement instruments scored appropriately on content validity and reliability, but only the subjective GHSI scored well on responsiveness.

Key words: nasal obstruction, quality of life, rhinomanometry, reproducibility of results, validation studies

Introduction

Nasal obstruction is one of the most common reasons to consult an otorhinolaryngologist⁽¹⁾. Patients who suffer from nasal obstruction experience discomfort due to insufficient airflow through the nose, which can have a negative impact on health-related quality of life^(2,3). The etiology of nasal obstruction

comprises a multitude of mucosal and anatomical conditions, with nasal septal deviation being the most frequent anatomical cause⁽⁴⁾. Treatment depends on underlying pathogenesis. In clinical practice, a variety of therapies are available both within non-surgical as well as surgical strategies⁽⁵⁾. Apart from watchful waiting, non-surgical management includes medical treat-

ment such as steroids (intranasal or oral), decongestants, and antihistamines. Surgical treatment may consist of septoplasty, turbinate surgery, or a combination of these interventions. Septoplasty aims to straighten the deviated nasal septum. Although septoplasty is the most frequently performed ear, nose, and throat (ENT) operation in adults, its effectiveness in patients with nasal obstruction and a deviated nasal septum remains uncertain⁽⁶⁾. There are no internationally accepted guidelines regarding indications for septoplasty and scientific evidence is inconclusive⁽⁷⁾. Moreover, the additional benefit of concurrent turbinate surgery is questioned⁽⁸⁾. The evaluation of treatments for nasal obstruction in adults with a deviated nasal septum is hampered by the lack of clearly validated measurement instruments. Both subjective (e.g., health-related quality of life) as well as objective (e.g., nasal patency) outcome measures have been described, but clinimetric properties of most instruments have never been sufficiently assessed⁽⁹⁾. Furthermore, subjective and objective outcome measures do not always correlate, and there is no general agreement on which outcome measure to prefer⁽⁹⁻¹¹⁾.

According to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN), assessment of clinimetric properties of measurement instruments is pivotal for proper instrument selection – which, in turn, may stimulate evidence-based practice by preventing incomparable studies and incorrect conclusions⁽¹²⁾. The quality of a measurement instrument is determined by its performance on three quality domains, each consisting of one or more measurement properties. The three quality domains distinguished by COSMIN are: a) validity, i.e., the degree to which the instrument measures the construct(s) it intends to measure; b) reliability, i.e., the degree to which the instrument is free from measurement error; and c) responsiveness, i.e., the ability of an instrument to detect change over time in the construct(s) to be measured⁽¹²⁾.

The aim of this study is to evaluate the clinimetric properties of four frequently used measurement instruments for assessing nasal obstruction in terms of all three quality domains, following COSMIN standards. The two subjective outcome measures under study are the Glasgow Health Status Inventory (GHSI) and the Glasgow Benefit Inventory (GBI). Both the GHSI and the GBI are specifically developed for health problems in otorhinolaryngology. Where the GHSI aims at assessing quality of life as a health status, the GBI is designed to detect changes in quality of life as a result of ENT interventions⁽¹³⁾. The two objective outcome measures under study are Peak Nasal Inspiratory Flow (PNIF) and 4-Phase Rhinomanometry (4PR). Both measurement instruments are used to assess nasal patency. PNIF measures airflow during maximum inspiration, whereas 4PR measures effective resistance separately for each side of the nose, taking

intranasal pressure, flow, and the factor time into account⁽¹⁴⁻¹⁷⁾. We will describe the validity, reliability, and responsiveness of these measurement instruments in a population of adults under treatment for nasal obstruction due to a deviated nasal septum.

Materials and methods

Setting and participants

Interviews on validity were performed with patients and ENT-surgeons. Interviewed patients were adults with nasal obstruction due to a deviated nasal septum, who had completed the four measurements under study at least once. ENT-surgeons selected for the interviews were experienced specialists in the field of rhinology and were recruited from both tertiary as well as secondary referral hospitals.

Reliability was determined in a sample of adults attending the outpatient clinic because of nasal obstruction due to a deviated nasal septum, either on a first visit or after receiving surgical or non-surgical treatment.

Construct validity and responsiveness were assessed using data collected from participants in our concurrent randomised controlled trial on the effectiveness of septoplasty (trial identifying number NTR3868, Dutch Trial Registry, www.trialregister.nl). In this multicentre trial, adults with an indication to have septoplasty performed (with or without turbinate surgery) because of nasal obstruction due to a deviated nasal septum are eligible to participate⁽⁷⁾.

In all patients participating in the current study, the nasal septal deviation was primarily diagnosed by an internal exam of the nose, performed by the ENT-surgeon and consisting of anterior rhinoscopy as well as nasal endoscopy. The internal exam of the nose had to demonstrate that the deviation was causing a mechanical nasal airway obstruction, leading to impaired nasal breathing. Accordingly, symptomatic impairment of the nasal passage due to nasal septal deviation had to be the primary indication to perform septoplasty. If clinical evaluation revealed nasal septal perforation, untreated allergic rhinitis, or allergic rhinitis unresponsive to medical treatment, patients were not eligible to participate – as were patients with a history of previous nasal septal surgery, patients with an indication to have a cosmetic rhinoplasty procedure performed, and cleft lip and/or palate patients.

Subjective measurement instruments: GHSI and GBI

The GHSI and GBI comprise self-administered, digital questionnaires developed by the Medical Research Council's Institute of Hearing Research in collaboration with the Glasgow Royal Infirmary, Scotland. Both questionnaires consist of 18 questions scored on a 5-point Likert scale. Health-related quality of life is expressed as a total score ranging from 0 (poorest health status) to plus 100 (best health status) on the GHSI and from minus 100 (maximum deterioration), through 0 (no change), to plus 100

(maximum improvement) on the GBI⁽¹⁸⁾.

Objective measurement instruments: PNIF and 4PR

Nasal patency was expressed by PNIF as nasal airflow in litre per minute (L/min) through both nostrils, and by 4PR as effective resistance in pascal per cubic centimetre per second (Pa/cm³/s) separately for each side of the nose. All objective measurements were performed both before and after decongestion of nasal mucosa. As prescribed in the manual, every PNIF measurement consisted of three attempts; scores on all attempts were registered and the mean score was used for analyses⁽¹⁹⁾. Data from 4PR were obtained as logarithmic transformations and were analysed as such in accordance with recommendations from Vogt et al.⁽²⁰⁾. The inspiratory flow meter used for PNIF was produced by Clement Clarke International Ltd, Essex, United Kingdom. The rhinomanometer used for 4PR was produced by RhinoLab GmbH, Freiburg im Breisgau, Germany.

Content, construct, and criterion validity

Validity is composed of: a) content validity, i.e., the degree to which the content of an instrument adequately reflects the construct to be measured; b) construct validity, i.e., the degree to which scores on an instrument are consistent with prior hypotheses; and c) criterion validity, i.e., the degree to which scores on an instrument are an adequate reflection of a 'gold standard'⁽¹²⁾.

In this study, content validity was evaluated from the perspective of researchers, patients, and ENT-surgeons. First, the component of content validity known as face validity was determined by discussing among ourselves whether the four instruments under study indeed gave the impression of adequately reflecting the construct to be measured. Second, content validity was evaluated by conducting semi-structured interviews with both patients and ENT-surgeons. Respondents were invited to appraise all four measurement instruments on relevance and comprehensiveness. Input from the interviews was summarised and notable similarities and differences in opinion among respondents were registered.

Construct validity was addressed during the interviews by asking patients and ENT-surgeons to reflect on hypothesised associations between scores on the four measurement instruments. Afterwards, the hypothesised associations were compared to actual correlations, using data collected during the three-month follow-up visit of our trial patients.

Criterion validity could not be determined, since a gold standard for assessing nasal obstruction in adults with a deviated nasal septum is currently lacking.

Reliability

Reliability was assessed by performing repeated measurements under varying circumstances, i.e., by different raters on the same

occasion (inter-rater reliability) and by the same rater on different occasions (intra-rater reliability)⁽¹²⁾.

Inter-rater reliability of the two subjective measurement instruments was not applicable, as the GHSI and GBI are patient-reported outcome measures. Inter-rater reliability of the two objective measurement instruments (PNIF and 4PR) was assessed by having two experienced raters (i.e., researchers) perform the measurements independently and consecutively on the same day in the same sample of patients.

Intra-rater reliability of all four measurement instruments was determined by administering the instruments twice within a one- to two-week interval by the same rater (i.e., patient or researcher) in the same sample of patients. It was verified that the condition of patients had remained stable during the interval. Test conditions (e.g., instructions and setting) were similar for all measurements performed.

Responsiveness

For the assessment of responsiveness two administrations of a measurement instrument are needed to evaluate the instrument's ability to detect change over time. When scores on an instrument show no change, it may be difficult to decide whether the instrument is not responsive or the patients' condition simply remained stable. Therefore, it is recommended to select a study sample in which at least a proportion of patients has improved or deteriorated in the construct to be measured. It has been stated that change is likely to have occurred in patients who underwent an intervention in the interim between two administrations⁽¹²⁾. In this study, responsiveness of the GHSI, PNIF, and 4PR was assessed by administering these measurement instruments at baseline and after a three-month interval, in which either surgical intervention (i.e., septoplasty with or without concurrent turbinate surgery) or non-surgical management had taken place. The mean scores at baseline and after three months were calculated and a comparison was made between the septoplasty arm and the non-surgical management arm. The GBI expresses change in quality of life as a single score and can be completed only after surgical intervention. For this reason, administration at baseline or in the non-surgical management arm was not applicable and responsiveness could not be determined.

Statistical analyses

To determine construct validity, Pearson's correlation coefficients between scores on the four measurement instruments were calculated. Coefficients less than -0.70 or greater than 0.70 were defined as strong correlation; coefficients in the range of -0.70 to -0.30 or in the range of 0.30 to 0.70 expressed moderate correlation; and coefficients greater than -0.30 but less than 0.30 were considered as indicators of weak correlation. Inter-rater as well as intra-rater reliability were determined by calculating the

Intraclass Correlation Coefficient (ICC). As opposed to Pearson's correlation coefficient, the ICC is a parameter of reliability for continuous variables which takes systematic errors between repeated measurements into account, using a two-way random effects model to calculate absolute agreement^(21,22). When the ICC was at least 0.70, measurement instruments were considered to be reliable⁽²²⁾. To assess responsiveness, the change between baseline and 3 months in the septoplasty arm was compared to the non-surgical management arm with a Student's t-test. Analyses were performed using IBM SPSS Statistics software (version 22.0, IBM Corp., Armonk, NY, USA).

Results

Participants' characteristics

A total of 8 subjects, i.e., 3 patients (37.5%) and 5 ENT-surgeons (62.5%), participated in interviews on content and construct validity.

For assessing inter-rater reliability, 30 patients were recruited: 21 men (70.0%) and 9 women (30.0%); ages ranged between 18 and 68 years (mean 42, SD 15 years). Most patients (n=12, 40.0%) had not yet received treatment for nasal obstruction; 11 patients (36.7%) had undergone septoplasty, and 7 patients (23.3%) were treated with nasal medication.

Intra-rater reliability was also evaluated in 30 patients: 24 men (80.0%) and 6 women (20.0%); ages ranged between 18 and 68 years (mean 44, SD 16 years). In this sample, 13 patients (43.3%) had not yet received treatment for nasal obstruction; 9 patients (30.0%) had undergone septoplasty, and 8 patients (26.7%) were treated with nasal medication.

To determine construct validity and responsiveness, a total of 111 patients were analysed; 76 men (68.5%) and 35 women (31.5%) with ages ranging between 18 and 67 years (mean 38, SD 14 years). The septoplasty arm consisted of 60 patients (54.1%) and the other 51 patients (45.9%) received non-surgical management.

No missing data were present in all analyses performed.

Content and construct validity

The component of content validity known as face validity was considered to be appropriate for all four measurement instruments. The GHSI and GBI appeared to address all relevant aspects of health-related quality of life in adults with nasal obstruction due to a deviated nasal septum, as the questions covered not only general but also physical and social well-being. Also, nasal airflow measured by PNIF and effective resistance measured by 4PR gave the impression of adequately indicating nasal patency in the target population. With respect to content validity in general, interviewed patients and ENT-surgeons expressed an overall positive opinion regarding relevance and comprehensiveness of the four measurement instruments and their evaluative purposes. Nonetheless, the social domain of

Table 1. Overview of correlations between subjective and objective outcome measures.

Measurement instruments	Pearson's correlation coefficient (95% CI)
GHSI and GBI	0.28 (0.03 – 0.48)
PNIF and 4PR right (both before decongestion)	-0.29 (-0.44 – -0.13)
PNIF and 4PR left (both before decongestion)	-0.15 (-0.33 – 0.03)
PNIF and 4PR right (both after decongestion)	-0.08 (-0.26 – 0.08)
PNIF and 4PR left (both after decongestion)	-0.17 (-0.37 – 0.02)
GHSI and PNIF before decongestion	0.18 (-0.03 – 0.39)
GHSI and PNIF after decongestion	0.13 (-0.07 – 0.31)
GHSI and 4PR right before decongestion	-0.01 (-0.17 – 0.16)
GHSI and 4PR left before decongestion	-0.12 (-0.30 – 0.08)
GHSI and 4PR right after decongestion	-0.03 (-0.21 – 0.14)
GHSI and 4PR left after decongestion	0.00 (-0.18 – 0.16)
GBI and PNIF before decongestion	-0.14 (-0.35 – 0.08)
GBI and PNIF after decongestion	-0.09 (-0.33 – 0.13)
GBI and 4PR right before decongestion	0.03 (-0.26 – 0.32)
GBI and 4PR left before decongestion	0.13 (-0.10 – 0.33)
GBI and 4PR right after decongestion	-0.06 (-0.31 – 0.17)
GBI and 4PR left after decongestion	0.17 (-0.05 – 0.39)

the GHSI and GBI raised concerns by some respondents, as they regarded nasal obstruction as part of physical rather than social well-being. Furthermore, it was mentioned that 4PR seemed more complex to perform than PNIF, stressing the need for proper instructions.

Construct validity was determined by comparing hypothesised associations, based on input derived from interviews, to actual correlations between the four measurement instruments. Even though the GHSI and GBI both measure health-related quality of life, health status was perceived as distinct from post-interventional change, since a large improvement in nasal complaints may have only a limited effect on overall well-being. Taking this into account, it was hypothesised that a moderate correlation would exist between the GHSI and GBI. Also, the correlation between PNIF and 4PR was hypothesised to be moderate, as both assess nasal patency, but only with 4PR each side of the nose was measured separately – which

Table 2. Overview of inter-rater reliability of the two objective measurements.

Measurement instrument	Intraclass correlation coefficient (95% CI)
PNIF before decongestion	0.96 (0.90 – 0.98)
PNIF after decongestion	0.95 (0.90 – 0.98)
4PR before decongestion right	0.69 (0.35 – 0.85)
4PR before decongestion left	0.89 (0.76 – 0.95)
4PR after decongestion right	0.82 (0.63 – 0.91)
4PR after decongestion left	0.86 (0.70 – 0.93)

was regarded as important, because nasal obstruction resulting from nasal septal deviation may be unilateral. Moreover, the correlation between PNIF and 4PR was expected to be negative, as high airflow is likely to be accompanied by low resistance. Correlations between the two subjective and the two objective measurement instruments were hypothesised to be weak, since health-related quality of life and nasal patency were considered to be different constructs. After calculating actual correlations between the four measurement instruments, prior hypotheses were partly confirmed. As expected, correlations between PNIF and 4PR were negative, and correlations between subjective and objective outcomes were weak. However, the GHSI was found to correlate weakly with the GBI, as was PNIF with 4PR – where in both cases, moderate correlations were hypothesised. Comparing the GHSI with GBI resulted in a Pearson’s correlation coefficient of 0.28. Comparing PNIF with 4PR separately for each side of the nose, both before and after decongestion, yielded Pearson’s correlation coefficients ranging from -0.29 to -0.08. For each pair of outcomes, the differences between the two instruments that had already been acknowledged, appeared to be even more important than accounted for. Table 1 provides an overview of Pearson’s correlation coefficients.

Reliability

Inter-rater reliability of PNIF was excellent both before and after decongestion of nasal mucosa, with ICCs of 0.96 (95% CI 0.90 – 0.98) and 0.95 (95% CI 0.90 – 0.98), respectively. As shown in Table 2, ICCs of inter-rater reliability of 4PR also remained above the cut-off value of 0.70, apart from the right-sided measurement before decongestion, which scored marginally lower (ICC 0.69, 95% CI 0.35 – 0.85).

Intra-rater reliability of all four measurement instruments was assessed with a mean interval of 12.5 days (SD 8.5 days). An overview of ICCs and confidence intervals can be found in Table 3, indicating good intra-rater reliability of the GHSI, GBI, and

Table 3. Overview of intra-rater reliability of all four measurement instruments.

Measurement instrument	Intraclass correlation coefficient (95% CI)
GHSI	0.87 (0.73 – 0.94)
GBI	0.91 (0.60 – 0.98)
PNIF before decongestion	0.83 (0.64 – 0.92)
PNIF after decongestion	0.91 (0.82 – 0.96)
4PR before decongestion right	0.72 (0.42 – 0.87)
4PR before decongestion left	0.71 (0.39 – 0.86)
4PR after decongestion right	0.69 (0.34 – 0.85)
4PR after decongestion left	0.88 (0.76 – 0.95)

PNIF (before and after decongestion of nasal mucosa). Again, ICCs of 4PR remained consistently above the cut-off value, apart from one measurement scoring slightly under 0.70 (Table 3).

Responsiveness

An overview of mean scores and 95% confidence intervals on the GHSI, PNIF, and 4PR at baseline in both the septoplasty as well as the non-surgical management arm is provided in Table 4. The mean change scores between baseline and three months were calculated and a Student’s t-test was performed to study the difference in both arms (Table 4). Although patients who had undergone septoplasty appeared to have improved on almost every outcome, the same was true for patients who had received non-surgical management – apart from scores on the Glasgow Health Status Inventory, which showed a clear improvement in the septoplasty arm (mean difference baseline and three months 6.8, 95% CI 4.4 – 9.2) versus a deterioration in the non-surgical management arm (mean difference baseline and three months -3.6, 95% CI -6.2 – -0.98). Only for the GHSI, the difference in treatment arms was statistically significant (mean difference 10.4, 95% CI 6.9 – 13.9). Mean change scores on PNIF showed improvement irrespective of mucosal decongestion or treatment received. The difference in scores on 4PR was merely minor, and this too was comparable for both treatment arms.

Results per measurement instrument

Both subjective outcome measures (GHSI and GBI) generally scored well on content validity and intra-rater reliability. Responsiveness was found to be adequate for the GHSI, but could not be determined for the GBI since this questionnaire can be completed only after surgical intervention. The objective measurement instruments (PNIF and 4PR) scored appropriately

Table 4. Overview of scores on the GHSI, PNIF, and 4PR in both treatment arms.

Measurement instrument	Septoplasty		Non-surgical management		Comparing septoplasty and non-surgical management Difference in treatment arms (95% CI)
	Mean score at baseline (95% CI)	Mean difference baseline and three months (95% CI)	Mean score at baseline (95% CI)	Mean difference baseline and three months (95% CI)	
GHSI	63.1 (60.4 – 65.7)	6.8 (4.4 – 9.2)	62.9 (59.2 – 66.6)	-3.6 (-6.2 – -0.98)	10.4 (6.9 – 13.9)
PNIF before decongestion	93.3 (83.3 – 103.2)	12.8 (3.3 – 22.2)	85.5 (74.3 – 96.7)	11.5 (2.9 – 20.2)	1.3 (-11.6 – 14.1)
PNIF after decongestion	106.8 (96.4 – 117.3)	12.7 (4.3 – 21.1)	102.5 (90.8 – 114.3)	13.6 (5.8 – 21.4)	-0.90 (-12.4 – 10.6)
4PR before decongestion right	1.09 (0.96 – 1.21)	-0.01 (-0.12 – 0.10)	1.25 (1.11 – 1.39)	-0.14 (-0.28 – 0.00)	0.13 (-0.05 – 0.30)
4PR before decongestion left	1.14 (0.98 – 1.29)	0.01 (-0.12 – 0.12)	1.10 (0.96 – 1.24)	0.05 (-0.06 – 0.17)	-0.05 (-0.21 – 0.12)
4PR after decongestion right	1.05 (0.94 – 1.17)	-0.07 (-0.19 – 0.06)	1.06 (0.96 – 1.16)	-0.05 (-0.14 – 0.04)	-0.02 (-0.18 – 0.13)
4PR after decongestion left	1.06 (0.93 – 1.19)	0.02 (-0.08 – 0.13)	1.01 (0.90 – 1.12)	-0.02 (-0.12 – 0.08)	0.04 (-0.10 – 0.19)

on content validity and inter- as well as intra-rater reliability, but analyses of responsiveness showed less favourable results. For all four measurement instruments, a definite conclusion regarding construct validity cannot be drawn yet, as only some of our findings were in accordance with prior expectations. Considering their performance on all three quality domains, the GHSI seems to score best out of the four measurement instruments under study. Although clinimetric properties of the two objective outcome measures appear to be comparable, the user-friendliness of PNIF offers a practical advantage over 4PR. A graphical overview of validity, reliability, responsiveness, and overall performance per outcome measure is provided in Table 5.

Discussion

Key results

This study describes clinimetric properties of frequently used subjective (GHSI and GBI) and objective (PNIF and 4PR) measurement instruments for assessing nasal obstruction in adults with a deviated nasal septum. Content validity was considered to be appropriate for all four outcome measures. Analyses of construct validity showed that prior hypotheses were partly confirmed by our data. Inter-rater and intra-rater reliability of all outcome measures was adequate with ICCs above the cut-off value of 0.70, except for 4PR which scored slightly lower (0.69) on one measurement in both analyses. After calculating the difference in outcome scores at baseline and after three months in patients who had received either septoplasty or non-surgical management, both treatment arms were compared and a statistically significant difference was found for the Glasgow Health Status Inventory only.

Comparison with the literature

As far as we are aware, we are the first studying the validity, reliability, and responsiveness of these frequently used outcome measures in adults under treatment for nasal obstruction due to a deviated nasal septum. Previous clinimetric assessment of a modified version of the Glasgow Benefit Inventory presented initial evidence of adequate validity and reliability in children undergoing tonsillectomy or ventilation tube insertion⁽²³⁾. Several studies have compared PNIF with other subjective as well as objective outcome measures, but most were performed in patients with (allergic) rhinosinusitis and none involved a comprehensive evaluation of validity, reliability, and responsiveness⁽²⁴⁻²⁶⁾. The significance of rhinomanometry for the objective assessment of nasal patency has been acknowledged by many authors, some of whom even considered rhinomanometry to be the gold standard^(27,28). Remarkably, the scientific debate appears to have focused on different techniques for performing rhinomanometry rather than clinimetric properties of the measurement instruments involved⁽²⁹⁾.

Strengths and limitations

The major strength of our study is that we are the first to provide an extensive evaluation of the clinimetric properties of both subjective and objective measurement instruments in this target population. Furthermore, analyses were conducted in adequately sized samples according to the COSMIN checklist, in which a sample of 30 patients is considered 'fair' and more than 100 patients is regarded as 'excellent'⁽³⁰⁾. Some possible limitations should also be discussed. First, although inter-rater and intra-rater reliability of 4PR generally appeared to be appropriate, two ICCs just under the cut-off

Table 5. Graphical overview of validity, reliability, responsiveness, and overall performance per measurement instrument.

	GHSI	GBI	PNIF	4PR
Content validity	+	+	+	+
Construct validity	?	?	?	?
Inter-rater reliability	NA	NA	+	+
Intra-rater reliability	+	+	+	+
Responsiveness	+	NA	-	-
Overall performance	+	+/-	+	+/-

+ = adequate; +/- = suboptimal/moderate; - = weak; ? = unclear; NA = not applicable.

value of 0.70 were found. The confidence interval in both cases was rather broad, demonstrating that lower ICCs require larger samples to obtain a more precise estimate⁽³¹⁾. Nonetheless, the effect of a limited sample size is likely to be an underestimation, since the ICC is a ratio of the variability between subjects to the total variability. A higher number of subjects allows more precise estimation of subject variability, which composes the majority of total variability as the ICC approaches 1.00^(32,33). Second, as only the GHSI, GBI, PNIF, and 4PR were selected for this study, clinimetric properties of other outcome measures frequently applied in the field of rhinology were not assessed. These four measurement instruments were chosen for reasons of feasibility, but also because other relevant outcome measures appeared to be validated in previous studies. For example, the validity, reliability, and responsiveness of the Nasal Obstruction Symptom Evaluation (NOSE) scale, a 5-item instrument for determining disease-specific health status, were addressed by Stewart et al. in 2004⁽⁹⁾. However, the authors did not follow the now well-accepted COSMIN standards, as these were not yet available at that time⁽³⁴⁾. Other subjective measurement instruments such as the Sino-Nasal Outcome Test (SNOT) and the Rhinosinusitis Quality of Life (RhinoQOL) survey are also administered frequently in studies evaluating septoplasty, but they have been developed for and validated in another target population, i.e., patients with chronic rhinosinusitis⁽³⁵⁻³⁹⁾. Also in these patients, nasal obstruction is a common complaint – among other symptoms such as a runny nose, facial pressure or headache, and cough⁽³⁸⁾. Third, responsiveness was assessed using data from our ongoing trial on the effectiveness of septoplasty so the lack of responsiveness may imply that the instrument is indeed not responsive or that the intervention is not effective⁽¹²⁾. Consequently, it would be preferable to assess responsiveness in a sample of patients undergoing an intervention which is known to be

effective, but determining treatment effectiveness requires an outcome which is known to be responsive. Thus, this appears to be an example of circular reasoning that is not easily avoided. As an alternative to assessing responsiveness by means of an effect size, COSMIN recommends to compare change scores of the instrument under study to changes on a gold standard (criterion approach) or to compare hypothesised change scores to actual changes on the instrument under study (construct approach)⁽¹²⁾. As a gold standard is lacking, the criterion approach was not applicable for this study. Furthermore, the current lack of evidence for the effectiveness of septoplasty hampers prior formulation of detailed hypotheses on the magnitude and direction of the intervention's effect and resulting change scores, as demanded by the construct approach for assessing responsiveness⁽⁷⁾. Considering that the Glasgow Health Status Inventory did show a notable change after septoplasty compared to non-surgical management, we expect the non-significant differences in scores on the objective outcome measures to indicate weak responsiveness rather than absence of treatment effect. This is supported by Glasgow Benefit Inventory scores in the septoplasty arm, which suggest an improvement three months after surgery (mean score 11.9, 95% CI 8.8 – 15.1). Evidently, the ability to detect change over time is an important property of measurement instruments used for evaluative purposes. Lack of evidence for the validity of an instrument's change score does not, however, detract from the validity and reliability of its single score⁽⁴⁰⁾. Furthermore, it should be mentioned that assessing responsiveness is an ongoing process of gathering evidence and as a result, definite conclusions cannot be drawn yet. In our ongoing randomised controlled trial, the effectiveness of septoplasty is being assessed using these and other measurement instruments for nasal obstruction in adults, as well as a clinical evaluation both before and after treatment⁽⁷⁾. Results of this trial will likely contribute to determining treatment effects and by doing so, facilitate further clinimetric assessment of outcome measures in this target population.

Clinical implications

In patients with nasal obstruction due to nasal septal deviation, septoplasty aims to widen the nasal passage by straightening the deviated septum, as a result of which nasal breathing is assumed to improve. Accordingly, both the subjective sensation of airflow through the nose, as well as an objective quantification of nasal patency are important to take into account in the evaluation of treatment effect. Out of the two subjective outcome measures under study, the GHSI scored best on the three quality domains. Since PNIF appeared to have clinimetric properties similar to 4PR, its user-friendliness may be of decisive importance for the application in clinical practice.

Conclusion

Our results suggest adequate validity and reliability of all four measurement instruments, i.e. the Glasgow Health Status Inventory, Glasgow Benefit Inventory, Peak Nasal Inspiratory Flow, and 4-Phase Rhinomanometry, whereas only the GHSI appeared to be responsive to change in symptoms of nasal obstruction after septoplasty.

Acknowledgements

The authors would like to thank all participating patients and the ENT-surgeons Godelieve Damen, Koen Ingels, Ingrid van Rooy, and Cor Stengs for their contribution to the interviews; student assistants Charlotte Michels and Sanne Leuwerink for their help in data collection; trial coordinator Carine Hendriks for her support in organising the study; and the Department of Otorhinolaryngology at Canisius Wilhelmina Hospital Nijmegen for participating in patient recruitment. Analyses of construct validity and responsiveness were conducted using data from our concurrent randomised controlled trial on the effectiveness of septoplasty. This trial is funded by ZonMw (The Netherlands Organization for Health Research and Development) and Radboud university medical center Nijmegen.

Authorship contribution

MvE, NvH, and MR conceived the study and initiated the study design. MvE and EtH conducted data collection. MvE, EtH, and MR performed data analyses. All authors drafted this manuscript and approved the final version.

Conflict of interest

MvE, NvH, and MR are currently conducting a randomised controlled trial on the effectiveness of septoplasty, in which all four measurement instruments under study are being administered (among other outcome measures). However, in the authors' opinion this does not pose a conflict of interest, since results of the current study will not have any consequences for the measurement instruments used in the randomised controlled trial; nor will future results of the randomised controlled trial affect the current manuscript. All primary and secondary outcomes of the randomised controlled trial have been predefined in the registered and published trial protocol⁽⁷⁾. Moreover, the randomised controlled trial is still recruiting and thus, no data can be retrieved or analysed yet. Therefore, the authors declare that they do not have any competing interests.

References

- Chandra RK, Patadia MO, Raviv J. Diagnosis of nasal airway obstruction. *Otolaryngol Clin North Am*, 2009. 42(2): 207-25, vii.
- Jessen M, Malm L. Definition, prevalence and development of nasal obstruction. *Allergy*, 1997. 52(40 Suppl): 3-6.
- Udaka T, Suzuki H, Fujimura T, et al. Chronic nasal obstruction causes daytime sleepiness and decreased quality of life even in the absence of snoring. *Am J Rhinol*, 2007. 21(5): 564-9.
- Sipila J, Suonpaa J. A prospective study using rhinomanometry and patient clinical satisfaction to determine if objective measurements of nasal airway resistance can improve the quality of septoplasty. *Eur Arch Otorhinolaryngol*, 1997. 254(8): 387-90.
- Recker C, Hamilton GS, 3rd. Evaluation of the Patient with Nasal Obstruction. *Facial Plast Surg*, 2016. 32(1): 3-8.
- Manoukian PD, Wyatt JR, Leopold DA, Bass EB. Recent trends in utilization of procedures in otolaryngology-head and neck surgery. *Laryngoscope*, 1997. 107(4): 472-7.
- van Egmond MM, Rovers MM, Hendriks CT, van Heerbeek N. Effectiveness of septoplasty versus non-surgical management for nasal obstruction due to a deviated nasal septum in adults: study protocol for a randomized controlled trial. *Trials*, 2015. 16: 500.
- Illum, P. Septoplasty and compensatory inferior turbinate hypertrophy: long-term results after randomized turbinoplasty. *Eur Arch Otorhinolaryngol*, 1997. 254 Suppl 1: S89-92.
- Stewart MG, Witsell DL, Smith TL, Weaver EM, Yueh B, Hannley MT. Development and validation of the Nasal Obstruction Symptom Evaluation (NOSE) scale. *Otolaryngol Head Neck Surg*, 2004. 130(2): 157-63.
- Kahveci OK, Miman MC, Yucel A, Yucedag F, Okur E, Altuntas A. The efficiency of Nose Obstruction Symptom Evaluation (NOSE) scale on patients with nasal septal deviation. *Auris Nasus Larynx*, 2012. 39(3): 275-9.
- Baumann, I. Quality of life before and after septoplasty and rhinoplasty. *GMS Curr Top Otorhinolaryngol Head Neck Surg*, 2010. 9: Doc06.
- COSMIN Group. COSMIN checklist manual. Amsterdam, the Netherlands; 2012.
- Robinson K, Gatehouse S, Browning GG. Measuring patient benefit from otorhinolaryngological surgery and therapy. *Ann Otol Rhinol Laryngol*, 1996. 105(6): 415-22.
- Jones AS, Viani L, Phillips D, Charters P. The objective assessment of nasal patency. *Clin Otolaryngol Allied Sci*, 1991. 16(2): 206-11.
- Vogt K, Jalowayski AA, Althaus W, et al. 4-Phase-Rhinomanometry (4PR)--basics and practice 2010. *Rhinol Suppl*, 2010(21): 1-50.
- Ottaviano G, Scadding GK, Iacono V, Scarpa B, Martini A, Lund VJ. Peak nasal inspiratory flow and peak expiratory flow. Upright and sitting values in an adult population. *Rhinology*, 2016. 54(2): 160-3.
- Bermuller C, Kirsche H, Rettinger G, Riechelmann H. Diagnostic accuracy of peak nasal inspiratory flow and rhinomanometry in functional rhinosurgery. *Laryngoscope*, 2008. 118(4): 605-10.
- Medical Research Council Institute of Hearing Research and Glasgow Royal Infirmary. The Glasgow Health Status Questionnaires Manual. Glasgow, Scotland; 1998.
- Clement Clarke International Ltd. Portable Inspiratory Flow Meter.
- Vogt K, Wernecke KD, Behrbohm H, Gubisch W, Argale M. Four-phase rhinomanometry: a multicentric retrospective analysis of 36,563 clinical measurements. *Eur Arch Otorhinolaryngol*, 2016. 273(5): 1185-98.
- Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? *Injury*, 2011. 42(3): 236-40.
- Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*, 2007. 60(1): 34-42.
- Kubba H, Swan IR, Gatehouse S. The Glasgow Children's Benefit Inventory: a new instrument for assessing health-related benefit after an intervention. *Ann Otol Rhinol Laryngol*, 2004. 113(12): 980-6.
- Proimos EK, Kiagiadaki DE, Chimona TS, Seferlis FG, Maroudias NJ, Papadakis CE. Comparison of acoustic rhinometry and nasal inspiratory peak flow as objective tools for nasal obstruction assessment in patients with chronic rhinosinusitis. *Rhinology*, 2015. 53(1): 66-74.
- Starling-Schwanz R, Peake HL, Salome CM, et al. Repeatability of peak nasal inspiratory flow measurements and utility for assessing the severity of rhinitis. *Allergy*, 2005. 60(6): 795-800.
- Holmstrom M, Scadding GK, Lund VJ, Darby YC. Assessment of nasal obstruction. A

- comparison between rhinomanometry and nasal inspiratory peak flow. *Rhinology*, 1990. 28(3): 191-6.
27. Ottaviano G, Fokkens WJ. Measurements of nasal airflow and patency: a critical review with emphasis on the use of peak nasal inspiratory flow in daily practice. *Allergy*, 2016. 71(2): 162-74.
28. Clement PA, Halewyck S, Gordts F, Michel O. Critical evaluation of different objective techniques of nasal airway assessment: a clinical review. *Eur Arch Otorhinolaryngol*, 2014. 271(10): 2617-25.
29. Clement PA, Gordts F. Consensus report on acoustic rhinometry and rhinomanometry. *Rhinology*, 2005. 43(3): 169-79.
30. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*, 2012. 21(4): 651-7.
31. de Vet HC, Terwee CB, Mokkink LB, Knol DL. Sample size for reliability studies, in *Measurement in Medicine: A Practical Guide*. New York, United States of America: Cambridge University Press; 2011.
32. Giraudeau B, Mary JY. Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med*, 2001. 20(21): 3205-14.
33. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med*, 1990. 20(5): 337-40.
34. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*, 2010. 63(7): 737-45.
35. Buckland JR, Thomas S, Harries PG. Can the Sino-nasal Outcome Test (SNOT-22) be used as a reliable outcome measure for successful septal surgery? *Clin Otolaryngol Allied Sci*, 2003. 28(1): 43-7.
36. Hytonen ML, Lilja M, Makitie AA, Sintonen H, Roine RP. Does septoplasty enhance the quality of life in patients? *Eur Arch Otorhinolaryngol*, 2012. 269(12): 2497-503.
37. Mondina M, Marro M, Maurice S, Stoll D, de Gabory L. Assessment of nasal septoplasty using NOSE and RhinoQoL questionnaires. *Eur Arch Otorhinolaryngol*, 2012. 269(10): 2189-95.
38. Piccirillo JF, Merritt MG, Jr., Richards ML. Psychometric and clinimetric validity of the 20-Item Sino-Nasal Outcome Test (SNOT-20). *Otolaryngol Head Neck Surg*, 2002. 126(1): 41-7.
39. 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(5): 846-54.
40. de Vet HC, Terwee CB, Mokkink LB, Knol DL. The concept of responsiveness, in *Measurement in Medicine: A Practical Guide*. New York, United States of America: Cambridge University Press; 2011.

M.M.H.T. van Egmond, MD, MSc
Radboud university medical center
Department of Otorhinolaryngology,
route 377
P.O. Box 9101
6500 HB Nijmegen
the Netherlands

Tel: +31-24-3614178,
F: +31-24-3540251
E: Machteld.vanEgmond
@radboudumc.nl