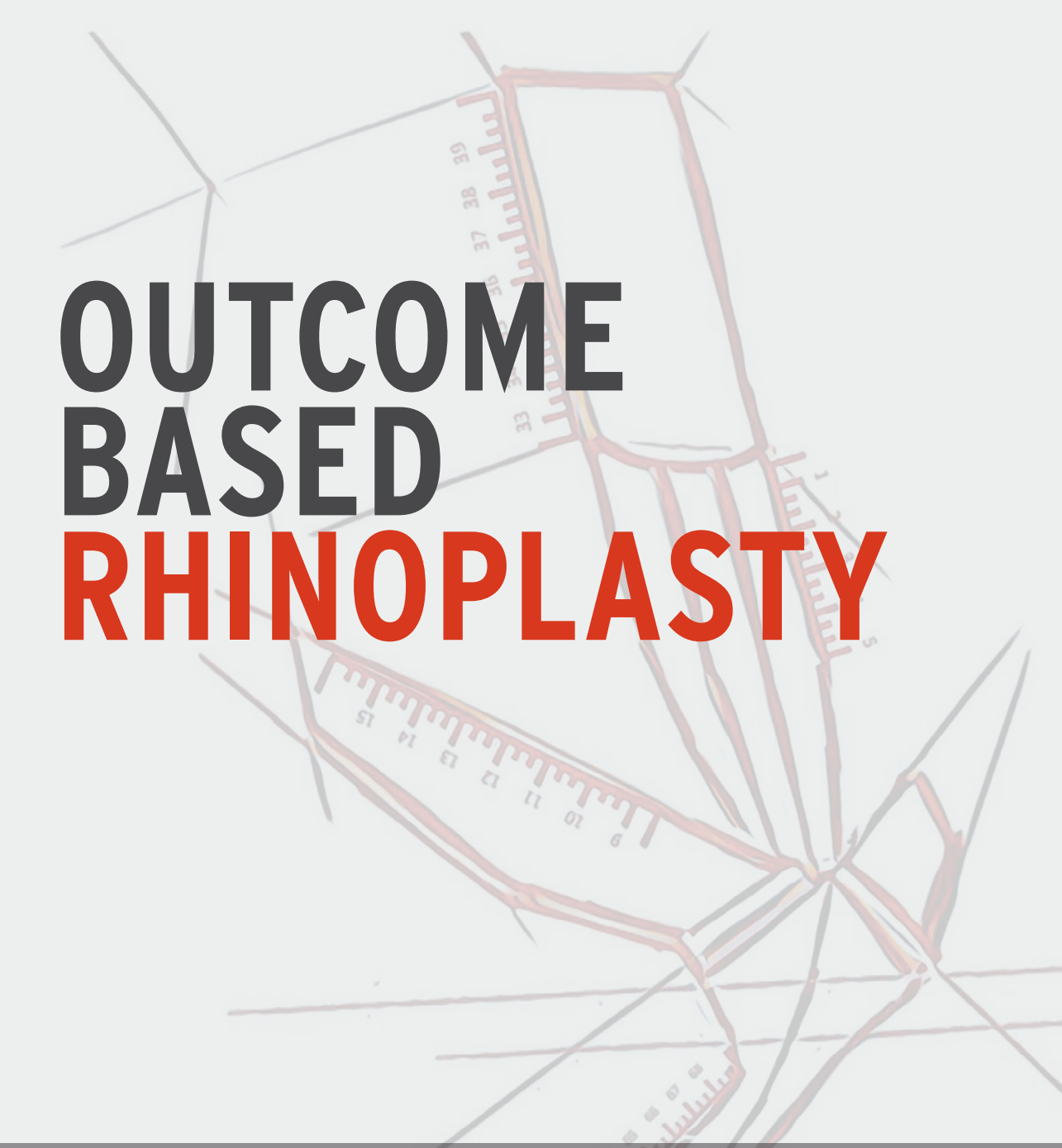


OUTCOME BASED RHINOPLASTY

TOWARDS DATA-DRIVEN DECISION-MAKING IN FUNCTIONAL AND AESTHETIC NASAL SURGERY

FLORIS V.W.J. VAN ZIJL

OUTCOME BASED RHINOPLASTY

Towards data-driven decision-making
in functional and aesthetic nasal surgery

F.V.W.J. van Zijl

COLOFON

Layout: © evelienjagtman.com
Cover and Printing: ridderprint.nl

ISBN: 978-94-6458-140-9

OUTCOME BASED RHINOPLASTY

Towards data-driven decision-making in functional and aesthetic nasal surgery

Uitkomstgerichte neuschirurgie

Op weg naar data gestuurde besluitvorming
bij functionele en esthetische neuscorrecties

Academisch proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag de
rector magnificus

Prof.dr. A.L. Bredenoord

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

woensdag 22 juni 2022 om **10:30 uur**

door

Floris Vincent Willem Joseph van Zijl

geboren te Maastricht

Erasmus University Rotterdam

Promotiecommissie

Promotor: Prof.dr. R.J. Baatenburg de Jong

Overige leden: Prof.dr. J.A. Hazelzet
Prof.dr. I.M.J. Mathijssen
Prof.dr. H.F. Lingsma

Copromotor: Dr. F.R. Datema

Paranimfen: Drs. H.F.E. van der Toom
Drs. J.A.W.M. van Zijl

Table of contents

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Chapter 1

General introduction and aims of this thesis

1. Outcome-based healthcare

Over the past years, increasing pressure is exerted on healthcare providers to communicate outcome data. Outcomes reflect the change in health status following treatment, and are essential to evaluate the effectiveness of healthcare. Primarily, outcome information is of great importance to governments and regulators. Healthcare systems are facing a sustainability crisis: the ageing population and consequently rising chronic disease burden are placing an increasing demand on healthcare, at a backdrop of limited resources. To contain costs, insights in precisely what interventions achieve what health benefit are critical to ensure that every euro spent produces the best outcomes. Whereas traditionally these outcomes encompassed more clinical metrics such as mortality and readmissions, the focus is now shifting towards a more patient-centered approach concentrating on results that matter to patients. Patient-relevant outcomes are expected to play an increasingly important role in reimbursements, accreditation programs, quality inspections and clinical trials.[1-3]

Apart from policy makers and payers, the demand for more relevant outcome information also arises from patients. In order to make educated decisions about their health, patients rely on meaningful and comprehensible outcome information. Patients are keen to participate in the decision-making process, but are often not provided with the information they need.[4-6] This particularly concerns outcome information that they can relate to, such as whether they can climb stairs rather than whether their spirometry has improved.[7] In addition, data on how well their individual doctor performs is absent, information that patients consider important in choosing where to receive treatment.[8-10] Patients are unlikely to receive data-driven answers from their clinicians: although clinicians acknowledge the value of outcome information, very few monitor their own outcomes.[11] Most clinicians rely on gut feeling to determine how well they perform. Since clinicians commonly judge their own performance as unusually succesful, the validity of this gut feeling is questionable; 90% rate themselves in the upper quartile and no-one sees themselves as below average in relation to peers.[12,13]

Certainly, patients, clinicians and interested third parties can consult the scientific literature to retrieve information about the expected effect of interventions. Due to their design, randomized controlled clinical trials (RCTs) are considered the gold standard to study the effect of an intervention.[14] As the impact of bias is minimized by strictly controlling the study population, setting and methods, RCTs provide the highest level of evidence on the efficacy of treatment. For many therapies however, including surgery, this evidence is missing because attaining such evidence poses significant technical, ethical or financial challenges.[15] Moreover, even if efficacy has

been demonstrated in well-controlled experimental settings, interventions could perform differently in general clinical practice.[16] Patient characteristics may be more diverse, many interventions are subject to substantial procedural variability, and the background, experience or interests of providers may differ significantly.[17] For example, most doctors quote scientific literature to inform and council patients. But can a surgeon who performs a certain procedure twice a year reliably quote the results of a study from surgeons performing that procedure 50 times a year? A discrepancy therefore exists between efficacy; demonstrated health benefit in an ideal, controlled situation, and effectiveness: the benefit of an intervention in the uncontrolled setting of daily practice. Evidence-based results may differ substantially from practice-based results, and it is precisely this practice-based evidence that is lacking.

These considerations are the foundation of an increasing interest in routinely measuring patient-relevant outcomes in healthcare. Over the past two decades, clinicians have acknowledged the value of outcome-based healthcare and have started to incorporate systematic outcome measurements in practice. In functional and aesthetic nasal surgery, or rhinoplasty, numerous outcome instruments that measure the effect of rhinoplasty have recently emerged.[18,19] Still, these instruments have primarily been used in a scientific setting and consensus on what instruments to use, how to incorporate outcome measurements in practice and how to analyze collected outcomes is absent. Consequently, rhinoplasty outcome data remain scarce. This thesis studies the selection and implementation of routine outcome measures in rhinoplasty, in an attempt to advance from gut feeling and intuition to data-driven decision-making.

The history of outcome measurements in healthcare

Although outcome measurements began to gain momentum over the past years, measuring outcomes in healthcare is not new. Florence Nightingale was the first to systematically measure outcomes: her revolutionary system of recording outcomes among soldiers during the Crimean War (1853-1856) enabled her to link death to poor hygiene, a finding that effectuated a spectacular decline of mortality.[20] Another protagonist of outcome measurements is pioneering surgeon Ernest Codman (Boston, 1869-1940). In order to evaluate treatment effects, learn from unwanted outcomes and enable comparisons between surgeons and hospitals, he proposed to systematically measure surgical results.[21] Unfortunately, Codman was far ahead of his time; his ideas received little appreciation by his colleagues and the evolution of outcomes measurement was brought to a halt. It was not until after the Second World War that, as a result of rapidly expanding health services, third-party payers demanded evidence of healthcare effectiveness and containment of costs.[22] Primarily, traditional measures of clinical outcomes such as mortality and complications were used as endpoints. During the 1960s and 1970s, it became eminent that these traditional indices were not always the most relevant outcomes. A more holistic approach of outcomes that are important to patients was adopted, in order to capture the multidimensionality of health status.[23] Most treatments do not aim to reduce mortality but rather aim to decrease specific symptoms, minimize disability or improve quality of life; examples of which are chronic conditions or elective procedures. Therefore, an understanding of the full range of different outcomes was necessary if outcomes relevant to patients were to be measured. A widely accepted model describing the different dimensions of patient outcomes is that of Wilson and Cleary.[24] In this model, five conceptually distinct measures of health status are identified: biological and physiological factors (e.g. blood pressure or serum hemoglobin level), symptoms (such as pain or fatigue), functioning (e.g. reduced ability to climb stairs), general health perceptions and overall quality of life.

Patient-reported outcome measures

With the exception of biological and physiological factors, the dimensions of Wilson and Cleary's model are aspects that require the perspective of patients. To capture this perspective, patient-reported outcome measures (PROMs) were introduced.[25] A PROM aims to measure the patients' view of his or her own health experience in a reliable, valid and feasible manner.[26] This self-reported assessment, usually a questionnaire, provides a quantitative evaluation of otherwise subjective results. PROMs can be subdivided into disease-specific and generic instruments. Disease-specific PROMs examine symptomatic or functional manifestations specifically related to a disease or treatment, generic PROMs are targeted at measuring

functional ability or quality of life in general.[27] Disease specific PROMs are usually considered more sensitive to measure change, e.g. after treatment, whereas generic PROMs are more suitable when comparisons across conditions are necessary.

Value-based healthcare

Measuring health outcomes has gained exponential interest through the voice of prof. Michael Porter (Harvard Business School, USA). As a result of rising and unsustainable healthcare costs, variable quality and ineffective competition among healthcare providers and payers, many consider a rigorous transformation of healthcare organization inevitable. Healthcare as we know it is a fee-for-service system; providers are rewarded for quantity of care, such as the number of consultations or procedures completed, irrespective of what those services achieve for patient health. In 2010, Porter proposed the value-based healthcare (VBHC) model: a radically different system in which reimbursements are based on outcomes.[28] In this model, the 'value' of delivered healthcare is determined by achieved patient-relevant outcomes, relative to the costs necessary to deliver those outcomes. In VBHC, the patient perspective is the yardstick of treatment success. For example, increased blood flow after coronary bypass surgery is not considered a successful outcome if the patient's primary concern, pain and shortness of breath during exercise, has not improved. By focusing on such patient-relevant outcomes, VBHC should be profitable for all three pillars of the healthcare system: clinicians can focus on improving outcomes relevant to patients, payers only pay for services that achieve results important to patients, and patients receive better care for the best price.

From theory to practice

Whether VBHC truly effectuates a reduction of expenditures remains to be seen, but even critics agree that VBHC offers credible and legitimate principles of maximizing patient outcomes. Given their dependency on PROMs, specialties like rheumatology (dealing with chronic conditions) or psychiatry (managing mental disorders) have emerged as frontrunners in routinely measuring patient-relevant outcomes on a large-scale basis.[29,30] Furthermore, numerous successful projects, for instance among patients with diabetes, prostate cancer and physiotherapy, have implemented the principles of VBHC to monitor and maximize patient outcomes, build outcome-based registries, create a learning community of providers, and tie reimbursements to achieved outcomes.[31] Governments increasingly acknowledge that apart from using outcome data to decide which healthcare bills to pay and which not to pay, routinely measuring patient-relevant outcomes provides transparency that both patients and providers can use to their advantage. The 'Hoofdlijnenakkoord', a national agreement for specialist medical care in the Netherlands, addresses the importance of healthcare professionals

focusing on outcomes and the development of outcome indicators.[32] In this agreement, all stakeholders support the ambition to define relevant outcomes for 50% of the disease burden by 2022. The overarching goals are to promote benchmarking between health care professionals and institutions, provide better access for patients to relevant and up-to-date outcomes information, and facilitate more outcome-based organization and payment.

Still, significant barriers remain. What exactly are the outcomes that are relevant to patients? How do we measure these outcomes? And how do we use these outcomes to improve healthcare? In this thesis, we attempt to address these questions for patients in need of rhinoplasty.

2. What is rhinoplasty?

Rhinoplasty is a surgical procedure that aims to restore functional or aesthetic deformities of the nose. It ranks among the most frequently performed facial plastic operations, with approximately 33.000 cases performed in the United Kingdom and over 200.000 cases in the United States in 2019.[33,34] The goal of functional rhinoplasty is to correct nasal obstruction caused by collapse or narrowing of the anterior nasal airway, whereas aesthetic rhinoplasty aims to improve nasal shape. Function and appearance of the nose are closely interrelated: aesthetic deformities can cause nasal obstruction and rhinoplasties intended to improve function may alter nasal appearance. Given the wide variation in individual nasal anatomy and the endless number of available interrelated operative techniques to manage similar problems, rhinoplasty is considered one of the most difficult facial plastic surgical procedures.[35] For the surgeon, challenges are to master these techniques and develop an artistic sense of aesthetics to balance the nose to the rest of the face, while improving or at least preserving the nasal airway.

Anatomy of the nose

The external nose consists of a framework of bony and cartilaginous structures, covered by a skin-soft tissue envelope. The bony framework of the nose is formed by paired nasal bones that communicate with the frontal bone cephalically, frontal process of the maxilla laterally, and the upper lateral cartilages caudally. Inferiorly and posteriorly, this bony pyramid fuses with the perpendicular plate of the ethmoid bone, the bony part of the nasal septum. The cartilaginous nasal septum, anterior to its bony counterpart, forms a cartilaginous nasal pyramid together with the upper lateral, or triangular, cartilages. Caudally, the cartilaginous septum rests on the maxillary crest and anterior nasal spine. The lower lateral cartilages, or alar cartilages, form the structural components of the nasal tip (**Figure 1**). The nose contains two nasal valves that create physiological resistance to nasal airflow. The external valve is bounded by the alar lobule laterally, nasal base or sill inferiorly and columella medially. The internal nasal valve is located a few millimeters posterior to the external valve, and is delineated by the septum medially, caudal edge of the upper lateral cartilage superiorly and head of the inferior turbinate inferolaterally (**Figure 2**). The cross-sectional area of the internal nasal valve is the narrowest part of the nasal airway.[36,37]

The nasal airway and nasal obstruction

Nasal airflow is a complex phenomenon. The task of the nasal airway is to heat, humidify and filter inhaled air, as well as direct inhaled air towards the olfactory groove, cephalad in the nasal cavity, to provide smell. Simultaneously, air must be forwarded to the

remaining upper and lower airway. To accomplish these goals, the nasal airway produces two types of airflow. Turbulent airflow, caused by airway resistance, ensures sufficient contact between the mucosa of the nose so that inhaled air can be conditioned, whereas laminar airflow facilitates unobstructed transmission of air to the pulmonary system.[38] To maintain both functions, an intricate balance between turbulent and laminar airflow is necessary. Narrowing of the nasal airway can increase nasal resistance, which may result in decreased laminar airflow. This narrowing can roughly be divided into mucosal causes (e.g. allergic or infectious rhinitis, polyps or tumours of the nasal cavity) and structural causes (i.e. framework pathology of the bony and cartilaginous parts of the nose causing insufficiency of the nasal valves).

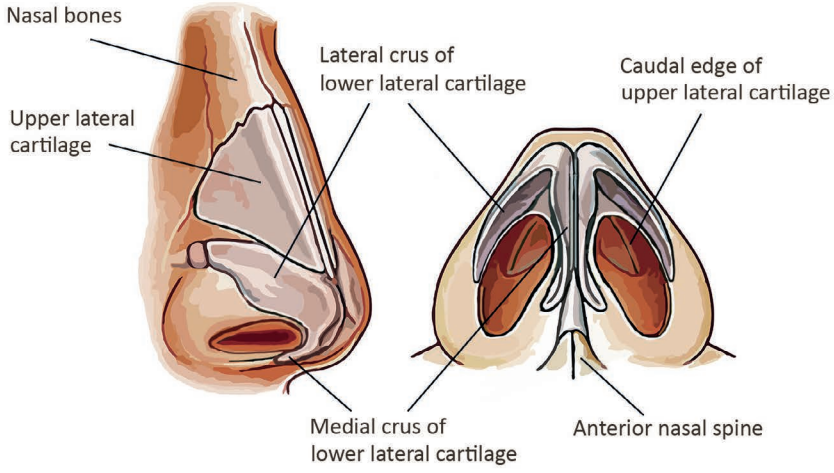


Figure 1. Nasal anatomy.

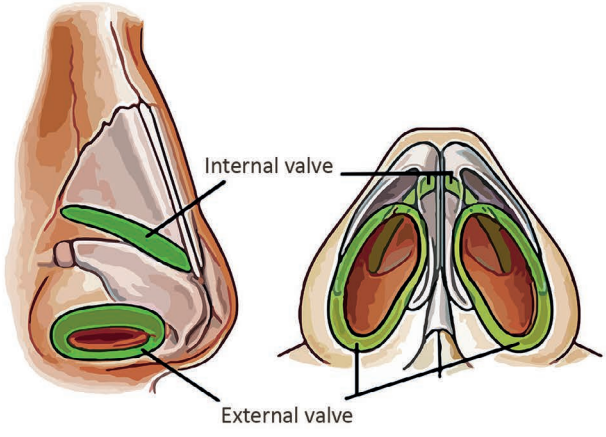


Figure 2. The internal and external nasal valve.

It is important to realize that physiological nasal airflow and the subjective perception of nasal breathing are two different constructs, that do not necessarily correlate with one another. Nasal obstruction is defined as a sensation of insufficient airflow through the nose, and is hence per definition a subjective complaint.[39] Nasal obstruction is not always accompanied by physical parameters of increased nasal airway resistance or decreased peak nasal flow, and reversely, decreased parameters of physiological nasal airflow do not necessarily induce nasal obstruction. Most recent studies suggest that the primary physiological mechanism that produces the sensation of sufficient nasal airflow is nasal mucosal cooling, causing activation of trigeminal thermoreceptors.[40] This would explain, for instance, why menthol spray creates a sensation of improved airflow without affecting nasal airway morphology.[41] A thorough history taking and a detailed examination by an experienced rhinologist remains the key combination to determine whether nasal obstruction is related to a specific anatomical cause.

Impact of nasal obstruction on quality of life

Although not life-threatening, nasal obstruction is a distressing symptom that can have a significant impact on the quality of life. Studies reveal that patients with chronic nasal obstruction, irrespective of the cause, attain lower scores on generic quality of life instruments compared to patients without nasal obstruction.[42,43] The burden of nasal obstruction has been studied extensively in patients with allergic rhinitis, attributing nasal obstruction to impaired daily functioning, adverse emotional and mental effects and sleep disorders.[44-46] For patients with structural causes of nasal obstruction such as turbinate hypertrophy, septal deviation or nasal valve collapse, quality of life is also negatively affected.[47] Given the fact that nasal obstruction is a common complaint in otorhinolaryngology, its impact likely poses a socially relevant health problem. Although there is no data on the economic burden of nasal obstruction itself, costs associated with allergic rhinitis and chronic rhinosinusitis are significant and nasal obstruction is the most frequent and disturbing complaint of these diseases.[48]

Functional rhinoplasty

If nasal obstruction is caused by insufficiency of the nasal valves, functional rhinoplasty may be indicated. Airflow through the narrow nasal valves causes an increase in velocity, leading to a decrease in intraluminal pressure, consistent with Bernoulli's principle.[49] If the cartilages and soft tissues surrounding the nasal valves do not possess sufficient rigidity and strength or are not adequately positioned, the nasal valves may collapse, resulting in obstruction of airflow. Given the anatomical borders of the nasal valves, septal deviations in the nasal valve area and hypertrophic inferior turbinates may also contribute to incompetency of the nasal valves. Accordingly, the goal of functional

rhinoplasty is to restore these structures and their relationships, ultimately leading to improved nasal breathing. A wide variety of surgical options, mainly encompassing autologous cartilage grafting and suture techniques, exist to accomplish this.

Aesthetic rhinoplasty

The nose is one of the most prominent features of the face. Due to its central and protruding position, the nose draws much unconscious attention and is a major determinant of facial attractiveness.[50-52] As a result of its static nature during speech and facial expression (as opposed to eyes, eyebrows and mouth), the nose is the most important feature used to discriminate faces.[53] Concordantly, deformities of the nose tend to attract attention. Pronounced nasal deformities are often conspicuous and negatively affect facial attractiveness. This can have a significant impact on body image and psychological functioning and may even influence the perception of personality characteristics.[54,55] However, the gradual difference between 'normal' and 'deformity' is subtle and what one might interpret as a deformity someone else might perceive as normal. Furthermore, nasofacial aesthetic ideals are culturally variable and time sensitive. Hence, it has proven difficult to express the exact definition of what constitutes an aesthetically pleasing nose. Pre-established aesthetics standards for the nose have primarily been based on classical Greek and Egyptian ideals of beauty, without any support from population-based studies.[56] Remarkably, attractive faces do not resemble these neoclassical canons more than do less attractive faces.[57,58] Even though no exact definition of the ideal, these canons are not rendered useless: the proportions and rules of thumb for nasofacial proportions have proven useful in analyzing the nose and offer rhinoplasty surgeons guidance in developing an artistic sense of aesthetics.[59]

Generally speaking, aesthetic rhinoplasty aims to correct nasal deformities in order to draw attention away from the nose. Given the cosmetic nature of the procedure, the goal of aesthetic rhinoplasty is largely dependent on patient's concerns and expectations, and the aesthetic judgement of the surgeon plays a very pronounced role in determining surgical goals. A satisfactory aesthetic rhinoplasty creates a nose that is aesthetically pleasing, particularly to the patient, without compromising nasal function. Thorough nasofacial analysis is essential to identify deformities and establish goals for surgery. Alterations to the bony or cartilaginous vault can be done to correct nasal crookedness or remove a dorsal hump, and the appearance of the nasal tip can be altered by a wide variation of techniques usually encompassing cartilage resection, transposition, grafting and suturing.

Surgical approach

To access the bony-cartilaginous framework of the nose, a rhinoplasty can be approached in an external (open) approach or endonasal (closed) approach. In the open approach, a small skin incision is made in the columella that extends to incisions on the inside of the nose (marginal incisions). This allows elevation of the complete skin-soft tissue envelope, providing maximum exposure of the cartilaginous and bony structures. In the closed approach, the endonasal incisions provide less surgical exposure, but prevent a visible scar. The choice of approach depends largely on the goal of the rhinoplasty, but preference for an open or closed approach also varies between surgeons. Furthermore the preference for approach appears to be time dependent; In the beginning of the 20th century rhinoplasties were performed almost exclusively using the closed approach, whereas near the end of the 20th century the open approach was popularized, and currently is the most preferred procedure.[60] The wide surgical exposure and direct visualization of anatomical deformities that the open approach provides facilitates precise corrections, and the procedure is technically less challenging compared to the closed approach. In recent years, however, the closed approach is increasingly being advocated by several rhinoplasty surgeons in selected cases in order to preserve the continuity of certain structures, such as the nasal dorsum when correcting dorsal hump deformities.[61]

3. Outcomes in rhinoplasty

Rhinoplasty is both science and craftmanship. There is a tremendous variability between surgeons regarding aesthetic ideals, surgical indications and preferred surgical maneuvers to tackle similar aesthetic and functional nasal problems. Each rhinoplasty surgeon may have his or her own set of techniques that one is entrusted with, usually based on training and experience. Consensus on principles and techniques is limited, as the rhinoplasty literature largely consists of low-level evidence.[62] Surgeons have typically evaluated the effect of a rhinoplasty by examining the nasal airway and comparing pre- and postoperative photographs on an individual patient basis. However, such assessments do not provide quantifiable, patient-relevant and reliable outcome data that are necessary to elevate the rhinoplasty evidence-base from expert opinion to cohort-level studies. Furthermore, outcome data is important to rhinoplasty surgeons to justify the clinical effectiveness of rhinoplasty towards third parties. Health insurance carriers nowadays increasingly demand evidence of effectiveness, and absence of level 1 or 2 evidence can have far stretched consequences, including the decline of expense coverages for functional rhinoplasty. Rhinoplasty surgeons have acknowledged the importance of outcome measurements and as a result, measuring outcomes in rhinoplasty has gained momentum over the past years.[63,64] Various outcome instruments that aim to measure the functional or aesthetic effect of rhinoplasty are currently available.

Instruments measuring functional outcome

Functional rhinoplasty outcome instruments attempt to measure patency of the nasal airway. The measurement of nasal patency is multifaceted: it can refer to a subjective interpretation of a physician examining the nose, it can refer to a patient's subjective evaluation of his or her ability to breathe through the nose, it can refer to cross-sectional measures of the nasal airway or it can refer to dynamic measures of flow, volume or resistance through the nasal cavity. Consequently, outcome instruments may measure any of these constructs. Several objective instruments measuring anatomic or physiologic parameters of the nasal airway exist. Using acoustic rhinometry or imaging studies such as computed tomography (CT) the width of the nasal cavity can be measured, expressed as acoustic reflections or direct measurements of the cross-sectional area of the nasal airway, respectively.[65,66] One of the major limits of these techniques is that they are not able to measure airway narrowing that occurs during inspiration, in case of dynamic lateral nasal wall collapse. Methods such as rhinomanometry or peak nasal inspiratory flow (PNIF) do not possess this limitation. Rhinomanometry measures pressure and flow during in- and expiration through the nose, detecting nasal obstruction by means of increased resistance and pressure or reduced flow (**Figure 3**).[67,68] PNIF measures the

amount of nasal airflow during a deep and quick forced inspiration through the nose.[69] Compared to rhinomanometry, PNIF is significantly easier and cheaper to perform, but disadvantages of PNIF are that it depends on patient cooperation and pulmonary function.[70] A more recent technique, computational fluid dynamics, uses imaging studies to create computational 3-dimensional models of the nasal airway to calculate nasal airflow, resistance and heat flux.[71]

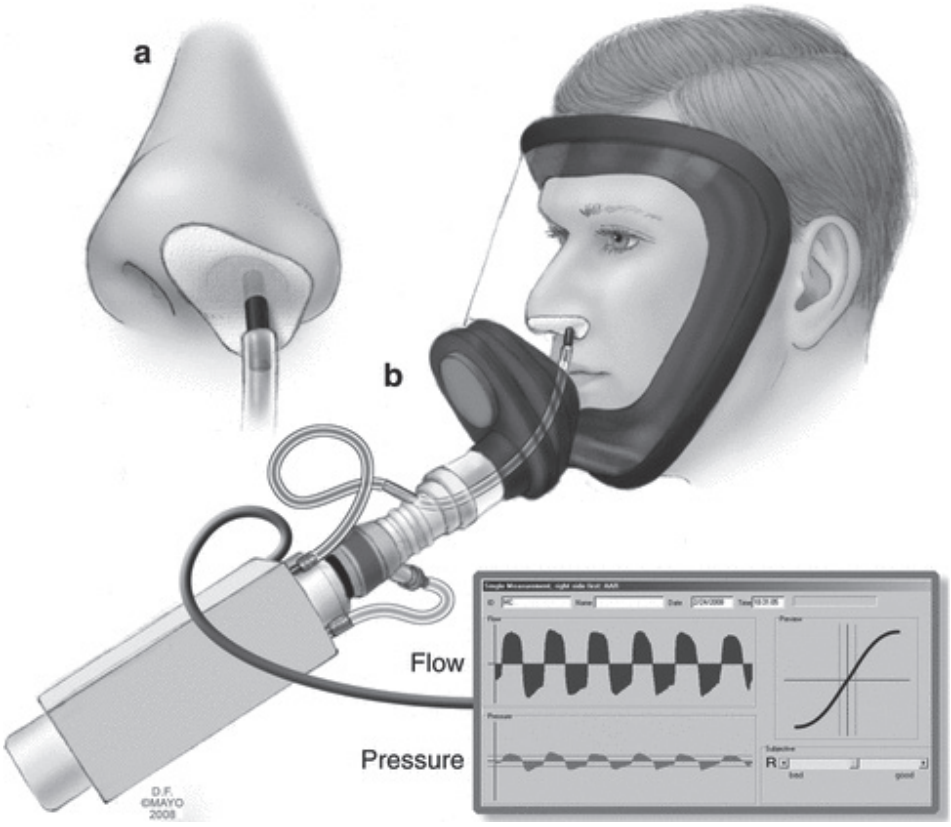


Figure 3. Setup for rhinomanometry. *From: Pallanch J. (2013) Physiology: Rhinomanometry. In: Önerci T. (eds) Nasal Physiology and Pathophysiology of Nasal Disorders. Springer, Berlin, Heidelberg.*

Apart from possible technical or financial issues, a major drawback of objective instruments measuring the nasal airway is that the correlation with patient-reported outcomes is controversial. Although some investigations report significant correlations,[69,72-76] other studies fail to find correlations between objective nasal airway parameters and subjective nasal patency.[77-82] Possibly, the variables that objective instruments measure may not explain the physiological mechanism of nasal airflow sensation. As previously mentioned,

perception of nasal airflow is a complex phenomenon that is not yet fully understood. Therefore, as long as no validated objective clinical instrument that correlates with symptoms exists, routine measurement of objective nasal airway parameters is not advocated.[83] Since the success of functional rhinoplasty is primarily dictated by the patient's perception of nasal patency, patient-reported outcomes that specifically assess nasal patency (disease-specific) are considered the most important to define functional outcome.[84] This is in line with consensus-based clinical practice guidelines for nasal valve insufficiency and rhinoplasty, recommending the use of PROMs, not objective instruments, to routinely measure the benefit of nasal surgery.[85,86]

Various PROMs have been used to evaluate the functional result of rhinoplasty. Whereas some PROMs have been validated extensively with compelling evidence of validity and reliability, the quality of others is questionable. Examples of PROMs that specifically measure nasal patency are the Nasal Obstruction Symptom Evaluation (NOSE) scale, Non-invasive Assessment and Symptomatic Improvement of the Obstructed Nose (NASION), Standardized Cosmesis and Health Nasal Outcomes Survey (SCHNOS) functional scale, Rhinologic Quality of Life (RhinoQol) and Sino Nasal Outcome Test (SNOT-22).[87-91] In addition, several authors have attempted to measure the effect of functional rhinoplasty on generic health-related quality of life, for instance using the Short Form (SF-12 or SF-36), General Health Questionnaire (GHQ-28), or EuroQoL-5 dimensions.[92-94]

Although consensus exists that PROMs are the outcome instrument of choice to measure the functional effect of rhinoplasty, there is no consensus on which of these PROM to use. Considering that numerous PROMs measuring functional outcome in rhinoplasty are available, with variable quality and without consensus on which of these PROMs to use, two problems arise. First, research performed with poor-quality outcome instruments is considered a waste of resources.[95,96] Measuring treatment effects using outcome instruments with insufficient quality may produce meaningless or unreliable data and should be avoided. Second, the arbitrary use of different outcome instrument hampers the use of one universal 'language' to measure a specific outcome, thus limiting the aggregation and comparison of outcome data. To select the best outcome instrument, a comprehensive systematic review that evaluates the quality of all available instruments is required. To date, such a review has not been conducted.

Instruments measuring aesthetic outcome

As with functional outcomes, there is a wide variety of different types of outcomes that have been measured to evaluate the aesthetic effect of rhinoplasty. Traditionally, aesthetic rhinoplasty outcomes have been evaluated using pre- and postoperative photographs, in which the interpretation or grading of results was primarily surgeon-

based.[97,98] However, given the cosmetic nature of the procedure, the most important parameter of success in aesthetic rhinoplasty is patient satisfaction. The first attempts to measure patient satisfaction with nasal appearance involved the use of ad-hoc questionnaires that fell short of proper development or validation.[99,100] In 2001, the first validated patient-reported disease-specific outcome instrument specifically designed for aesthetic rhinoplasty, the Rhinoplasty Outcome Evaluation, was published.[101] This 6-item questionnaire evaluates patient satisfaction, confidence and social acceptance related to nasal appearance. Examples of other disease-specific questionnaires that are currently available are the Utrecht Questionnaire, assessing concern with body image in relation to nasal appearance, and the FACE-Q Rhinoplasty Module, assessing patient satisfaction with several dimensions of shape (e.g. size, width, straightness).[102,103] Since it has long been recognized that psychological disturbances are relatively frequent in patients seeking aesthetic rhinoplasty[104-106], another outcome assessed in aesthetic rhinoplasty is psychosocial functioning or body image. Examples of such questionnaires are the Derriford Appearance Scale (DAS-24 or DAS-59), Multidimensional Body-Self Relations Questionnaire or Rosenberg Self Esteem Scale.[107-111] Also generic health or quality-of-life instruments such as the General Health Questionnaire, EuroQoL and Short Form (SF-12 or SF-36) have been used to measure the effect of aesthetic rhinoplasty.[110-112] Similar to functional PROMs, the quality of these aesthetic PROMs is variable and no consensus exists on which of these PROMs to use to measure the aesthetic result of a rhinoplasty.

Although patient satisfaction remains the major determinant of aesthetic rhinoplasty success, patient-reported outcomes are subjective in nature. Scores may be influenced by confounding factors not under the control of the surgeon. An objective measure ignores these influences and could therefore complement patient-reported outcomes. Numerous papers report on the objective quantification of changes in nasal shape after surgery. Traditionally, such quantifications encompassed the use of a caliper or ruler directly to the patient's skin to measure anatomical proportions and angles of the nose.[113,114] Since photography has become the mainstay of facial analysis, the use of calipers has largely been abandoned. Particularly the availability of digital photography and imaging software has enabled detailed and accurate analyses of nasal shape. Several authors describe the use of computer-assisted measurements to quantify postoperative changes after rhinoplasty, for instance regarding nasal tip projection, dorsal aesthetic lines or dorsal width.[115-117] Following the notion that a nose should be interpreted as a 3-dimensional rather than a 2D structure, 3D-photography and optical surface scanners have recently been utilized to quantify differences after rhinoplasty (**Figure 4**).[118-123] What is often lacking in order to qualify these instruments as outcome measures, however, is the definition of 'ideal'. Measuring change alone is

not a parameter of surgical success: it requires an interpretation of how successful that change was in order to function as an outcome instrument. Given the importance of patient satisfaction in the assessment of success, an objective outcome instrument would have to correlate with subjective (patient-reported) outcomes. To date, only specific nasal ratios or parameters have been studied in relation to subjective outcomes; whereas some authors have found correlations[124], others did not encounter correlations between nasal ratios and subjective facial attractiveness.[125] No studies exist objectively capturing the nose as a whole and correlating this measurement to subjective outcomes.

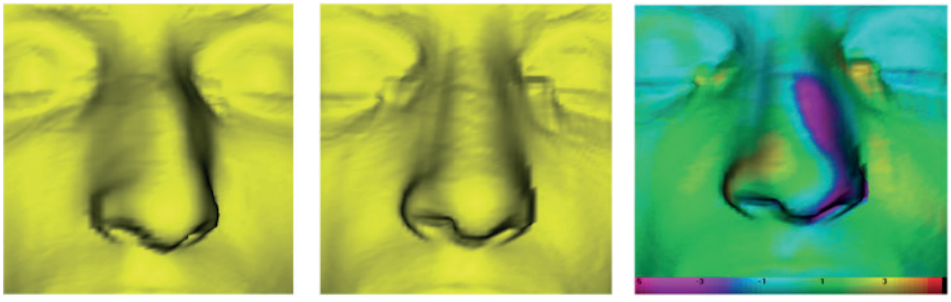


Figure 4. Optical surface scanner. *Left,* preoperative optical surface scan; *center*, postoperative optical surface scan; *right*, difference in volume from pre- to postoperative scan. Warm colours (green to red) represent positive surface displacement (in millimeters), cold colours (blue to pink) represent negative displacement. Turquoise areas represent surfaces that have not changed. *From: Chau et al, Use of an optical surface scanner in assessment of outcome following rhinoplasty surgery. J Laryngol Otol, 122(9);972-977, 2008.*

4. Aims of this thesis

Systematically collecting relevant outcomes generates valuable data that allow patients, surgeons and other stakeholders involved in rhinoplasty to make informed decisions. However, routinely measuring outcomes in rhinoplasty is currently hampered by large heterogeneity of available outcome instruments and lack of know-how on the implementation and analysis of outcome measurements. The goals of this thesis are to (1) provide an evidence-based recommendation on which rhinoplasty outcome instruments to use, (2) develop a methodology to implement routine outcome measurements into daily practice, and (3) explore additional advantages resulting from such measurements. Furthermore, this thesis attempts to quantify nasal attractiveness, which could aid in objectively measuring aesthetic rhinoplasty outcome and complement patient-reported outcomes.

To improve the quality of rhinoplasty outcome data and counter heterogeneity in outcome reporting, the goal of **chapter 2** is to select the most suitable instrument to measure functional and aesthetic outcomes after rhinoplasty. Using a consensus-based methodology, the measurement properties of all PROMs that measure functional or aesthetic outcomes are critically appraised, summarized and compared. The result is an evidence-based recommendation on the most suitable instruments. In addition, promising instruments worthy of further research are identified.

An important aspect of using a PROM is that the instrument of choice is linguistically and culturally adapted to the target audience. An accepted PROM to quantify subjective nasal obstruction is the NOSE scale, published in 2004 by Stewart et al.[87] The NOSE scale is a valid, reliable and short PROM that measures the severity and impact of nasal obstruction. Although the NOSE scale has been validated in many languages, no validated Dutch version of the scale is available. **Chapter 3** aims to translate and validate the NOSE scale into the Dutch language.

Although the theoretical value of outcome data is evident for many, the anticipated difficulty of integrating routine outcome measurements in the clinical environment has been a common obstacle for health care providers, including rhinoplasty surgeons. Some of the suggested barriers include the workload associated with routinely collecting outcomes, the lack of a protocol on the data collection process (including patient eligibility, timing, and frequency of administration) and missing know-how on how to correctly analyze and interpret the data.[126,127] In **chapter 4**, the functionality of a short and practical rhinoplasty outcome routine is investigated,

in an attempt to demonstrate that routinely measuring rhinoplasty outcomes in a busy practice is feasible and beneficial. **Chapter 5** aims to discuss advantages of routine outcome measurements in secondary cleft lip rhinoplasty.

A particular advantage of routine outcome measurements is evaluating performance of the individual rhinoplasty surgeon. Self-assessment provides valuable feedback in the life-long process of mastering rhinoplasty, but means to measure and evaluate personal performance are scarce. Rhinoplasty surgeons typically self-reflect on an ad-hoc basis using pre- and postoperative photographs. Such evaluations are useful, but also subjective and difficult to quantify. Instead, datasets with patient-relevant surgical outcomes of all encountered patients would facilitate a substantially more accurate and data-based reflection of surgical performance. The aim of **chapter 6**, therefore, is to demonstrate the self-evaluative potential of routine outcome measurements in rhinoplasty.

The final aim of this thesis is to lay a basis for an objective measure to evaluate aesthetic outcome. Developing an objective aesthetic outcome instrument requires a definition and objective quantification of ideal nasal shape. Based on previous research on facial attractiveness, **chapter 7** explores the role of averageness on nasal attractiveness. In this experimental study the correlation between mathematical averageness of nasal shape and subjective attractiveness is investigated, in an attempt to identify objectively measurable traits of appealing noses.

In **chapter 8**, the main findings of this thesis are discussed, methodological considerations are addressed, and suggestions for future research are raised. A summary of the research presented in this thesis is provided in English and in Dutch in **chapter 9**.

References

1. Desomer A, van den Heede K, Triemstra M, et al. Use of Patient-Reported Outcome and Experience Measures in Patient Care and Policy. 2018. Available at: https://www.nivel.nl/sites/default/files/bestanden/KCE_use_of_PROM_PREM.pdf. Accessed April 1, 2021
2. Bruins B. Ministry of Health Welfare and Sport. Letter to House of Representatives: 'Outcome based healthcare 2018-2022'. July 2, 2018. Available at: https://www.government.nl/binaries/government/documents/letters/2018/07/02/letter-to-parliament-about-outcome-based-healthcare-2018-2022/Letter+to+the+Parliament+Outcome+based+healthcare+2018-2022.pdf. Accessed August 19, 2020
3. Coles TM, Hernandez AF, Reeve BB, et al. Enabling patient-reported outcome measures in clinical trials, exemplified by cardiovascular trials. Health Qual Life Outcomes 2021;19(1):164
4. Sepucha K, Mulley Jr AG. A perspective on the patient's role in treatment decisions. Med Care Res Rev 2009;66(1 Suppl):53S-74S
5. Stand van de patient. Patiëntenfederatie Nederland, 2018 (in Dutch). Available at: https://www.patientenfederatie.nl/downloads/brochures/176-stand-van-de-patient/file. Accessed April 1, 2021
6. Price Transparency and Physician Quality Report Card 2017 - Report Altarum's Center for Payment Innovation. Available at: https://altarum.org/sites/default/files/uploaded-publication-files/2017%20Price%20Transparency%20and%20Physician%20Quality%20Report%20Card_0.pdf Accessed April 1, 2021
7. Noonan VK, Lyddiatt A, Ware P, et al. Montreal Accord on Patient-Reported Outcomes (PROs) use series - Paper 3: patient-reported outcomes can facilitate shared decision-making and guide self-management. J Clin Epidemiol 2017;89:125-135
8. Ranganathan M, Hibbard J, Rodday AM, et al. Motivating public use of physician-level performance data: an experiment on the effects of message and mode. Med Care Res Rev 2009;66(1):68-81
9. Yegian JM, Dardess P, Shannon M, Carman KL. Engaged patients will need comparative physician-level quality data and information about their out-of-pocket costs. Health Aff (Millwood) 2013;32(2):328-337
10. Rechel B, McKee M, Haas M, et al. Public reporting on quality, waiting times and patient experience in 11 high-income countries. Health Policy 2016;120(4):377-383
11. Trebble TM, Carder C, Paul M, et al. Determining doctors' views on performance measurement and management of their clinical practice. Future Hosp J 2015;2(3):166-170
12. de Boer TA, Gietelink DA, Vierhout ME. Discrepancies between physician interview and a patient self-assessment questionnaire after surgery for pelvic organ prolapse. Int Urogynecol J Pelvic Floor Dysfunct 2008;19(10):1349-1352
13. Walfish S, McAlister B, O'Donnell P, Lambert MJ. An investigation of self-assessment bias in mental health providers. Psychol Rep 2012;110(2):639-644
14. Jones DS, Podolsky SH. The history and fate of the gold standard. Lancet 2015;385(9977):1502-1503
15. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. BMJ 2002;324(7351):1448-1451
16. Kent DM, Nelson J, Dahabreh IJ, et al. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. Int J Epidemiol 2016;45(6):2075-2088
17. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996;312(7040):1215-1218

18. Wahmann MS, Bulut OC, Bran GM, Veit JA, Riedel F. Systematic Review of Quality-of-Life Measurement After Aesthetic Rhinoplasty. Aesthetic Plast Surg 2018 Dec;42(6):1635-1647
19. Spataro E, Most SP. Measuring Nasal Obstruction Outcomes. Otolaryngol Clin North Am 2018 Oct;51(5):883-895
20. Aravind M, Chung KC. Evidence-based medicine and hospital reform: tracing origins back to Florence Nightingale. Plast Reconstr Surg 2010;125(1):403-409
21. Kaska SC, Weinstein JN. Historical perspective. Ernest Amory Codman, 1869-1940. A pioneer of evidence-based medicine: the end result idea. Spine (Phila Pa 1976) 1998;23(5):629-633
22. Relman AS. Assessment and accountability: the third revolution in medical care. N Engl J Med 1988;319(18):1220-1222
23. Donabedian A. Evaluating the quality of medical care. Milbank Mem Fund Q 1966;44(3):Suppl:166-206
24. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA 1995;273(1):59-65
25. Laine C, Davidoff F. Patient-centered medicine. A professional evolution. JAMA 1996;275(2):152-156
26. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2(14):i-iv, 1-74
27. Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013;346:f167
28. Porter ME. What is value in health care? N Engl J Med 2010;363(26):2477-2481
29. Roe D, Mazor Y, Gelkopf M. Patient-reported outcome measurements (PROMs) and provider assessment in mental health: a systematic review of the context of implementation. Int J Qual Health Care 2021;33(1)
30. Oude Voshaar MAH, Das Gupta Z, Bijlsma JWJ, et al. International Consortium for Health Outcome Measurement Set of Outcomes That Matter to People Living With Inflammatory Arthritis: Consensus From an International Working Group. Arthritis Care Res (Hoboken) 2019;71(12):1556-1565
31. EIT Health. Implementing Value-Based Health Care in Europe: Handbook for Pioneers (Director: Gregory Katz), 2020.
32. Bestuurlijk akkoord medisch-specialistische zorg 2019 t/m 2022 (in Dutch). Available at: https://www.rijksoverheid.nl/documenten/brieven/2018/06/04/bestuurlijk-akkoord-medisch-specialistische-zorg-2019-t-m-2022. Accessed November 11, 2020
33. Hospital Episode Statistics, NHS Digital. Hospital Admitted Patient Care Activity, 2019-2020: Procedures and interventions. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2019-20. Accessed December 21, 2020
34. American Society of Plastic Surgeons. National Clearinghouse of Plastic Surgery Procedural Statistics. 2019 Plastic Surgery Statistics Report. Available at: https://www.plasticsurgery.org/documents/News/Statistics/2019/plastic-surgery-statistics-full-report-2019.pdf Accessed December 21, 2020
35. Lohuis PJFM. Advanced Caucasian and Mediterranean Rhinoplasty. In. Amsterdam: Kugler Publications; 2014:2-19
36. Bridger GP. Physiology of the nasal valve. Arch Otolaryngol 1970;92(6):543-553
37. Grymer LF, Hilberg O, Pedersen OF, Rasmussen TR. Acoustic rhinometry: values from adults with subjective normal nasal patency. Rhinology 1991;29(1):35-47
38. Ballert JA, Park SS. Functional considerations in revision rhinoplasty. Facial Plast Surg 2008;24(3):348-357

39. Jessen M, Malm L. Definition, prevalence and development of nasal obstruction. Allergy 1997;52(40 Suppl):3-6
40. Sozansky J, Houser SM. The physiological mechanism for sensing nasal airflow: a literature review. Int Forum Allergy Rhinol 2014;4(10):834-838
41. Eccles R, Jones AS. The effect of menthol on nasal resistance to air flow. J Laryngol Otol 1983;97(8):705-709
42. Udaka T, Suzuki H, Kitamura T, et al. Relationships among nasal obstruction, daytime sleepiness, and quality of life. Laryngoscope 2006;116(12):2129-2132
43. Gadkaree SK, Fuller JC, Justicz NS, et al. Health Utility Values as an Outcome Measure in Patients Undergoing Functional Septorhinoplasty. JAMA Facial Plast Surg 2019;21(5):381-386
44. Shedden A. Impact of nasal congestion on quality of life and work productivity in allergic rhinitis: findings from a large online survey. Treat Respir Med 2005;4(6):439-446
45. Gurevich F, Glass C, Davies M, et al. The effect of intranasal steroid budesonide on the congestion-related sleep disturbance and daytime somnolence in patients with perennial allergic rhinitis. Allergy Asthma Proc 2005;26(4):268-274
46. Bousquet PJ, Demoly P, Devillier P, Mesbah K, Bousquet J. Impact of allergic rhinitis symptoms on quality of life in primary care. Int Arch Allergy Immunol 2013;160(4):393-400
47. Rhee JS, Book DT, Burzynski M, Smith TL. Quality of life assessment in nasal airway obstruction. Laryngoscope 2003;113(7):1118-1122
48. Stewart M, Ferguson B, Fromer L. Epidemiology and burden of nasal congestion. Int J Gen Med 2010;3:37-45
49. Bloching MB. Disorders of the nasal valve area. GMS Curr Top Otorhinolaryngol Head Neck Surg 2007;6:Doc07
50. Hsiao JH, Cottrell G. Two fixations suffice in face recognition. Psychol Sci 2008;19(10):998-1006
51. Jones D. Sexual selection, physical attractiveness, and facial neoteny. Current Anthropology 1995;36:723-748
52. Meerdink JE, Garbin CP, Leger DW. Cross-gender perceptions of facial attributes and their relation to attractiveness: do we see them differently than they see us? Percept Psychophys 1990;48(3):227-233
53. Logan AJ, Gordon GE, Loffler G. Contributions of individual face features to face discrimination. Vision Res 2017;137:29-39
54. van Schijndel O, Tasman AJ, Litschel R. The Nose Influences Visual and Personality Perception. Facial Plast Surg 2015;31(5):439-445
55. Andretto Amodeo C. The central role of the nose in the face and the psyche: review of the nose and the psyche. Aesthetic Plast Surg 2007;31(4):406-410
56. Farkas LG, Hreczko TA, Kolar JC, Munro IR. Vertical and horizontal proportions of the face in young adult North American Caucasians: revision of neoclassical canons. Plast Reconstr Surg 1985;75(3):328-338
57. Torsello F, Mirigliani L, D'Alessio R, Deli R. Do the neoclassical canons still describe the beauty of faces? An anthropometric study on 50 Caucasian models. Prog Orthod 2010;11(1):13-19
58. Farkas LG, Kolar JC. Anthropometrics and art in the aesthetics of women's faces. Clin Plast Surg 1987;14(4):599-616
59. Larrabee Jr. WF. Facial analysis for rhinoplasty. Otolaryngol Clin North Am 1987;20(4):653-674
60. Berghaus A. Modern Rhinoplasty: Is There a Place for the Closed Approach? Facial Plast Surg 2016;32(4):402-408

61. Saban Y, Daniel RK, Polselli R, Trapasso M, Palhazi P. Dorsal Preservation: The Push Down Technique Reassessed. Aesthet Surg J 2018;38(2):117-131
62. Eggerstedt M, Brown HJ, Shay AD, et al. Level of Evidence in Facial Plastic Surgery Research: A Procedure-Level Analysis. Aesthetic Plast Surg 2020;44(5):1531-1536
63. Lee MK, Most SP. Evidence-Based Medicine. Rhinoplasty. Facial Plastic Surgery Clinics of North America 2015;23(3):303-312
64. Wahmann MS, Bulut OC, Bran GM, Veit JA, Riedel F. Systematic Review of Quality-of-Life Measurement After Aesthetic Rhinoplasty. Aesthetic Plast Surg 2018;42(6):1635-1647
65. Poetker DM, Rhee JS, Mocan BO, Michel MA. Computed tomography technique for evaluation of the nasal valve. Arch Facial Plast Surg 2004;6(4):240-243
66. Hilberg O. Objective measurement of nasal airway dimensions using acoustic rhinometry: methodological and clinical aspects. Allergy 2002;57 Suppl 70:5-39
67. Kern EB. Standardization of rhinomanometry. Rhinology 1977;15(3):115-119
68. Clement PA, Gordts F, Standardisation Committee on Objective Assessment of the Nasal Airway IRS, Ers. Consensus report on acoustic rhinometry and rhinomanometry. Rhinology 2005;43(3):169-179
69. Cuddihy PJ, Eccles R. The use of nasal spirometry as an objective measure of nasal septal deviation and the effectiveness of septal surgery. Clin Otolaryngol Allied Sci 2003;28(4):325-330
70. 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
71. Kimbell JS, Garcia GJM, Frank DO, et al. Computed nasal resistance compared with patient-reported symptoms in surgically treated nasal airway passages: A preliminary report. Am J Rhinol and Allergy 2012;26(3):e94-e98
72. Menger DJ, Swart KMA, Nolst Trenité GJ, Georgalas C, Grolman W. Surgery of the external nasal valve: The correlation between subjective and objective measurements. Clin Otolaryngol 2014;39(3):150-155
73. Murrell GL. Correlation between subjective and objective results in nasal surgery. Aesthet Surg J 2014;34(2):249-257
74. Tompos T, Garai T, Zemplén B, Gerlinger I. Sensation of nasal patency compared to rhinomanometric results after septoplasty. Eur Arch Otorhinolaryngol 2010;267(12):1887-1891
75. Kimbell JS, Frank DO, Laud P, Garcia GJM, Rhee JS. Changes in nasal airflow and heat transfer correlate with symptom improvement after surgery for nasal obstruction. Journal of Biomechanics 2013;46(15):2634-2643
76. Kjaergaard T, Cvancarova M, Steinsvag SK. Does nasal obstruction mean that the nose is obstructed? Laryngoscope 2008;118(8):1476-1481
77. Kahveci OK, Miman MC, Yucel A, et al. The efficiency of Nose Obstruction Symptom Evaluation (NOSE) scale on patients with nasal septal deviation. Auris Nasus Larynx 2012;39(3):275-279
78. Toyserkani NM, Frisch T, Von Buchwald C. Postoperative improvement in acoustic rhinometry measurements after septoplasty correlates with long-term satisfaction. Rhinology 2013;51(2):171-175
79. Dadgarnia MH, Baradaranfar MH, Mazidi M, Meibodi SMRA. Assessment of septoplasty effectiveness using acoustic rhinometry and rhinomanometry. Iranian Journal of Otorhinolaryngology 2013;25(71):71-78
80. Sánchez HL, Nuño CA, Sañudo EGC, et al. Correlation between rhinomanometry and nose scale and measurement of therapeutic success in nasal obstruction. Otolaryngol Head Neck Surg 2014;151(1):P115-P116

81. Lam DJ, James KT, Weaver EM. Comparison of anatomic, physiological, and subjective measures of the nasal airway. Am J Rhinol 2006;20(5):463-470
82. Andrews PJ, Choudhury N, Takhar A, et al. The need for an objective measure in septorhinoplasty surgery: are we any closer to finding an answer? Clin Otolaryngol 2015;40(6):698-703
83. Andre RF, Vuyk HD, Ahmed A, Graamans K, Nolst Trenite GJ. Correlation between subjective and objective evaluation of the nasal airway. A systematic review of the highest level of evidence. Clin Otolaryngol 2009;34(6):518-525
84. Cannon DE, Rhee JS. Evidence-based practice: functional rhinoplasty. Otolaryngol Clin North Am 2012;45(5):1033-1043
85. Rhee JS, Weaver EM, Park SS, et al. Clinical consensus statement: Diagnosis and management of nasal valve compromise. Otolaryngol Head Neck Surg 2010;143(1):48-59
86. Ishii LE, Tollefson TT, Basura GJ, et al. Clinical Practice Guideline: Improving Nasal Form and Function after Rhinoplasty Executive Summary. Otolaryngol Head Neck Surg 2017;156(2):205-219
87. Stewart MG, Witsell DL, Smith TL, et al. Development and validation of the Nasal Obstruction Symptom Evaluation (NOSE) scale. Otolaryngol Head Neck Surg 2004;130(2):157-163
88. Nouraei SA, Virk JS, Kanona H, et al. Non-invasive Assessment and Symptomatic Improvement of the Obstructed Nose (NASION): a physiology-based patient-centred approach to treatment selection and outcomes assessment in nasal obstruction. Clin Otolaryngol 2016;41(4):327-340
89. Moubayed SP, Ioannidis JPA, Saltychev M, Most SP. The 10-Item Standardized Cosmesis and Health Nasal Outcomes Survey (SCHNOS) for Functional and Cosmetic Rhinoplasty. JAMA Facial Plast Surg 2017
90. Poirrier AL, Ahluwalia S, Goodson A, et al. Is the Sino-Nasal Outcome Test-22 a suitable evaluation for septorhinoplasty? Laryngoscope 2013;123(1):76-81
91. Atlas SJ, Gallagher PM, Wu YA, et al. Development and validation of a new health-related quality of life instrument for patients with sinusitis. Qual Life Res 2005;14(5):1375-1386
92. Phillips PS, Stow N, Timperley DG, et al. Functional and cosmetic outcomes of external approach septoplasty. Am J Rhinol and Allergy 2011;25(5):351-357
93. Arunachalam PS, Kitcher E, Gray J, Wilson JA. Nasal septal surgery: Evaluation of symptomatic and general health outcomes. Clin Otolaryngol Allied Sci 2001;26(5):367-370
94. Fuller JC, Levesque PA, Lindsay RW. Assessment of the EuroQol 5-Dimension Questionnaire for Detection of Clinically Significant Global Health-Related Quality-of-Life Improvement Following Functional Septorhinoplasty. JAMA Facial Plast Surg 2017;19(2):95-100
95. Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 2014;383(9912):166-175
96. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009;374(9683):86-89
97. Anderson JR. A scale for evaluating results of rhinoplasty. Arch Otolaryngol Head Neck Surg 1985;111(8):520-523
98. Mahler D, Moses S, Last U. A measuring scale for objective evaluation of the nasal shape. Aesthetic Plast Surg 1983;7(4):223-226
99. Guyuron B, Bokhari F. Patient satisfaction following rhinoplasty. Aesthetic Plast Surg 1996;20(2):153-157
100. Dziewulski P, Dujon D, Spyriounis P, Griffiths RW, Shaw JD. A retrospective analysis of the results of 218 consecutive rhinoplasties. Br J Plast Surg 1995;48(7):451-454
101. Alsarraf R, Larrabee Jr WF, Anderson S, Murakami CS, Johnson Jr CM. Measuring cosmetic facial plastic surgery outcomes: a pilot study. Arch Facial Plast Surg 2001;3(3):198-201

102. Lohuis PJ, Hakim S, Duivesteijn W, Knobbe A, Tasman AJ. Benefits of a short, practical questionnaire to measure subjective perception of nasal appearance after aesthetic rhinoplasty. Plast Reconstr Surg 2013;132(6):913e-923e
103. Pusic AL, Klassen AF, Scott AM, Cano SJ. Development and Psychometric Evaluation of the FACE-Q Satisfaction with Appearance Scale. A New Patient-Reported Outcome Instrument for Facial Aesthetics Patients. Clinics in Plastic Surgery 2013;40(2):249-260
104. Hay GG. Psychiatric aspects of cosmetical nasal operations. Br J Psychiatry 1970;116:85-97
105. Wright MR. A psychological study of patients undergoing cosmetic surgery. Arch Otolaryngol 1975;101:145-151
106. Linn L, Goldman IB. Psychiatric observations concerning rhinoplasty. Psychosom Med 1949;11:307-315
107. Gunel C, Omurlu IK. The effect of rhinoplasty on psychosocial distress level and quality of life. Eur Arch Otorhinolaryngol 2015 Aug;272(8):1931-5
108. Litner JA, Rotenberg BW, Dennis M, Adamson PA. Impact of cosmetic facial surgery on satisfaction with appearance and quality of life. Arch Facial Plast Surg 2008;10(2):79-83
109. Cingi C, Songu M, Bal C. Outcomes research in rhinoplasty: Body image and quality of life. Am J Rhinol and Allergy 2011;25(4):263-267
110. Klassen A, Jenkinson C, Fitzpatrick R, Goodacre T. Patients' health related quality of life before and after aesthetic surgery. Br J Plast Surg 1996;49(7):433-438
111. Fatemi MJ, Rajabi F, Moosavi SJ, Soltani M. Quality of life among Iranian adults before and after rhinoplasty. Aesthetic Plast Surg 2012;36(2):448-452
112. Cingi C, Eskiizmir G. Deviated nose attenuates the degree of patient satisfaction and quality of life in rhinoplasty: A prospective controlled study. Clin Otolaryngol 2013;38(2):136-141
113. Hormozi AK, Toosi AB. Rhinometry: An important clinical index for evaluation of the nose before and after rhinoplasty. Aesthetic Plast Surg 2008;32(2):286-293
114. Petroff MA, McCollough EG, Hom D, Anderson JR. Nasal tip projection. Quantitative changes following rhinoplasty. Arch Otolaryngol Head Neck Surg 1991;117(7):783-788
115. Ingels K, Orhan KS. Measurement of preoperative and postoperative nasal tip projection and rotation. Arch Facial Plast Surg 2006;8(6):411-415
116. Mojallal A, Ouyang D, Saint-Cyr M, et al. Dorsal aesthetic lines in rhinoplasty: a quantitative outcome-based assessment of the component dorsal reduction technique. Plast Reconstr Surg 2011;128(1):280-288
117. Kortbus MJ, Ham J, Fechner F, Constantinides M. Quantitative analysis of lateral osteotomies in rhinoplasty. Arch Facial Plast Surg 2006;8(6):369-373
118. Chau H, Dasgupta R, Sauret V, Kenyon G. Use of an optical surface scanner in assessment of outcome following rhinoplasty surgery. J Laryngol Otol 2008;122(9):972-977
119. Largo RD, Wettstein R, Fulco I, et al. Three-dimensional laser surface scanning in rhinosurgery. Facial Plast Surg 2013;29(2):116-120
120. Szychta P, Witmanowski H, Rykala J. Assessment of the usefulness of three-dimensional scanner in aesthetic evaluation of post-traumatic rhinoplasty. Journal of Plastic Surgery and Hand Surgery 2013;47(2):106-112
121. Toriumi DM, Dixon TK. Assessment of Rhinoplasty Techniques by Overlay of Before-and-After 3D Images. Facial Plastic Surgery Clinics of North America 2011;19(4):711-723
122. Van Heerbeek N, Ingels KJAO, Van Loon B, Plooij JM, Bergé SJ. Three dimensional measurement of rhinoplasty results. Rhinology 2009;47(2):121-125

123. Bared A. Lower lateral cartilage repositioning: objective analysis using 3-dimensional imaging. JAMA Facial Plast Surg 2014;16(4):261-267
124. Ozturk K, Gode S, Karahan C, Midilli R. Assessing the rhinoplasty outcome: inter-rater variability of aesthetic perception in the light of objective facial analysis. Eur Arch Otorhinolaryngol 2015 Dec;272(12):3709-13
125. Devcic Z, Rayikanti BA, Hevia JP, et al. Nasal tip projection and facial attractiveness. Laryngoscope 2011;121(7):1388-1394
126. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. BMJ Qual Saf 2014;23(6):508-518
127. Philpot LM, Barnes SA, Brown RM, et al. Barriers and Benefits to the Use of Patient-Reported Outcome Measures in Routine Clinical Care: A Qualitative Study. Am J Med Qual 2018;33(4):359-364

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Chapter 2

Evaluation of measurement properties of patient reported outcome measures in rhinoplasty: a systematic review

F.V.W.J. van Zijl MD, L.B. Mokkink PhD, J.A. Haagsma PhD, F.R. Datema MD PhD

JAMA Facial Plast Surg. 2019 Mar 1;21(2):152-162.

Abstract

Importance: The number of available rhinoplasty outcome measurement instruments has increased rapidly over the past years. A large heterogeneity of instruments of different quality now exists, causing difficulty in pooling and comparing outcome data.

Objective: To critically appraise, summarize and compare the measurement properties of all patient reported outcome instruments (PROMs) that measure functional or aesthetic symptoms of patients undergoing rhinoplasty, using consensus-based methodology and guidelines. This facilitates an evidence-based recommendation on the most suitable instrument to measure rhinoplasty outcomes and identifies promising instruments worthy of further research.

Evidence Review: A systematic literature search of EMBASE, Medline and Web-of-Science was conducted from the databases' respective inception dates to May 18th, 2018. Thirty-three articles evaluating one or more measurement properties of instruments measuring symptoms related to nasal breathing or satisfaction with nasal appearance in patients who had undergone septoplasty and/or rhinoplasty were included. Measurement properties were graded according to the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines for systematic reviews of PROMs.

Findings: The search strategy identified 33 studies that used 12 different instruments. In general, high-quality studies on measurement properties of instruments measuring aesthetic and/or functional symptom-specific outcome of rhinoplasty are scarce. The Nasal Obstruction Symptom Evaluation (NOSE) scale demonstrated high-quality evidence for sufficient structural validity, internal consistency, reliability, construct validity and responsiveness, along with favorable interpretability and feasibility aspects, and was therefore selected as most suitable instrument to measure functional outcome. Amongst instruments measuring aesthetic outcome, the FACE-Q and Standardized Cosmesis and Health Nasal Outcomes Survey (SCHNOS) aesthetic subscale are recommended for further study. Future studies on the measurement properties of the identified PROMs, in particular content validity studies, are necessary.

Conclusions and Relevance: Three instruments with high potential for further use were identified in a systematic review of rhinoplasty outcome instruments using a standardized, consensus-based methodology: the NOSE, FACE-Q and SCHNOS. These findings may contribute to standardized collection of outcome data in rhinoplasty.

Introduction

The number of available rhinoplasty outcome measurement instruments has increased rapidly in recent years.[1,2] As a result, great heterogeneity and a lack of standardization of rhinoplasty outcome measurement exists. The arbitrary use of instruments of different quality to measure different outcomes causes difficulties in pooling and comparing outcome data and hinders the transparent communication of outcomes to patients and third parties in a time of need.

Although rhinoplasty outcome research is generally accepted to encompass the measurement of treatment effectiveness in functional and aesthetic terms, it can be expressed in numerous ways. Examples are anatomical and physiological factors (e.g. peak nasal inspiratory flow, rhinomanometry and acoustic rhinometry to measure nasal airflow and obstruction), patient-reported nasal symptoms, psychological functioning, satisfaction with health care, quality of life in general, or rhinoplasty revision rates.[3] While multiple endpoints are likely important to rhinoplasty patients, nasal symptoms are generally considered amongst the most important factors to determine the functional and aesthetical success of rhinoplasty.[4-6] Symptoms are usually considered to be a 1-dimensional domain and are directly related to disease and treatment effects.[7]

An accepted method of measuring nasal symptoms is the use of patient reported outcome measures (PROMs).[8] Although many different PROMs have been developed, no consensus exists as to which PROM is optimal in the assessment of symptoms before and after rhinoplasty.[2,9] Selecting the best PROM to evaluate both functional and aesthetic changes following rhinoplasty requires the evaluation of specific instrument characteristics, such as interpretability of results (e.g. normative values, minimal important change), feasibility of use (e.g. availability, patient compliance) and measurement properties. Measurement properties are quality aspects of an outcome measure, such as reliability, validity or responsiveness (**Table 1**).[10] To make an evidence-based recommendation on the most suitable PROMs, the measurement properties of each available PROM must be evaluated in a high-quality systematic review of studies on measurement properties.

In 2017, Barone et al. performed a valuable systematic review on measurement properties of rhinoplasty outcome instruments.[11] However, this review included original development studies only, while other available studies on measurement properties of the included PROMs were excluded. Furthermore, the guidelines proposed by the Scientific Advisory Committee[12] lack comprehensive, detailed and consensus-based descriptions of what constitutes an adequate measurement property. A complementary

comprehensive systematic review that includes all available PROMs and corresponding measurement property studies, evaluated using consensus-based standards, could improve the selection of the most suitable rhinoplasty outcome instruments.

In this systematic review, the COSMIN (Consensus-based Standards for the selection of health Measurement Instruments) methodology and guidelines are used to critically appraise, summarize and compare the measurement properties of all PROMs that measure functional or aesthetic symptoms of patients undergoing rhinoplasty.[13] The primary aim is to make an evidence-based recommendation on the most suitable PROMs to measure these rhinoplasty outcomes or to identify PROMs that are potentially suitable after further research.

Methods

Search Strategy

A systematic literature search was performed by a senior biomedical information specialist (Wichor M. Bramer, BSc, Medical Library, Erasmus Medical Center, Rotterdam, the Netherlands) on November 1st 2017, with an update on May 18th, 2018. The search was conducted in EMBASE, Medline and Web-of-Science, without time limit. The search strategy consisted of three search elements: (1) construct: nasal function/obstruction or nasal shape/deformity (this term was kept as broad as possible to minimize missing relevant studies); (2) target population: those undergoing rhinoplasty or nasal surgery; and (3) a sensitive and validated search filter for instrument measurement properties (validity, reliability and/or responsiveness) in Medline and EMBASE.[14] Reference lists of articles were screened for potentially eligible studies. The full search strategy is available on request.

Inclusion and Exclusion Criteria

One of us (F.V.W.J. van Z.) screened all titles and abstracts from the search result, and then two of us (F.V.W.J. van Z. and F.R.D.) assessed the full texts for eligibility. Original articles written in the English language were included if the following criteria were met:

(1) Construct of interest: the aim of the PROM was to measure symptoms related to nasal breathing or satisfaction with nasal appearance as judged by the patient. Instruments measuring other end points (e.g. screening or prediction tools, general quality of life, tools to measure complication and revision rates or graft survival) were excluded.
(2) Study sample: the study sample consisted of patients who had a septoplasty or rhinoplasty. Rhinoplasty was defined as any surgical manipulation of the bony and cartilaginous nasal framework, targeted to improve nasal shape or nasal breathing. Septoplasty populations were included, since the main treatment goal is similar to the goal of a functional rhinoplasty. Excluded were articles describing the effect of turbinate surgery, cadaveric or animal studies, reconstructive (following trauma) or filler rhinoplasty studies and studies performed on a pediatric, cleft or other craniofacial population.
(3) Study aim: the aim of the study concerned the evaluation of one or more measurement properties (validity, reliability and/or responsiveness). Additionally, articles describing the development of a PROM were included, as well as articles aiming to evaluate interpretability (e.g. distribution of scores, change scores for relevant subgroups, minimal important change [MIC]). Articles simply using a PROM (e.g. to display outcomes of a specific clinic, population or surgical technique),

without the specific aim of evaluating measurement properties, were excluded. For example, instrument score change after rhinoplasty in longitudinal studies was not considered evidence of responsiveness.

Assessment of Methodological Quality of Included Studies
The methodological quality of each single study on measurement properties was assessed using the COSMIN Risk of Bias checklist.[15] Studies were stratified to having very good, adequate, doubtful or inadequate methodological quality. The COSMIN taxonomy was used to determine which measurement properties were assessed in each study (**Table 1**). The full data extraction process was performed independently by two of us (F.V.W.J. van Z. and J.A.H.). A third reviewer (L.B.M.) was consulted in case of uncertainties or disagreements. Relevant characteristics of included studies and instruments were extracted.

Assessment of Measurement Property Results
The result of each study on a measurement property was rated independently by two of us (F.V.W.J. van Z. and J.A.H.) against criteria for good measurement properties (+, - or ?).[16] Of note, "criterion validity" (**Table 1**) was not assessed because a gold standard for the measurement properties of PROMs is lacking.

The content validity of a questionnaire arises from assessment of relevance, comprehensiveness and comprehensibility of items (and instructions). Evidence of these parameters can be derived from PROM development studies or content validity studies. Additionally, the COSMIN manual for the assessment of content validity advises the reviewers to assess relevance, comprehensiveness and comprehensibility of questionnaire items themselves.[17] This requires a predefined definition of what aspects should or should not be included in a PROM measuring a specific construct to label it as relevant or comprehensive. Unfortunately, there is no consensus in the rhinoplasty literature on what specific symptoms or aspects are relevant and need to be included in a symptom outcome instrument. It is therefore impossible to predefine standards about relevance or comprehensiveness of items in a questionnaire. Hence, we confined the assessment of content validity to the quality of the development study of a specific PROM in which patient input was a major determinant.

Regarding hypotheses testing and responsiveness, we used the following predefined hypotheses about expected correlations and directions:[16]

1. Correlations with (changes in) instruments measuring similar constructs (i.e. other instruments identified in this review, or visual analogue scale [VAS] measuring similar construct) should be 0.50 or greater.

2. There should be significant differences in scores between relevant subgroups (e.g. patients with nasal obstruction should score significantly worse than controls without nasal obstruction).
3. It is expected that any instrument measuring the effect of nasal breathing-related symptoms after functional rhinoplasty should find a large effect size, defined as Cohen's d > 0.80. This hypothesis was based on findings of two reviews reporting large effect sizes of nasal valve surgery measured using the Nasal Obstruction Symptom Evaluation (NOSE) questionnaire and a VAS ("how well can you breathe through your nose", range 0-10).[18,19] Since the literature holds no data on the effect of septoplasty alone on this construct, we were unable to define hypotheses for septoplasty.
4. It is expected that any instrument measuring the effect of satisfaction with nasal appearance after aesthetic rhinoplasty should measure an effect size of Cohen's d > 0.80 as well. Yang et al. pooled change scores of the Rhinoplasty Outcome Evaluation (ROE) scale after rhinoplasty and reported large effect sizes.[20]

Evidence Synthesis and Generating Recommendations

All results per measurement property of each PROM were checked for consistency, and the results were qualitatively summarized by two of us (F.V.W.J. van Z. and J.A.H.). These summarized results were evaluated against the criteria for good measurement properties to get an overall rating (i.e. sufficient, insufficient, inconsistent, or indeterminate) for the measurement property.[16] The focus was on the PROM specifically, while in the previous steps the focus was on the single studies. Using the GRADE approach (Grading Recommendations Assessment, Development and Evaluation[21]), we graded the quality of evidence, determining the trustworthiness of the summarized result, based on 4 factors: (1) risk of bias as determined using the COSMIN Risk of Bias checklist, (2) inconsistency of results across studies, (3) imprecision (i.e. low sample size), and (4) indirectness (referring to evidence coming from different populations).[16] Subsequently, information on interpretability and feasibility was appraised. Important aspects of feasibility were defined as length of the instrument and completion time, ease of score calculation or access fee of an instrument. The aim of this methodology was to determine the most suitable PROM for each construct on the basis not only of the evaluation of measurement properties, but of the instrument's interpretability and feasibility aspects as well.

Table 1. COSMIN definitions of domains, measurement properties, and aspects of measurement properties.

Term		
Domain	**Measurement property**	**Aspect of a measurement property**
Reliability		
Reliability (extended definition)		
	Internal consistency	
	Reliability	
	Measurement error	
Validity		
	Content validity	
		Face validity
	Construct validity	
		Structural validity
		Hypotheses testing
		Cross-cultural validity
	Criterion validity	
Responsiveness		
	Responsiveness	
Interpretability		

Abbreviations: COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments[10]; HR-PRO, health-related patient-reported outcomes

	Definition
	The degree to which the measurement is free from measurement error
	The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same health related-patient reported outcomes (HR-PRO) (internal consistency); over time (test-retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e. raters or responders) on different occasions (intra-rater)
	The degree of the interrelatedness among the items
	The proportion of the total variance in the measurements which is due to 'true' differences between patients
	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
	The degree to which an HR-PRO instrument measures the construct(s) it purports to measure
	The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured
	The degree to which (the items of) an HR-PRO instrument indeed seems to be an adequate reflection of the construct to be measured
	The degree to which the scores of an HR-PRO instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured
	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured
	Idem construct validity
	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a 'gold standard'
	The ability of an HR-PRO instrument to detect change over time in the construct to be measured
	Idem responsiveness
	Interpretability is the degree to which one can assign qualitative meaning – that is, clinical or commonly understood connotations – to an instrument's quantitative scores or change in scores. Note that interpretability is not considered a measurement property, but an important characteristic of a measurement instrument.

Results

The search strategy identified 6057 unique publications, of which only 74 remained eligible for further review after title and abstract screening. Forty-four articles were excluded based on different study sample characteristics (N = 27), no assessment of measurement properties (N = 8), evaluation of an instrument that does not measure symptoms (N =7) and review articles (N = 2). Reference checking of included papers identified 3 measurement property studies that met inclusion criteria. A total of 33 articles, evaluating 12 different outcome instruments, were included in this review (**Figure 1**).

In 28 of the 33 included articles, development and/or other measurement properties were described. Most articles measured more than one measurement property; these 28 articles included a total of 98 separate studies. Five articles yielded additional information on interpretability. Of the 12 instruments identified, 5 measured symptoms related to nasal obstruction or other rhinologic complaints (hereinafter referred to as functional outcome); 2 focused on satisfaction or concern with nasal appearance (hereinafter referred to as aesthetic outcome); and 5 included items on both domains. The general characteristics of the included instruments and articles are summarized in **Table 2**[19,22-26] and **Table 3**,[25-52] respectively. Table 2 also includes interpretability and feasibility aspects of each PROM.

Supplement 1 summarizes the results of the included studies on measurement properties per instrument. For every included study the methodological quality and the result of that study is detailed. **Table 4** provides an overview of these findings and corresponding quality of evidence. Of note, the ROE, SNOT-23 (Sino-Nasal Outcome Test), FROI-17 (Functional Rhinoplasty Outcome Inventory), and RHINO (Rhinoplasty Health Inventory and Nasal Outcomes) included both functional and aesthetic items without separate sum scores for the separate domains; these PROMs have therefore been categorized as measuring both outcomes in one scale. The SCHNOS (Standardized Cosmesis and Health Nasal Outcomes Survey) also includes both domains, but subscale scores for each domain are reported, and therefore we present the results of these subscales individually.[44]

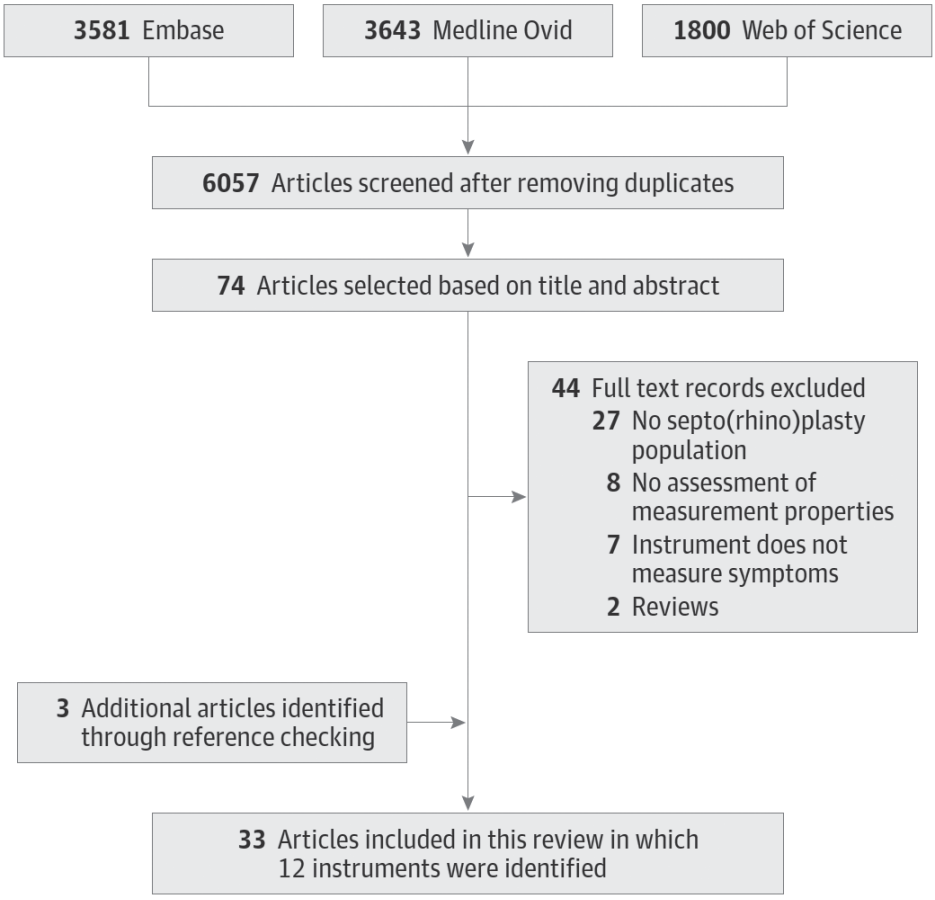


Figure 1. Flowchart of article inclusion.

Table 2. PROM characteristics.

Instrument	Construct(s) (according to reviewers)	Target population (according to author)	Conceptual model used (if stated)	(Sub)scale(s); number of items	Symptoms or domains assessed
ROE (2001)	Satisfaction with nasal appearance, nasal breathing	Facial plastic surgery patients	WHO; physical, mental and social well-being	Single (sub)scale, 6 items	Satisfaction with nasal appearance, social acceptance, confidence, desire for change
NOSE (2004)	Symptoms related to nasal obstruction	Patients with nasal obstruction	Items on nasal obstruction extracted from existing PROMs (mostly on rhinosinusitis)	Single (sub)scale; 5 items	Nasal congestion and obstruction; trouble with nasal breathing (regular and during exercise), sleeping
RhinoQoL (2005)	Symptoms related to rhinosinusitis and its physical and emotional impact	Patients with sinusitis	Existing PROMS reviewed, expert opinion, patient focus groups	Symptom frequency, 5 items; symptom bothersomeness, 3 items; symptom impact, 9 items	Facial pain, blocked nose, post-nasal drip, nasal discharge. Trouble sleeping and concentration; frustration/irritation/ depression/embarrassment due to nasal symptoms
SNOT-22 (2009)	Symptoms related to rhinosinusitis and its physical and emotional impact	Patients with sinusitis	Based on SNOT-20, 2 items added following expert opinion	Single (sub)scale, 22 items	Sneezing, nasal discharge, coughing, ear pain, dizziness, facial pain, trouble sleeping/sleepiness, reduced productivity/ concentration, frustration/ sadness/embarrassment, taste/smell, blockage/ congestion of the nose
FACE-Q rhinoplasty module (2010)	Satisfaction with nasal appearance	Rhinoplasty patients	Interview with patients, expert opinion, literature review	Satisfaction with nose, 10 items; satisfaction with nostrils, 5 items	Several dimensions of shape (e.g. size, width, straightness), views from different angles (in profile, photographs), how well the nose matches the face
UQ (2013)	Concern with nasal appearance	Aesthetic rhinoplasty patients	Designed based on ROE, no further description	Single (sub) scale, 5 items; additional VAS rating nasal appearance	Concern and bothersomeness with nasal appearance, professional/ social impact and stress due to this concern
SNOT-23 (2014)	Symptoms related to rhinosinusitis and its physical and emotional impact, including aesthetic item	Septorhinoplasty patients	Based on SNOT-22, 1 item added (no conceptual model)	Single (sub)scale, 23 items.	As SNOT-22, with additionally concern with nasal shape.

	Recall period	Response options	Range of scores	Interpretation	Administration time	Available languages (validated for rhinoplasty/ septoplasty)	Access fee
	'Current'	5-point Likert, 0–4	0-100 (dividing raw score by 24 and multiplying by 100), higher score indicates better outcome	Cutoff for normal scores: 12 and higher (distribution based)[22]	No reported data	English, German, Portuguese	No
	Last 4 weeks	5-point Likert, 0–4	0-100 (multiplying raw score by 5), higher score indicates worse outcome	MIC, 4.2-7.5 (distribution based)[23,24] Cutoff for normal scores, less than 30 (distribution and review based)[19,24] Severity groups[24]	No reported data	English, Spanish, Chinese, Portuguese, Italian, Greek, French, Dutch, Arabic, Slovenian, Turkish	No
	7 days	14 5-point Likert scales, 3 VAS, 0-10	Scores normalized to 0-100 for each subscale	MIC, 3.8-6.1 (distribution based)[23]	No reported data	English, French	No
	Past 2 weeks	6-point Likert, 0-5	0-110, higher score indicates worse outcome	Subgroup scores included[25]	No reported data	English, Persian	No
	Past week	4-point Likert, 1–4	Rasch logit scores transformed to range 0-100, higher score indicates better outcome, using conversion table	Cutoff for normal scores: 47 and higher, satisfaction with nose; 64 and higher, satisfaction with nostrils[26]	No reported data	English	No for non-profit users
	Not specified	5-point Likert, 1-5; 1 VAS	5-25, additional VAS, higher score indicates worse outcome	No norm- or subgroup scores or MIC available	No reported data	English, German	No
	Past 2 weeks	6-point Likert, 0-5	0-115, higher score indicates worse outcome	No norm- or subgroup scores or MIC available	No reported data	English	No

Table 2. Continued.

Instrument	Construct(s) (according to reviewers)	Target population (according to author)	Conceptual model used (if stated)	(Sub)scale(s); number of items	Symptoms or domains assessed
FROI-17 (2014)	Nasal symptoms including physical and emotional impact, including aesthetic item	Septorhinoplasty patients	Expert opinion	Single (sub)scale; 17 items.	Nasal obstruction, nasal discharge, dry throat, ear pressure, olfactory impairment, sleeping, sleepiness/decreased energy, concentration, irritability, depression, self-esteem, embarrassment with nasal shape, overall nasal adverse effects.
NSQ (2015)	Nasal symptoms, use of nasal medication	Septoplasty patients	Not stated.	Unknown, 13 items and 3 VAS.	Nasal breathing (including at night and during exercise), crusting, bleeding, sneezing, secretion, pain, use of nasal sprays/ antihistamines, smoking, allergy
NASION (2016)	Postoperative change of symptoms related to nasal obstruction	Nasal surgery patients	Expert opinion and focus group with patients	Single (sub)scale, 6 items	Nasal congestion and obstruction; trouble with nasal breathing (regular and during exercise), sleeping, quality of life as a result of nasal treatment
RHINO (2016)	Satisfaction with nasal breathing and nasal appearance	Rhinoplasty patients, both functional and aesthetic	WHO; physical, mental and social well-being	Single (sub)scale, 10 items	Nasal attractiveness/ proportionality/symmetry, trouble with nasal breathing (regular, during exercise, while sleeping), professional/social impact of nasal appearance, smell
SCHNOS (2017)	Symptoms related to nasal obstruction, concern with nasal shape	Rhinoplasty patients, both functional and aesthetic	Existing PROMs, patient interviews, expert opinion	Nasal obstruction domain, 4 items; nasal cosmesis domain, 6 items	Nasal obstruction and congestion; trouble with nasal breathing (during exercise and sleeping), mood/self-esteem related to nose, nasal shape (straightness, tip, symmetry), profile view, how well the nose matches the face

Abbreviations: FROI-17, Functional Rhinoplasty Outcome Inventory; MIC, minimal important change; NASION, Noninvasive Assessment and Symptomatic Improvement of the Obstructed Nose; NOSE, Nasal Obstruction Symptom Evaluation; NSQ, Nasal Surgical Questionnaire; PROM, patient-reported outcome measure;

	Recall period	Response options	Range of scores	Interpretation	Administration time	Available languages (validated for rhinoplasty/ septoplasty)	Access fee
	Not specified	6-point Likert, 0–5	0–100, dividing raw score by 85 and multiplying by 100, higher score indicates worse outcome	No norm- or subgroup scores or MIC available	5-10 minutes	English	No
	Not specified	1 5-point Likert, 3 VAS, 8 4-point Likert, 2 3-point Likert	No total range of scores, VAS, 0-10	No norm- or subgroup scores or MIC available	No reported data	English	No
	Past 4 weeks	5-point Likert, -2 to +2	-12 to 0 to +12 (maximum worsening)	No norm- or subgroup scores or MIC available	No reported data	English	No
	Not specified	5-point Likert, 1-5	20-100, multiplying raw score by 2, higher score indicates better outcome	No norm- or subgroup scores or MIC available	No reported data	English	No
	Past 4 weeks	6-point Likert, 0-5	0-20, nasal obstruction domain; 0-30, cosmesis domain	No norm- or subgroup scores or MIC available	No reported data	English	No

RHINO, Rhinoplasty Health Inventory and Nasal Outcomes; RhinoQoL, Rhinologic Quality of Life; ROE, Rhinoplasty Outcome Evaluation; SCHNOS, Standardized Cosmesis and Health Nasal Outcomes Survey; SNOT, SinoNasal Outcome Test; UQ, Utrecht Questionnaire; VAS, visual analogue scale; WHO, World Health Organization.

Table 3. Study characteristics.

Article	PROM	Population	Patient *n*, control *n*
Alsarraf,[27] 2000	ROE	Rhinoplasty (aesthetic facial plastic surgery)	Development; no patients included
Alsarraf et al., [28] 2001	ROE	Rhinoplasty (aesthetic facial plastic surgery)	26
Amer et al.,[29] 2017	NOSE	Septoplasty	101, 102 controls
Bezerra et al.,[30] 2011	NOSE	Septoplasty	33, controls unknown
Bulut et al.,[31] 2014	FROI-17	Septorhinoplasty (functional and aesthetic)	41 alpha version, 103 patients
Bulut et al.,[32] 2016	ROE	Primary septorhinoplasty (not specified)	100, 30 controls
Haye et al.,[33] 2015	NSQ	Septoplasty	55 (test retest), 75 (responsiveness)
Izu et al.,[34] 2014	ROE	Aesthetic rhinoplasty w/wo functional surgery	56, 100 controls
Jalessi et al.,[35] 2013	SNOT-22	Septoplasty candidates	30, 30 controls
Karahatay et al.,[36] 2018	NOSE	Septoplasty	168, 88 controls
Klassen et al.,[37] 2010	FACE-Q	Plastic surgery	50 qualitative interview, 35 cognitive debriefing
Klassen et al.,[38] 2016	FACE-Q	Rhinoplasty (plastic surgery clinics)	158
Radulesco et al.,[26] 2018	FACE-Q	Cosmetic rhinoplasty	52 patients, 52 controls.
Lachanas et al.,[39] 2014	NOSE	Septoplasty	132, 123 control
Larrosa et al.,[40] 2014	NOSE	Septoplasty	58, 58 controls
Lee and Most,[41] 2016	RHINO	Functional and/or aesthetic rhinoplasty	22 (8 functional, 4 aesthetic, 10 both)
Lohuis et al.,[42] 2013	UQ	Aesthetic rhinoplasty	121
Marro et al.,[43] 2011	NOSE, RhinoQol	Septoplasty	50, 50 controls

	Gender % male	Age mean (SD) and/ or range	Country	Language
	No data for rhinoplasty alone. Whole cohort 19.0.	47.7	USA	English
	59.4	31.92 +- 8.83 (16-53)	Egypt	Arabic
	57.6	39.3 +/- 11.9 (21-65)	Brazil	Portuguese
	50.4	28.7 +- 11.4	Germany	Not reported
	50.0	24 (range 18-65)	Germany	German
	Test-retest 67.3, responsiveness 64.0	Test retest 40.4 (12.9), responsiveness 41.6 (13.8)	Norway	Presumably Norwegian (The exclusion criteria were inadequate command of the Norwegian language)
	Cases 41.07, controls 44.0	Cases 29.65 (9.86 range 14-53), controls 30.79 (9.07 range 18-66).	Brazil	Portuguese
	Cases 40.0, controls 36.7.	Cases 30.4 (7.1), controls 33 (6.7).	Iran	Persian
	67.6, controls 51.1	28.6 +/- 8.4, controls 28.06 +/- 10.	Turkey	Turkish
	Qi 12.0, cd 14.0	Qi 51 (range 20-79), cd 45 (20-68).	USA and Canada	English
	25.3	32.6 (11.4) range 18-70.	USA, Canada, England.	English
	30.7 both groups	32.9 (11.2), controls 34.5 (13.1)	France	French
	66.1	35.4 (18-64)	Greece	Greek
	62.1	43.9 ± 15.1	Spain	Spanish
	45.5	34.9 (range 18-67)	USA	English
	15.7	34 (range 17-66)	Netherlands	Not reported
	68.0, controls 50.0	42.5, controls 40.1.	France	French

Table 3. Continued.

Article	PROM	Population	Patient *n*, control *n*
Moubayed et al.,[44] 2017	SCHNOS	Functional and/or aesthetic rhinoplasty	18 (development), 191 (measurement properties), 107 cosmetic or both, 52 functional, 32 other facial patients.
Mozzanica et al.,[45] 2013	NOSE	Septoplasty	116, 232 controls
Nouraei et al.,[46] 2016	NASION	Functional: conservative treatment, septoplasty, septorhinoplasty	71 total. 12 medical, 28 septal/turbinate, 31 septorhinoplasty
Onerci Celebi et al.,[47] 2018	NOSE	Septoplasty	50, 50 controls
Poirrier et al.,[25] 2013	SNOT-22	Functional and reconstructive septorhinoplasty	76
Spiekermann et al.,[48] 2017	UQ	Aesthetic and functional rhinoplasty	Cases not clearly reported, 36 controls
Stewart et al.,[49] 2004	NOSE	Septoplasty	32
Takhar et al.,[50] 2014	SNOT-23	Functional and reconstructive septorhinoplasty	69, 10 controls
Urbancic et al.,[51] 2017	NOSE	Septoplasty	58, 58 controls
van Zijl et al.,[52] 2017	NOSE	Septoplasty or septorhinoplasty	129, 50 controls

Abbreviations: FROI-17, Functional Rhinoplasty Outcome Inventory; NASION, Noninvasive Assessment and Symptomatic Improvement of the Obstructed Nose; NOSE, Nasal Obstruction Symptom Evaluation; NSQ, Nasal Surgical Questionnaire; PROM, patient-reported outcome measure;

	Gender % male	Age mean (SD) and/ or range	Country	Language
	33.0	41.5 (15.8)	USA	English
	64.6	43.8 (18-85)	Italy	Italian
	50.7	33.	England	English
	62, controls 56	30.8 +/- 11, controls 33.4 +/- 10.	Turkey	Turkish
	55.3	35.53 (11.41)	England	English
	Cases not clearly reported, controls 40.0	Cases male 25 (15-63), female 26 (16-70), controls 30 (21-58).	Germany	German
	78.1	47, 19-78	USA	English
	66.7, controls not specified.	34.5 (12.9), controls 30.3 (6.3).	England	English
	25.7	37.8 +- 13.92	Slovenia	Slovenain
	63.6	34.6 +- 14.5 (17-74)	Netherlands	Dutch

RHINO, Rhinoplasty Health Inventory and Nasal Outcomes; RhinoQoL, Rhinologic Quality of Life; ROE, Rhinoplasty Outcome Evaluation; SCHNOS, Standardized Cosmesis and Health Nasal Outcomes Survey; SNOT-22 and SNOT-23, SinoNasal Outcome Test; UQ, Utrecht Questionnaire.

Table 4. Summary of findings[a]

Instrument	**Structural validity**	**Internal consistency**	**Cross-cultural validity**	**Measurement invariance**
Functional outcome				
NOSE	**+** (high)	**+** (high)	NA	NA
RhinoQol	NA	NA	NA	NA
SNOT-22	NA	NA	NA	NA
NSQ	NA	NA	NA	NA
NASION	NA	NA	NA	NA
SCHNOS F[b]	**+** (moderate)	**+** (high)	NA	NA
Aesthetic outcome				
FACE-Q	**+** (low)	**+** (high)	**+** (very low)	NA
UQ	NA	NA	NA	NA
SCHNOS A[c]	**+** (moderate)	**+** (high)	NA	NA
Both outcomes in one scale				
RHINO	NA	NA	NA	NA
ROE	NA	NA	NA	NA
SNOT-23	NA	NA	NA	NA
FROI-17	NA	NA	NA	NA

Abbreviations: Plus sign (+), sufficient overall rating measurement property; question mark (?), indeterminate overall rating measurement property; FROI-17, Functional Rhinoplasty Outcome Inventory; NA, data not available; NASION, Noninvasive Assessment and Symptomatic Improvement of the Obstructed Nose; NOSE, Nasal Obstruction Symptom Evaluation; NSQ, Nasal Surgical Questionnaire; RHINO, Rhinoplasty Health Inventory and Nasal Outcomes; RhinoQoL, Rhinologic Quality of Life; ROE, Rhinoplasty Outcome Evaluation; SCHNOS, Standardized Cosmesis and Health Nasal Outcomes Survey; SNOT-22 and SNOT-23, SinoNasal Outcome Test; UQ, Utrecht Questionnaire.

	Reliability	Measurement error	Criterion validity	Construct validity	Responsiveness
	+ (high)	NA	NA	**+** (high)	**+** (high)
	NA	NA	NA	**?** (moderate)	NA
	NA	NA	NA	**+** (low)	**+** (low)
	NA	NA	NA	NA	**?** (low)
	NA	NA	NA	**+** (moderate)	NA
	NA	NA	NA	**+** (high)	**?** (moderate)
	NA	NA	NA	**+** (high)	**+** (low)
	NA	NA	NA	**+** (high)	**+** (high)
	NA	NA	NA	**+** (low)	**?** (moderate)
	NA	NA	NA	**+** (low)	**+** (very low)
	+ (low)	NA	NA	**+** (moderate)	**+** (high)
	NA	NA	NA	**+** (moderate)	**+** (low)
	NA	NA	NA	NA	**+** (high)

[a] For the parenthetically reported quality level of evidence, high indicates that we are very confident that the true measurement property lies close to that of the estimate of the measurement property; moderate, we are moderately confident in the measurement property estimate – the true measurement property is likely to be close to the estimate of the measurement property, but there is a possibility that it is substantially different; low, our confidence in the measurement property estimate is limited – the true measurement property may be substantially different from the estimate of the measurement property; and very low, we have little confidence in the measurement property estimate – the true measurement property is likely to be substantially different from the estimate of the measurement property.

[b] Functional subscale

[c] Aesthetic subscale

According to the COSMIN guidelines, the most important measurement property of a measurement instrument is content validity, but without recognized standards, this cannot be assessed. Structural validity and internal consistency are the next most important measurement properties, and subsequently the results of the other measurement properties are considered. As detailed in Table 4, the NOSE scale, the FACE-Q Rhinoplasty Set, and the SCHNOS are the only instruments with evidence for sufficient structural validity and internal consistency, and these 3 instruments will therefore be discussed in more detail. For the Rhinologic Quality of Life (RhinoQoL), SNOT-22, SNOT-23, Nasal Surgical Questionnaire (NSQ), Non-invasive Assessment and Symptomatic Improvement of the Obstructed Nose (NASION), Utrecht Questionnaire (UQ), ROE, FROI-17, and RHINO, no information on structural validity and internal consistency was available.

NOSE Scale

The NOSE was developed by Stewart et al. in 2004 and assesses symptoms related to nasal obstruction (the degree of congestion, obstruction, trouble breathing, trouble sleeping, and nasal obstruction during exercise).[49] The unidimensional questionnaire consists of 5 items. The quality of the development study was rated inadequate because no cognitive interview study including patients was performed. Other content validity studies were not encountered. High-quality evidence for sufficient structural validity, internal consistency, reliability, construct validity and responsiveness was found. The limits of agreement were evaluated by the Greek translation, but the measurement error cannot be formally determined because the authors do not provide an estimate of the MIC.[39]

FACE-Q Rhinoplasty Module

In 2013, Pusic et al. introduced the FACE-Q, a set of over 40 subscales measuring various patient-reported outcomes of facial aesthetic treatments, such as satisfaction with facial appearance, health-related quality of life, satisfaction with care, and adverse effects.[53] We included the 10-item subscale measuring satisfaction with nasal appearance and the 5-item subscale measuring satisfaction with nostril appearance in this review. The FACE-Q was developed by conducting cognitive interviews with patients.[37] The quality of this development study was rated inadequate because data were not coded independently. No other content validity studies were identified. The nasal appearance and nostril appearance scales both carried low-quality evidence for sufficient structural validity and high-quality evidence for sufficient internal consistency. There was very low evidence for sufficient cross-cultural validity. Furthermore, high-quality evidence for sufficient construct validity and low-quality evidence for sufficient responsiveness was found for both subscales.

SCHNOS

Moubayed et al. developed and tested the SCHNOS in 2018.[44] The list consists of a functional domain (4 items: e.g. blocked nose, nasal breathing during sleep) and an aesthetic domain (6 items, 5 evaluating several aspects of nasal shape and 1 measuring mood and self-esteem due to the nose). For each domain, subscores are calculated. The design of the PROM was relatively comprehensive and included a cognitive interview study with patients; however, the quality of the PROM design was rated doubtful owing to missing information on how the interviews were performed and analyzed, and missing documentation on the results of the pilot test. The study contains moderate evidence for sufficient structural validity and high-quality evidence for good internal consistency. Furthermore, the authors report sufficient construct validity supported by high-quality evidence for the functional domain and low-quality evidence for the aesthetic domain. Responsiveness of the SCHNOS cannot accurately be assessed because data to calculate effect sizes were not provided.

Recommendations for the Most Suitable PROM to Measure Functional Outcome and Aesthetic Outcome

Regarding PROMs measuring functional symptoms, the NOSE scale carries the best evidence for sufficient measurement properties (Table 4). It has been studied extensively, resulting in high-quality evidence for sufficient structural validity, internal consistency, reliability, construct validity, and responsiveness. Normative values, symptomatic ranges, and relevant scores for subgroups have been calculated. Furthermore, the instrument is short and available in at least 11 languages. Of note is that content validity is yet to be investigated because the development study was of inadequate quality.

The most suitable instrument to measure aesthetic outcome is less evident. The FACE-Q rhinoplasty module has sufficient structural validity, internal consistency, cross-cultural validity, construct validity and responsiveness, but with varying degrees of evidence quality. Apart from recently published cutoffs for normal scores, no other information on interpretability is yet available, and validated translation studies have not yet been published, but these may emerge in the future. The ROE questionnaire is a popular instrument to measure results of aesthetic rhinoplasty, but information on structural validity and internal consistency is lacking. Also the SCHNOS, including both a functional and aesthetic domain, carries potential for further research.

Discussion

This is the first systematic review in which measurement properties of available rhinoplasty outcome instruments measuring symptoms are analyzed using a standardized, consensus-based methodology. Recently updated methodological guidance for these type of reviews, developed by the COSMIN initiative, was adhered to ascertain sufficient quality of the systematic review.[15,16] In general, high-quality studies on measurement properties of PROMs measuring aesthetic and/or functional symptom-specific outcome of rhinoplasty are scarce. This review identified 12 instruments, but a majority of these have not been evaluated thoroughly.

The two constructs that are important to patient-reported outcome in rhinoplasty, referred to as functional and aesthetic outcome, have deliberately been separated in this review. Some PROMs identified, such as the ROE, SNOT-23, FROI-17 and RHINO, have integrated questions on both domains in one sum score. However, it is preferable to measure functional and aesthetic outcome separately, especially since most rhinoplasty practices focus on either construct rather than the two equally distributed. Our recommendation on outcome instruments is therefore disaggregated into an instrument or subscale measuring function and an instrument or subscale measuring shape.

It is important to note that the results of this systematic review do not imply that the other identified PROMs, not suggested for further study, are inadequate. It merely demonstrates that the measurement properties of these PROMs have not or have barely been tested. An important factor to consider on this issue is the influence of publication bias, as is suggested by the data detailed in Table 4 and Supplement 1. Of 98 identified separate studies on measurement properties, only one result was rated insufficient (Supplement 1).

Apart from further studies on the measurement properties of the identified PROMs, an even more important suggestion for future research is defining which constructs should be measured in patients undergoing rhinoplasty. We have focused on symptoms in this review, but there are numerous other constructs that are potentially important to measure in rhinoplasty patients, such as quality of life or satisfaction with care. Consensus on the framework of constructs allows the development of core outcome sets, which are agreed minimum sets of outcomes that should be measured and reported in clinical trials in a specific area.[54] An agreed standardized collection of outcomes considered important by patients and clinicians counteracts heterogeneity in outcome reporting and enables inclusion of more studies in future systematic reviews

on rhinoplasty interventions. Especially in rhinoplasty, a field in which the success of an intervention is difficult to quantify, standardized outcomes collection facilitates a large step toward transparency for both patients, clinicians and third parties.

Limitations

An important limitation of this systematic review is the absence of content validity evaluation. Without consensus on the definition of nasal function and nasal shape symptoms, a judgement of whether the content of a PROM is comprehensive and relevant cannot be made. Therefore, establishing a well-substantiated framework on which a questionnaire measuring nasal functional or aesthetic symptoms should be built is crucial. Methodologically sound qualitative studies including both patients and clinicians could contribute to defining these constructs. After consensus on the definitions and operationalization of the constructs, content validity of an existing PROM can be assessed in a content validity study by systematically asking patients and professionals about the relevance, comprehensiveness and comprehensibility of the PROM. Although we have rated the quality of NOSE and FACE-Q design studies as inadequate according to COSMIN standards, future content validity studies could still establish sufficient content validity of these PROMs. Of additional practical value is the content analysis that Barone et al have included in their review, providing an overview of aspects covered per PROM[11]. This could also aid in selecting the proper instrument after the construct has been defined. A comment on using the COSMIN standards could be that the quality criteria are rather strict; development studies, especially, must be of excellent quality to obtain a sufficient score.

Conclusion

Using consensus-based methodology, we evaluated the measurement properties of 12 different PROMs that measure outcomes after rhinoplasty. In general, high-quality studies on measurement properties of the identified PROMs are scarce. However, the available evidence of the NOSE has given this PROM a substantial head start, and the combination with available interpretation scores and favorable feasibility aspects allows us to state that the NOSE currently carries the most potential for further use. The FACE-Q and SCHNOS are, to a lesser extent, currently most suitable to measure aesthetic outcome. Further studies on the measurement properties of the identified PROMs, in particular content validity studies, are crucial to consolidate these recommendations.

References

1. Wahmann MS, Bulut OC, Bran GM, Veit JA, Riedel F. Systematic Review of Quality-of-Life Measurement After Aesthetic Rhinoplasty. *Aesthetic Plast Surg*. 2018;42(6):1635-1647.
2. Spataro E, Most SP. Measuring Nasal Obstruction Outcomes. *Otolaryngol Clin North Am*. 2018.
3. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*. 1995;273(1):59-65.
4. Cannon DE, Rhee JS. Evidence-based practice: functional rhinoplasty. *Otolaryngol Clin North Am*. 2012;45(5):1033-1043.
5. Rhee JS, Weaver EM, Park SS, et al. Clinical consensus statement: Diagnosis and management of nasal valve compromise. *Otolaryngol Head Neck Surg*. 2010;143(1):48-59.
6. Ching S, Thoma A, McCabe RE, Antony MM. Measuring outcomes in aesthetic surgery: A comprehensive review of the literature. *Plastic and reconstructive surgery*. 2003;111(1):469-480.
7. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: A new era in clinical research. *Perspect Clin Res*. 2011;2(4):137-144.
8. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*. 2006;4:79.
9. Lee MK, Most SP. Evidence-Based Medicine. Rhinoplasty. *Facial Plastic Surgery Clinics of North America*. 2015;23(3):303-312.
10. 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):737-745.
11. Barone M, Cogliandro A, Di Stefano N, Tambone V, Persichetti P. A systematic review of patient-reported outcome measures after rhinoplasty. *Eur Arch Otorhinolaryngol*. 2017;274(4):1807-1811.
12. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. 2002;11(3):193-205.
13. https://www.cosmin.nl. Accessed July 1, 2018.
14. Terwee CB, Jansma EP, Riphagen, II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18(8):1115-1123.
15. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018;27(5):1171-1179.
16. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157.
17. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159-1170.
18. Floyd EM, Ho S, Patel P, Rosenfeld RM, Gordin E. Systematic Review and Meta-analysis of Studies Evaluating Functional Rhinoplasty Outcomes with the NOSE Score Review. *Otolaryngol Head Neck Surg*. 2017;156(5):809-815.
19. Rhee JS, Sullivan CD, Frank DO, Kimbell JS, Garcia GJ. A systematic review of patient-reported nasal obstruction scores: defining normative and symptomatic ranges in surgical patients. *JAMA facial plastic surgery*. 2014;16(3):219-225; quiz 232.

20. Yang F, Liu Y, Xiao H, Li Y, Cun H, Zhao Y. Evaluation of Preoperative and Postoperative Patient Satisfaction and Quality of Life in Patients Undergoing Rhinoplasty: A Systematic Review and Meta-analysis. *Plast Reconstr Surg.* 2017;141(3):603-611.
21. GRADE. GRADE Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. 2013. http://gdt.guidelinedevelopment.org/app/handbook/handbook.html. *Accessed July 1, 2018.*
22. Izu SC, Kosugi EM, Brandão KV, et al. Normal values for the rhinoplasty outcome evaluation (ROE) questionnaire. *Brazilian Journal of Otorhinolaryngology.* 2012;78(4):76-79.
23. Mondina M, Marro M, Maurice S, Stoll D, De Gabory L. Assessment of nasal septoplasty using NOSE and RhinoQoL questionnaires. *European Archives of Oto-Rhino-Laryngology.* 2012;269(10):2189-2195.
24. Lipan MJ, Most SP. Development of a severity classification system for subjective nasal obstruction. *JAMA Facial Plast Surg.* 2013;15(5):358-361.
25. Poirrier AL, Ahluwalia S, Goodson A, Ellis M, Bentley M, Andrews P. Is the Sino-Nasal Outcome Test-22 a suitable evaluation for septorhinoplasty? *Laryngoscope.* 2013;123(1):76-81.
26. Radulesco T, Mancini J, Penicaud M, Dessi P, Michel J. Assessing normal values for the FACE-Q rhinoplasty module: An observational study. *Clin Otolaryngol.* 2018;43(4):1025-1030.
27. Alsarraf R. Outcomes research in facial plastic surgery: A review and new directions. *Aesthetic Plastic Surgery.* 2000;24(3):192-197.
28. Alsarraf R, Larrabee Jr WF, Anderson S, Murakami CS, Johnson Jr CM. Measuring cosmetic facial plastic surgery outcomes: a pilot study. *Arch Facial Plast Surg.* 2001;3(3):198-201.
29. Amer MA, Kabbash IA, Younes A, Elzayat S, Tomoum MO. Validation and cross-cultural adaptation of the arabic version of the nasal obstruction symptom evaluation scale. *Laryngoscope.* 2017;127(11):2455-2459.
30. Bezerra TF, Padua FG, Pilan RR, Stewart MG, Voegels RL. Cross-cultural adaptation and validation of a quality of life questionnaire: the Nasal Obstruction Symptom Evaluation questionnaire. *Rhinology.* 2011;49(2):227-231.
31. Bulut C, Wallner F, Plinkert PK, Baumann I. Development and validation of the Functional Rhinoplasty Outcome Inventory 17 (FROI-17). *Rhinology.* 2014;52(4):315-319.
32. Bulut OC, Plinkert PK, Wallner F, Baumann I. Quality of life in functional rhinoplasty: rhinoplasty outcomes evaluation German version (ROE-D). *Eur Arch Otorhinolaryngol.* 2016;273(9):2569-2573.
33. Haye R, Tarangen M, Shiryaeva O, Dosen LK. Evaluation of the Nasal Surgical Questionnaire for Monitoring Results of Septoplasty. *Int J Otolaryngol.* 2015;2015:563639.
34. Izu SC, Kosugi EM, Lopes AS, et al. Validation of the Rhinoplasty Outcomes Evaluation (ROE) questionnaire adapted to Brazilian Portuguese. *Qual Life Res.* 2014;23(3):953-958.
35. Jalessi M, Farhadi M, Kamrava SK, et al. The reliability and validity of the persian version of sinonasal outcome test 22 (SNOT 22) questionnaires. *Iranian Red Crescent Medical Journal.* 2013;15(5):404-408.
36. Karahatay S, Tasli H, Karakoc O, Aydin U, Turker T. Reliability and validity of the Turkish Nose Obstruction Symptom Evaluation (NOSE) scale. *Turk J Med Sci.* 2018;48(2):212-216.
37. Klassen AF, Cano SJ, Scott A, Snell L, Pusic AL. Measuring patient-reported outcomes in facial aesthetic patients: development of the FACE-Q. *Facial Plast Surg.* 2010;26(4):303-309.
38. Klassen AF, Cano SJ, East CA, et al. Development and Psychometric Evaluation of the FACE-Q Scales for Patients Undergoing Rhinoplasty. *JAMA Facial Plast Surg.* 2016;18(1):27-35.

39. Lachanas VA, Tsiouvaka S, Tsea M, Hajiioannou JK, Skoulakis CE. Validation of the Nasal Obstruction Symptom Evaluation (NOSE) scale for greek patients. *Otolaryngology - Head and Neck Surgery (United States).* 2014;151(5):819-823.
40. Larrosa F, Roura J, Dura MJ, Guirao M, Alberti A, Alobid I. Adaptation and validation of the Spanish version of the Nasal Obstruction Symptom Evaluation (NOSE) Scale. *Rhinology.* 2015;53(2):176-180.
41. Lee MK, Most SP. A Comprehensive Quality-of-Life Instrument for Aesthetic and Functional Rhinoplasty: The RHINO Scale. *Plast. reconstr. surg., Glob. open.* 2016;4(2):e611.
42. Lohuis PJ, Hakim S, Duivesteijn W, Knobbe A, Tasman AJ. Benefits of a short, practical questionnaire to measure subjective perception of nasal appearance after aesthetic rhinoplasty. *Plastic and reconstructive surgery.* 2013;132(6):913e-923e.
43. Marro M, Mondina M, Stoll D, de Gabory L. French validation of the NOSE and RhinoQOL questionnaires in the management of nasal obstruction. *Otolaryngol Head Neck Surg.* 2011;144(6):988-993.
44. Moubayed SP, Ioannidis JPA, Saltychev M, Most SP. The 10-Item Standardized Cosmesis and Health Nasal Outcomes Survey (SCHNOS) for Functional and Cosmetic Rhinoplasty. *JAMA Facial Plast Surg.* 2017.
45. Mozzanica F, Urbani E, Atac M, et al. Reliability and validity of the Italian nose obstruction symptom evaluation (I-NOSE) scale. *Eur Arch Otorhinolaryngol.* 2013 Nov;270(12):3087-94
46. Nouraei SAR, Virk JS, Kanona H, Zatonski M, Koury EF, Chatrath P. Non-invasive Assessment and Symptomatic Improvement of the Obstructed Nose (NASION): A physiology-based patient-centred approach to treatment selection and outcomes assessment in nasal obstruction. *Clin Otolaryngol.* 2016.
47. Onerci Celebi O, Araz Server E, Yigit O, Longur ES. Adaptation and validation of the Turkish version of the Nasal Obstruction Symptom Evaluation scale. *Int Forum Allergy Rhinol.* 2018;8(1):72-76.
48. Spiekermann C, Rudack C, Stenner M. Reliability and validity of the German version of the Utrecht Questionnaire for Outcome Assessment in Aesthetic Rhinoplasty (D-OAR). *Eur Arch Otorhinolaryngol.* 2017;274(11):3893-3898.
49. 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):157-163.
50. Takhar AS, Stephens J, Randhawa PS, Poirrier AL, Andrews P. Validation of the sino-nasal outcome test-23 in septorhinoplasty surgery. *Rhinology.* 2014;52(4):320-326.
51. Urbancic J, Soklic Kosak T, Jenko K, et al. Cross-cultural adaptation and validation of nasal obstruction symptom evaluation questionnaire in Slovenian language. *Zdr Varst.* 2017;56(1):18-23.
52. van Zijl F, Timman R, Datema FR. Adaptation and validation of the Dutch version of the nasal obstruction symptom evaluation (NOSE) scale. *Eur Arch Otorhinolaryngol.* 2017;274(6):2469-2476.
53. Pusic AL, Klassen AF, Scott AM, Cano SJ. Development and Psychometric Evaluation of the FACE-Q Satisfaction with Appearance Scale. A New Patient-Reported Outcome Instrument for Facial Aesthetics Patients. *Clin Plast Surg.* 2013;40(2):249-260.
54. Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials.* 2012;13:132.

Supplement 1. Measurement properties results.

PROM (ref)	Country (language)	Structural validity		
		n	Meth. quality	Result (rating)
ROE Alsarraf 2001	USA (English)			
ROE Bulut 2016	Germany (German)			
ROE Izu 2014	Brazil (Portuguese)			
ROE Summary				
NOSE Amer 2017	Egypt (Arab)			
NOSE Bezerra 2011	Brazil (Portuguese)			
NOSE Karahatay 2018	Turkey (Turkish			
NOSE Lachanas 2014	Greece (Greek)			
NOSE Larrosa 2014	Spain (Spanish)			
NOSE Marro 2011	France (French)			
NOSE Mozzanica 2013	Italy (Italian)			
NOSE Onerci Celebi 2018	Turkey (Turkish)			
NOSE Stewart 2004	USA (English)	32	inadequate	No factor analysis on 5 definitive items (?)
NOSE Urbancic 2017	Slovenia (Slovenian)			

Internal consistency			Cross-cultural validity			Reliability		
n	**Meth. quality**	**Result (rating)**	**n**	**Meth. quality**	**Result (rating)**	**n**	**Meth. quality**	**Result (rating)**
60	doubtful	0.84 (? no evidence for unidimensionality)				26	doubtful	Pearson correlation 0.83 (?)
Preop 100, 6mo postop 54, 12mo postop 69.	doubtful	Pre 0.64, 6mo post 0.82, 12mo post 0.73(? no evidence for unidimensionality)						
56	doubtful	0.86 (? no evidence for unidimensionality)				56	adequate	Test-retest ICC 0.94 (+)
		0.64-0.86 (?)						**0.94 (+)**
101	very good	0.995 (+)				101	inadequate	Pearson correlation 0.97 (?)
33	very good	0.81 (+)				29	inadequate	Gamma coefficient 0.78 (?)
170	very good	Test 0.928, retest 0.942 (+)				85	doubtful	Pearson correlation 0.95 (?)
109	very good	0.74, retest group 0.76 (+)				109	doubtful	Pearson r >0.70 (?)
58	very good	0.955 (+)				58	inadequate	Gamma coefficient 0.96 (?)
50	very good	0.86 (+)				50	inadequate	Spearman correlation all >0.40 (?)
116	very good	0.81 (+)				86	adequate	ICC 0.85 (+)
100	Very good	0.966 (+)				50	inadequate	Wilcoxon nonparametric test: no significant difference. (?)
32	inadequate	0.785 (+)				21	inadequate	Gamma coefficient 0.70 (?)
116	very good	0.971 (+)				90	inadequate	Gamma coefficient 0.98 (?)

Supplement 1. Continued.

PROM (ref)	Country (language)	Structural validity		
		n	Meth. quality	Result (rating)
NOSE van Zijl 2017	Netherlands (Dutch)	Preop 126, postop 50, control 50	very good	Unidimensional scale preop (CFI 0.968, TLI 0.935) and postop (CFI 0.999, TLI 0.998) (+), control poor indices CFI 0.876, TLI 0.752).
NOSE Summary				**1 factor preop and postop (+)**
RhinoQoL Marro 2011	France (French)			
RhinoQol Summary				
SNOT-22 Jalessi 2013	Iran (Persian)			
SNOT-22 Poirrier 2013	UK (English)			
SNOT-22 Summary				
FACE-Q nose Klassen 2016	USA/Canada/UK (English)	158	doubtful	Item residual correlations <0.30, nonsignificant x2 P values (+)
FACE-Q nose Radulesco 2018	France (French)			
FACE-Q nose Summary				**(+)**
FACE-Q nostrils Klassen 2016	USA/Canada/UK (English)	158	doubtful	Item residual correlations <0.30, nonsignificant x2 P values (+)
FACE-Q nostrils Radulesco 2018	France (French)			
FACE-Q nostrils Summary				**(+)**

	Internal consistency			Cross-cultural validity			Reliability		
	n	**Meth. quality**	**Result (rating)**	**n**	**Meth. quality**	**Result (rating)**	**n**	**Meth. quality**	**Result (rating)**
	Preop 129, postop 50	very good	Preop 0.81, postop 0.91 (+)				77	Adequate	ICC 0.89 (+)
			0.74-0.995 (+)						**0.85-0.89 (+)**
	50 patients, 50 controls (not mentioned in which group Cronbach alpha was calculated)	doubtful	Frequency score 0.57, bothersomeness 0.67, impact scores 0.83 (? no evidence for unidimensionality in septoplasty population)				50	inadequate	Wilcoxon nonparametric test: no significant difference. (?)
			(?)						**(?) (no ICCs)**
	30	doubtful	0.89 (? no evidence for unidimensionality in septoplasty population)				30	doubtful	Pearson correlation 0.85 (?)
			(?)						**(?) (no ICCs)**
	158	Very good	0.96 (+)	158	inade-quate	No differential item function detected (+)			
	52	Very good	0.93 (+)						
			0.93-0.96 (+)			**(+)**			
	158	Very good	0.96 (+)	158	inade-quate	No differential item function detected (+)			
	52	Very good	0.93 (+)						
			0.93-0.96 (+)			**(+)**			

Supplement 1. Continued.

PROM (ref)	Country (language)	Structural validity		
		n	Meth. quality	Result (rating)
UQ Lohuis 2013	Netherlands (not reported)			
UQ Spiekermann 2017	Germany (German)			
UQ Summary				
SNOT-23 Takhar 2014	UK (English)			
SNOT-23 Summary				
FROI-17 Bulut 2014	Germany (not reported)	41	inadequate	No factor analysis on 17 definitive items (?)
FROI-17 Summary				**?**
NSQ Haye 2015	Norway (Norwegian)			
NSQ Summary				
NASION Nouraei 2016	UK (English)			
NASION Summary				
RHINO Lee 2016	USA (English)			
RHINO Summary				
SCHNOS Functional subscale Moubayed 2017	USA (English)	191	adequate	2 unidimensional factors (functional and cosmesis) (+)
SCHNOS F Summary				**2 factors (+)**
SCHNOS Aesthetic subscale Moubayed 2017	USA (English)	191	adequate	2 unidimensional factors (functional and cosmesis) (+)
SCHNOS A Summary				**2 factors (+)**

	Internal consistency			Cross-cultural validity			Reliability		
	n	Meth. quality	Result (rating)	n	Meth. quality	Result (rating)	n	Meth. quality	Result (rating)
	121	doubtful	Preop 0.857, postop 0.837, re-postop 0.846 (? no evidence for unidimensionality)				74	doubtful	Pearson correlation 0.30 (?)
	Not clearly reported	doubtful	Preop 0.91, postop 0.91-0.92 (? no evidence for unidimensionality)						
			(?)						**(?) (no ICCs)**
							69	inadequate	Correlation 0.81, method not reported (?)
									(?) (no ICCs)
	103	inadequate	Nasal symptoms 0.78, general symptoms 0.92, self-confidence 0.74 (? Not at least low evidence for sufficient structural validity)						
			?						
	55	doubtful	0.82 (? no evidence for unidimensionality)				55	inadequate	Wilcoxon signed rank test: no significant difference (?)
			?						**?**
	71	doubtful	0.90 (? no evidence for unidimensionality)						
			(?)						
	22	doubtful	0.92, (? no factor analysis)				22	doubtful	Pearson correlation 0.94 (?)
			(?)						**(?) (no ICCs)**
	191	very good	0.94 (+)						
			0.94 (+)						
	191	very good	0.94 (+)						
			0.94 (+)						

Supplement 1 Continued.

PROM (ref)	Country (language)	Measurement error		
		n	Meth. quality	Result (rating)
ROE Alsarraf 2001	USA (English)			
ROE Bulut 2016	Germany (German)			
ROE Izu 2014	Brazil (Portuguese)			
ROE Summary				
NOSE Amer 2017	Egypt (Arab)			
NOSE Bezerra 2011	Brazil (Portuguese)			
NOSE Karahatay 2018	Turkey (Turkish)			
NOSE Lachanas 2014	Greece (Greek)	109	doubtful	Substantial agreement, Limits of Agreement -11.17 to 11.27 (range 0-100) (?, no MIC)
NOSE Larrosa 2014	Spain (Spanish)			
NOSE Marro 2011	France (French)			
NOSE Mozzanica 2013	Italy (Italian)			
NOSE Onerci Celebi 2018	Turkey (Turkish)			

	Construct validity			Responsiveness		
	n	**Meth. quality**	**Result (rating)**	**n**	**Meth. quality**	**Result (rating)**
				26	doubtful	38.8 to 83.3 = +44.5. No SD provided (?)
	100 patients, 30 controls	adequate	Control group sign. difference (+)	6mo 54, 12mo 69.	very good	Only SRM available; after 6mo 1.28, after 12mo 0.96 (+)
	56 patients, 100 controls	adequate	Control group sign. difference (+)	56	very good	After 3m Cohen's d 3.93 (+)
			2+ Overall (+)			**Cohen's d 0.96-3.93 (+)**
	101 patients, 102 controls.	very good	Control group sign. difference (+)	101? Not clearly reported	adequate	Cohen's d 5.87 (+)
	33 patients, controls not reported.	doubtful	Control group sign. difference (+)	33	very good	Effect size (method not specified) 2.72 (+)
	85 patients, 88 controls.	Adequate doubtful	Control group sign. difference (+) Correlation with VAS ('regarding severity of nasal obstruction') r= 0.948 (+)	83	Very good	Effect size Wilcoxon Z = 7.9 (+)
	132 patients, 123 controls	very good	Control group sign. difference (+)	50	very good	Cohen's d 3.22 (+)
	58 patients, 58 controls	very good doubtful	Control group sign. difference (+) Correlation patient group with VAS (difficulty in breathing through your nose") r > 0.50 for all items (+)	58? Not clearly reported	very good	Effect size (method not specified) 2.65 (+)
	50 patients, 50 controls.	very good very good	Control group sign. difference (+) Correlation patient group with RhinoQol r = 0.30-0.48 (-) Correlation control group with RhinoQol r = 0.49-0.66 (+)			
	116 patients, 232 controls	very good doubtful	Control group sign. difference (+) Correlation patient group with VAS (subjective sensation of nasal obstruction) all items r > 0.50 except item 4 (+)	40	adequate	Cohen's d 2.20 (+)
	50 patients, 50 controls.	Very good	Control group sign. difference (+)			

Supplement 1 Continued.

PROM (ref)	Country (language)	Measurement error		
		n	Meth. quality	Result (rating)
NOSE Stewart 2004	USA (English)			
NOSE Urbancic 2017	Slovenia (Slovenian)			
NOSE van Zijl 2017	Netherlands (Dutch)			
NOSE Summary				**MIC not defined (?)**
RhinoQoL Marro 2011	France (French)			
RhinoQol Summary				
SNOT-22 Jalessi 2013	Iran (Persian)			
SNOT-22 Poirrier 2013	UK (English)			
SNOT-22 Summary				
FACE-Q nose Klassen 2016	USA/Canada/UK (English)			
FACE-Q nose Radulesco 2018				
FACE-Q nose Summary				
FACE-Q nostrils Klassen 2016	USA/Canada/UK (English)			
FACE-Q nostrils Radulesco 2018				

	Construct validity			Responsiveness		
	n	**Meth. quality**	**Result (rating)**	**n**	**Meth. quality**	**Result (rating)**
	32 patients, 12 controls	doubtful doubtful	Control group sign. difference (+) Correlation patient group with VAS ("difficulty in breathing through your nose") r > 0.50 item 1-3, r < 0.50 item 4-5 (?)	21	doubt-ful	Effect size (method not specified) 2.65 (+)
	58 patients, 58 controls.	very good	Control group sign. difference (+)	21	very good	Cohen's d 4.58 (+)
	129 patients, 50 controls	Adequate doubtful	Control group sign. difference (+) Correlation patient group with VAS (difficulty in breathing through your nose") r < 0.50 preop and r > 0.50 postop (?) Correlation control group with VAS r > 0.50 (+)	50	ade-quate	Effect size cannot be calculated (?)
			13+, 2? and 1- Overall (+)			**Cohen's d 2.20-5.87 (+)**
	50 patients, 50 controls	very good very good	Control group sign. difference (+) Correlation patient group with NOSE r = 0.30-0.48 (-) Correlation control group with NOSE r = 0.49-0.66 (+)			
			2+ and 1- Overall (?)			
	30 patients, 30 controls	adequate	Control group sign. difference (+)			
	76	inadequate	Correlation patient group with VAS (not defined) r = 0.82 (+)	76	ade-quate	Cohen's d 7.94 (+)
			2+ Overall (+)			**(+)**
	158	very good	Correlation with facial appeareance scale r = 0.81 (+)	23	very good	Cohen's d 1.7 (+)
			1+ Overall (+)			**(+)**
	158	very good	Correlation with facial appearance scale r = 0.63 (+)	23	very good	Cohen's d 1.1 (+)

Supplement 1 Continued.

PROM (ref)	Country (language)	Measurement error		
		n	Meth. quality	Result (rating)
FACE-Q nostrils Summary				
UQ Lohuis 2013	Netherlands (not reported)			
UQ Spiekermann 2017	Germany (German)			
UQ Summary				
SNOT-23 Takhar 2014	UK (English)			
SNOT-23 Summary				
FROI-17 Bulut 2014	Germany (not reported)			
FROI-17 Summary				
NSQ Haye 2015	Norway (Norwegian)			
NSQ Summary				
NASION Nouraei 2016	UK (English)			
NASION Summary				
RHINO Lee 2016	USA (English)			
RHINO Summary				
SCHNOS Functional subscale Moubayed 2017	USA (English)			

	Construct validity			Responsiveness		
	n	**Meth. quality**	**Result (rating)**	**n**	**Meth. quality**	**Result (rating)**
			1+ Overall (+)			**(+)**
				121	doubtful	Effect size cannot be calculated (?)
	Presumably 214 patients (only female n and ratio given), 36 controls	very good doubtful	Control group sign. difference (+) Correlation patient group with VAS (subjective nasal appearance) preop r = 0.63, 12 mo postop r = 0.69 (+)	214 (? Only female *n* and ratio given, number used in responsiveness not clearly reported)	very good	Cohen's d 1mo 1.05, 3mo 1.36, 12mo 0.91 (+)
			2+ Overall (+)			**Cohen's d 1mo-12mo > 0.80 (+)**
	69 patients, 10 controls	very good very good inadequate	Correlation patient group with NOSE r = 0.82 (+) Control group sign. difference (+) "overall" VAS to SNOT-23, aesthetic VAS to only 23rd item.	69	adequate	Cohen's d 1.07 (+)
			1+ Overall (+)			**Cohen's d 1.07 (+)**
				103	very good	Only SRMs, Cohen's d cannot be calculated. SRMs are +
						(+)
				75	adequate	Cohen's d 1.27 day VAS, 1.60 night vas (+)
						Only for VAS, not whole PROM: (?)
	71	very good	Correlation with nose r = 0.64 (+), with SNOT-22 r = 0.77 (+)			
			2+ Overall (+)			
	22	very good (functional items)	Correlation of functional items with NOSE r = 0.85 (+)	22	adequate	Cohen's d 2.44 (+)
			1+ Overall (+)			**(+)**
	159 patients, 32 control	very good doubtful	Correlation of functional items with NOSE r = 0.94 (+) Control group sign. difference (visually based on box plots) (+)	69	adequate	Effect size cannot be calculated (?)

Supplement 1 Continued.

PROM (ref)	Country (language)	Measurement error		
		n	Meth. quality	Result (rating)
SCHNOS F Summary				
SCHNOS Aesthetic subscale Moubayed 2017	USA (English)			
SCHNOS A Summary				

+ sufficient rating measurement property
? indeterminate rating measurement property
- insufficient rating measurement property

MIC: minimal important change
ICC: intraclass correlation coefficient
CFI: comparative fix index

	Construct validity			Responsiveness		
	n	**Meth. quality**	**Result (rating)**	**n**	**Meth. quality**	**Result (rating)**
			2+ **Overall (+)**			**(?)**
	159 patients, 32 control	doubtful	Control group sign. difference (visually based on box plots) (+)	69	ade-quate	Effect size cannot be calculated (?)
			1+ **Overall (+)**			**(?)**

TLI: Tucker-Lewis index
SD: standard deviation
SRM: standardized response mean

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Chapter 3

Adaptation and validation of the Dutch version of the Nasal Obstruction Symptom Evaluation (NOSE) scale

Floris V.W.J. van Zijl MD, Reinier Timman PhD, Frank R. Datema MD PhD.

Eur Arch Otorhinolaryngol. 2017 Jun;274(6):2469-2476.

Abstract

Background: The Nasal Obstruction Symptom Evaluation (NOSE) scale is a validated questionnaire for the assessment of quality of life related to nasal obstruction. The aim of this study was to validate the Dutch (NL-NOSE) questionnaire.

Methodology: Guidelines for the cross-cultural adaptation process from the original English language scale into a Dutch language version were followed. Patients undergoing functional septoplasty or septorhinoplasty and asymptomatic controls completed the questionnaire both before and three months after surgery to test reliability and validity. Additionally, we explored the possibility to reduce the NOSE scale even further using graded response models.

Results: 129 patients and 50 controls were included. Internal consistency and test-retest reliability were good. The instrument showed excellent between-group and high response sensitivity to change. The NL-NOSE correlated well with the score on a visual analog scale measuring the subjective sensation of nasal obstruction, with exception of item 4 (trouble sleeping). Item 4 provided least information to the total scale and item 3 (trouble breathing through nose) most, particularly in the postoperative group.

Conclusion: The Dutch version of the NOSE (NL-NOSE) demonstrated satisfactory reliability and validity. We recommend the use of the NL-NOSE as a validated instrument to measure subjective severity of nasal obstruction in Dutch adult patients.

Introduction

In 2004, Stewart et al. introduced the Nasal Obstruction Symptom Evaluation (NOSE) scale as a valid, reliable and responsive self-report instrument to quantify the subjective burden related to nasal obstruction.[1] Patients are asked to answer five 5-point Likert Scale questions related to nasal obstruction resulting in a sumscore, ranging from 0 to 20, which is then multiplied by 5. The instrument is easy to complete with a minimal respondent burden, likely contributing to its global popularity in outcome research and surgical technique evaluation. This is illustrated by validated adaptations of the NOSE scale for the Spanish, Chinese, Italian, French, Greek and Portuguese language.[2-7] Additionally normative and abnormal value ranges for the NOSE scale have been outlined, allowing a more precise definition of treatment success and meaningful clinical changes of numerical scores.[8] The primary aim of this study was to translate and validate the NOSE scale instrument into the Dutch language.

An important remark when using (extensive) questionnaires to evaluate patient satisfaction, quality of life and change herein following medical treatment, is the influence of 'respondent burden bias' on given answers when questionnaires are too extensive. Although the NOSE scale is a relatively short questionnaire with only 5 items, the risk of inaccurate or incomplete answers might become important when the NOSE scale is offered to patients in addition to other questionnaires used for routine outcome monitoring. The secondary aim of this study was therefore to explore the possibility to reduce the NOSE scale into a more concise version including only the most indicative items.

Materials and Methods

This single-center instrument validation study consisted of a cross-cultural adaptation phase and a statistical validation phase. All data were prospectively collected between April 1st 2015 and September 1st 2016 at the department of otorhinolaryngology and head and neck surgery and the department of urology of the academic Erasmus Medical Center, Rotterdam (the Netherlands). This study was approved by the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam, the Netherlands, documented by study number MEC-2015-361.

Phase 1: cross-cultural adaptation to the Dutch language

General accepted guidelines for the process of cross-cultural adaptation were followed.[9] Forward translation of the original NOSE questionnaire was performed by one bilingual Dutch-native otolaryngologist and one bilingual Dutch-native professional translator without medical background. The two bilingual investigators reconciled differences between the two forward translations and checked for semantic and conceptual equivalence, resulting in one single provisional Dutch translation of the NOSE scale. Two English-native translators without medical background then translated the provisional Dutch questionnaire back into the original language. These backward translations were compared with the original NOSE scale focusing on discrepancies and item content. The end result was a final version of the questionnaire (NL-NOSE, **Figure 1**).

Phase 2: NL-NOSE validation

Study populations

For this study, two separate populations were recruited prospectively. The first group included patients with nasal obstruction caused by a septal deviation and/or nasal valve insufficiencies. Patients were included when they were eligible for surgery, able to speak and read the Dutch language and experienced nasal obstruction longer than three months, without a noticeable response to intranasal steroid treatment for a minimum of four weeks. We excluded patients younger than 18 years, patients with nasal obstruction related to mucosal disorders, craniofacial patients or patients who had prior septoplasty/septorhinoplasty or turbinate surgery. The second group consisted of healthy asymptomatic controls recruited at the department of urology. Controls needed to be older than 18 years, be able to read and speak the Dutch language and have no history of nasal obstruction and/or use of intranasal medication.

In hoeverre waren de volgende situaties de afgelopen maand een probleem voor u?
Over the past 1 month, how much of a problem were the following conditions for you?

Omcirkel alstublieft het best passende antwoord
Please circle the most correct response

		Geen probleem *Not a problem*	licht probleem *very mild problem*	matig probleem *moderate problem*	redelijk ernstig probleem *fairly bad problem*	ernstig probleem *severe problem*
1.	Verstopte neus of vol gevoel *Nasal congestion or stuffiness*	0	1	2	3	4
2.	Neusblokkade of obstructie *Nasal blockage or obstruction*	0	1	2	3	4
3.	Moeite met ademen door de neus *Trouble breathing through my nose*	0	1	2	3	4
4.	Problemen met slapen *Trouble sleeping*	0	1	2	3	4
5.	Niet voldoende lucht door de neus krijgen bij sport of inspanning *Unable to get enough air through my nose during exercise or exertion*	0	1	2	3	4

Figure 1. NL-NOSE adapted from the original NOSE scale (original scale in italic).

Methods and statistical analysis

Generally accepted quality criteria for validation were used as a guideline.[10,11] Generally, in the various language NOSE validation studies correlations of at least 0.40 with criterion measures were reported.[2-6] In order to detect a significant correlation coefficient of at least 0.40 we considered 50 cases as sufficient.[12] In cases where one out of five NL-NOSE items was missing, the total score was calculated from the mean of the completed items. If more than one item was missing, the case was excluded.

Internal consistency

Internal consistency was investigated using Cronbach's alpha coefficient, which was considered fair when alpha was between 0.70 and 0.79, good between 0.80 and 0.89, and excellent above 0.90.[13] Corrected item-total and inter-item correlations were tested using Spearman correlations. For assessment of unidimensionality a confirmatory factor analyses (CFA) was performed in the preoperative, postoperative and control groups. These CFAs tested single factor models without allowing additional covariances between the items. All CFAs were applied using ordinary maximum likelihood, that excludes cases with missing values. Standards for a good fit were derived from Brown.[14] The recommended index values are presented in **Table 1**.

Table 1. Fit measures confirmatory one factor analysis.

	Preoperative N = 126	Postoperative N = 50	Control N = 50	All cases N = 303	Recommended Brown 2006
RMSEA	0.103	0.024	0.208	0.115	<0.05/<0.08*
pclose	0.112	0.471	0.015	0.008	>0.05
CFI	0.968	0.999	0.876	0.985	>0.95
TLI	0.935	0.998	0.752	0.971	>0.95
SRMR	0.033	0.022	0.070	0.014	<0.08

RMSEA - root mean squared error of approximation
pclose - probability of RMSEA ≤ 0.05
CFI - comparative fit index
TLI - Tucker-Lewis index
SRMR - standardised root mean squared residual
*<0.05 = good, <0.08 reasonable

Reproducibility

Test-retest reliability was investigated by administering a second NL-NOSE questionnaire two weeks after the first. This was carried out for the patient group only. Patients with any change in conservative treatment after completing the first questionnaire (medication, nasal steroids, other) or change of symptoms due to upper or lower

airway infections were excluded for the assessment of test-retest reliability. Test-retest reliability was calculated using 2 way random – average measures intra-class correlation coefficients (ICC), with a positive rating for reliability given at >0.70. Differences between responders and non-responders at the second test were analyzed with Mann-Whitney U tests and a χ^2 test.

Discriminant validity
Discriminant validity of the NL-NOSE was tested by comparison of the scores of the patient group with the asymptomatic control group with a Mann-Whitney U test, with a significant difference defined as $p < 0.05$.

Responsiveness
The response (sensitivity to change) was tested using a subgroup of patients who were asked to complete the NL-NOSE three months after surgery, assessed with the Wilcoxon Rank test and calculation of the mean and inter-quartile range.

Construct validity
In the absence of an objective gold standard to quantify nasal patency, construct validity was assessed with a Spearman correlation test between NL-NOSE item scores and scores on a 100 mm Visual Analog Scale indicating nasal airway patency, ranging from 0 (very bad) to 10 (very good). Our predefined hypothesis reads "patients with higher NL-NOSE scores, indicating more subjective burden of nasal obstruction, will have higher scores on the nasal airway patency VAS".

Graded Response Models
Although this study was not primarily set up to develop a shorter version of the NL-NOSE scale, an exploratory attempt was made to reduce the number of items. For this purpose, graded response models (GRMs) were fitted to assess the information provided by each individual item on the latent trait. We only utilized the samples for which the unidimensionality assumption was reasonably met. The likelihood method applied in these GRMs was mean and variance adaptive Gauss–Hermite quadrature.

CFA was performed with STATA version14.1 (StataCorp, College Station, Texas 77845 USA), all other statistical analyses were performed with SPSS 21.0 (IBM SPSS, Armonk, NY, USA).

Results

Based on inclusion and exclusion criteria, a total of 131 patients with an indication for functional septoplasty or septorhinoplasty and 51 asymptomatic controls completed the NL-NOSE questionnaire. A number of 129 patients and 50 controls gave valid answers on at least 4 items. Of these 129 patients, 77 completed an additional re-test questionnaire returned by postal mail; 47 did not respond and 5 were excluded for retest analysis due to an unintended change in conservative treatment. No significant baseline differences were observed between responders and non-responders for the total NOSE scale (Mann-Whitney U=1950, p=0.80), age (U=1925, p=0.71) and gender (χ^2=0.043, p=0.84). On November 1st 2016, 64 out of 129 patients were operated on, of whom 50 patients had sufficient follow-up time to complete an additional postoperative questionnaire three months after surgery. A total of 313 administrations had been performed, with a total of 13 missing values on individual items (0.83%). These missing values led to the exclusion of 4 cases (1.28%).

The patient population (N=129) consisted of 82 males (63.6%), with a mean age of 34.6 ± 14.5 (range 17-74). Mean sum score (0-100) was 70.5 ± 20.0 (SD). No significant correlations of the NL-NOSE with age were observed, and there were no significant differences between men and women (non-parametric tests, all p-values >0.30).

Internal consistency

Internal consistency of the NL-NOSE was high with a Cronbach alpha of 0.81 for the preoperative group (N=129), and 0.91 in the postoperative group (N=50). Item-total and inter-item correlations for both preoperative and postoperative measures are displayed in **Table 2**. In the preoperative group, all values were above 0.40 except for the correlation between items 'trouble sleeping' and 'nasal blockage or obstruction' (0.36) and the correlation between items 'trouble sleeping' and 'unable to get enough air through my nose during exercise' (0.32). The inter-item correlations within the control group were much lower, in particular for item 5, while the inter-item correlations for all participants combined were much higher. Relationships between the different variables were close with highly significant differences (P<0.01) for all correlations.

The confirmatory factor analysis in the preoperative group showed good indices for the CFI, TLI and SRMR, but a lesser value for the RMSEA, though the chance that the RMSEA (pclose) is not significant is acceptable (**Table 1**, abbreviations enlisted), generally indicating that the unidimensionality assumption is reasonably met. In the postoperative group all fit indices are excellent. The fit measures in the control group are poor, indicating that unidimensionality of the scale in this group is not satisfactorily established.

Table 2. Inter-item and corrected item-total Spearman correlations of NL-NOSE.

Item	1. Congestion	2. Obstruction	3. Breathing	4. Sleeping	5. Exercise	Corrected total	
1. Congestion		0.72	0.81	0.51	0.80	0.84	Postoperative
2. Obstruction	0.61		0.77	0.46	0.69	0.77	
3. Breathing	0.52	0.57		0.55	0.81	0.89	
4. Sleeping	0.52	0.36	0.49		0.53	0.56	
5. Exercise	0.48	0.46	0.54	0.32		0.86	
Corrected total	0.69	0.60	0.67	0.53	0.54		
			Preoperative				
1. Congestion		0.81	0.78	0.70	0.73	0.84	All participants
2. Obstruction	0.43		0.80	0.65	0.73	0.82	
3. Breathing	0.46	0.45		0.73	0.81	0.87	
4. Sleeping	0.33	0.33	0.35		0.94	0.74	
5. Exercise	0.14	0.31	0.54	0.16		0.79	
Corrected total	0.44	0.74	0.77	0.75	.084		
			Controls				

Reproducibility

Test-retest reliability (N=77) was good with an intra-class correlation of 0.89 (p >0.001).

Control group and discriminant validity

In the control group (N=50), nineteen (38.0%) controls were male and the average age was 47.9 ± 16.8 (range 19-80). Mean sum score was 8.5 with a standard deviation of 13.0 (**Figure 2**). The NL-NOSE showed excellent discrimination between groups with a mean rank of 114.3 for patients and a mean rank of 27.2 for controls (Mann-Whitney U = 85, p<0.001). Cronbach's alpha in the control group was 0.79.

Pre- and postoperative evaluation (responsiveness)

Patients that completed a questionnaire before and after surgery (N=50) were all operated on by one author (F.R.D.), performing either a septoplasty or (septo) rhinoplasty mainly aiming at restoring nasal patency. Postoperative mean sum scores were significantly lower compared to preoperative values (Wilcoxon rank P<0.001). All but two patients had lower scores after the operation; these two patients reported no change. The magnitude of surgery effect was large; median sum scores dropped from 70.0 preoperatively to 20.0 postoperatively (median change 40.0, interquartile range 25-63).

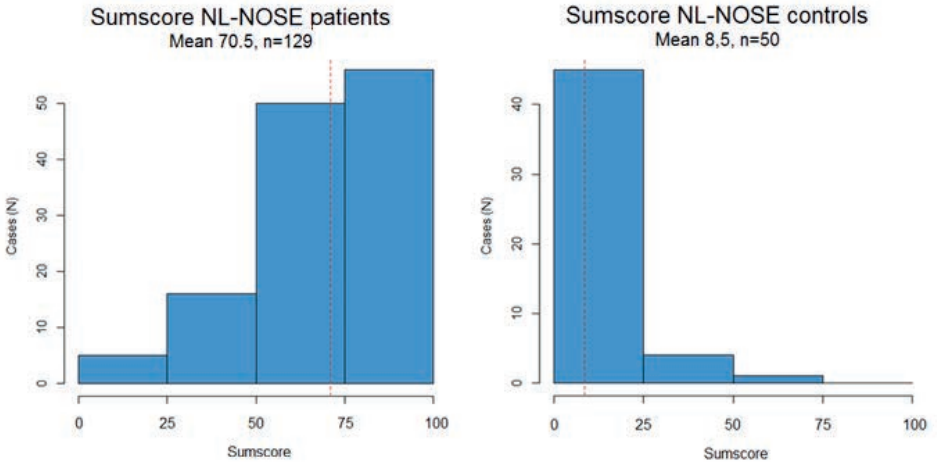


Figure 2. Sum scores of patients and controls.

Correlation with VAS (construct validity)

Correlation of the mean VAS score (left and right) with the NL-NOSE sum score and individual items is shown in **Table 3**. Sum scores correlated well with the VAS; both for the symptomatic cohort pre- and postoperatively and for the control cohort, confirming our hypothesis. Regarding the individual items, only the item 'trouble sleeping' did not correlate well with VAS.

Table 3. Spearman correlations of NL-NOSE with VAS.

	Preoperative N = 129		Postoperative N = 50		Control N = 50	
Item	rho	p	rho	p	rho	p
1. Congestion	0.46	<0.001	0.66	<0.001	0.57	<0.001
2. Obstruction	0.43	<0.001	0.79	<0.001	0.50	<0.001
3. Breathing	0.40	<0.001	0.74	<0.001	0.45	0.001
4. Sleeping	0.12	0.165	0.36	0.002	0.20	0.169
5. Exercise	0.46	<0.001	0.67	<0.001	0.36	0.011
Total	0.44	<0.001	0.78	<0.001	0.65	<0.001

rho = Spearman correlation

Graded Response Models

We fitted in GRMs for the pre- and postoperative patients, as the unidimensionality assumption was reasonably met in these groups. It must be noted that these models are explorative, as Reise & Yu reported that a GRM can be estimated with 250 cases but a sample of at least 500 is advised.[15] Our preoperative group included only 131 cases

for this analysis, and the postoperative group 51. In both samples item 4 (trouble with sleeping) provided least information to the total scale and item 3 most, particularly in the postoperative group (**Table 4, Figure 3**). These findings are confirmed with classical test theory analyses; the Mann-Whitney U values are largest for item 4 and smallest for item 3 (**Table 4**). Mann-Whitney Z-values are largest for item 3. These values for item 3 are about as large as for the total scale, suggesting that the total scale might not provide much more information than item 3.

Table 4. GRM item discrimination coefficients and differences in total NOSE scores between groups.

	GRM item coefficient (95% CI)		Difference pre- postoperative		Difference preoperative - controls	
	Preoperative N = 131	Postoperative N = 51	M-W U-value	M-W* Z-value	M-W U-value	M-W* Z-value
1. Congestion	2.51 (1.55, 3.47)	3.87 (1.92, 5.81)	1102.0	7.09	317.5	9.63
2. Obstruction	2.26 (1.46, 3.06)	3.46 (1.70, 5.22)	938.0	7.50	364.0	9.44
3. Breathing	2.60 (1.56, 3.63)	8.80 (-1.50, 19.11)	807.5	8.07	150.0	10.18
4. Sleeping	1.33 (0.83, 1.84)	1.69 (0.70, 2.68)	1095.0	7.20	597.0	8.78
5. Exercise	1.82 (1.15, 2.49)	4.57 (2.09, 7.04)	116.0	7.15	241.0	9.10
Total			670.0	8.23	85.0	10.12

M-W = Mann-Whitney U test

* all p-values <0.001

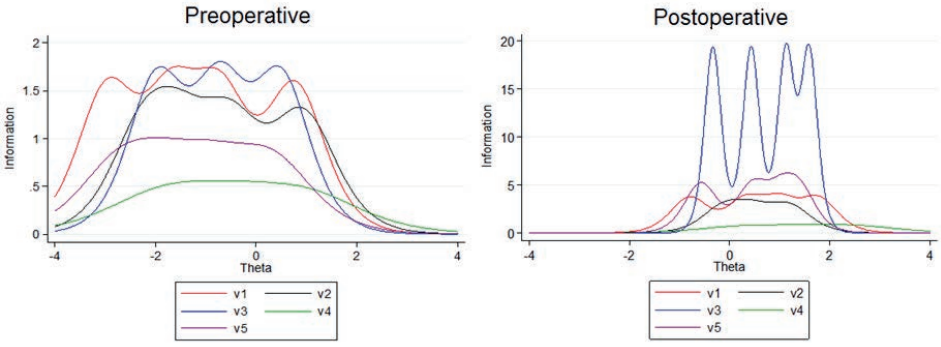


Figure 3. GRM item functions for pre- and postoperative patients.

Discussion

Routine outcome measurements have become an important indicator for medical performance. Transparent outcome reports assist the patient in making an educated guess between healthcare providers as long as the instruments used are comparable. The use of patient-reported outcome measures, in absence of globally accepted objective instruments, is feasible when the instruments used are validated. The NOSE scale is a validated, globally accepted instrument to quantify the burden related to nasal obstruction and change herein following nasal surgery. Cross-cultural adaptation of the NOSE scale makes it a valuable instrument to compare outcome results between institutions and organize multi-center studies. In that context, our need for a validated Dutch version of the NOSE scale became apparent.

Internal consistency measures the extent to which items in a questionnaire are correlated, which is an important measurement property for questionnaires that intend to measure a single underlying concept by using multiple items, like the NOSE questionnaire.[10] We found a Cronbach alpha of 0.81 for the NL-NOSE, which is within accepted ranges and comparable to previously reported NOSE validation studies.[1-5,7] When looking at the Cronbach alpha of the postoperative cohort, we found a value of 0.91. This is also demonstrated by **Table 1**, with item correlations in the postoperative group higher compared to the correlations of the preoperative group, and in **Table 2**, indicating that the fit for a unidimensional model is better for the post-operative group.

The reproducibility of the NL-NOSE was confirmed by performing a test-retest; correlating initial test and subsequent retest scores. We found an ICC coefficient of 0.89, indicating that of the questionnaire is stable over time. Normative data was generated by a cohort with no distinct complaints of nasal patency. This group scored a mean of 8.5 ± 13.0 compared to 70.5 ± 20.0 in the case cohort, suggesting that the NL-NOSE is a sensitive instrument to identify patients with nasal patency complaints. The correlations between the VAS and the total score of the NL-NOSE demonstrated good construct validity. We explored both pre-, postoperative and control group correlations and found that the correlations with VAS in these separate patient groups were lower compared to the correlations documented in the Spanish and Italian validations studies.[5,6] However, both the Spanish and Italian authors do not mention the composition of the tested group. When utilizing the overall cohort we encountered higher correlations with VAS, comparable to those reported in the Italian study. These increased correlations are caused by the larger variance induced by the combination of low scoring controls and postoperative patients and high scoring preoperative patients for the total scale and

their high respectively low VAS scores. Regarding the individual items, only the item 'trouble sleeping' did not correlate well with VAS, which is similar to the results of other validation studies. The GRM also pointed out that the contribution of item 4 to the total scale is limited.

Perhaps most importantly, in line with other validation studies, the NL-NOSE demonstrated excellent responsiveness after surgery, indicating that the instrument is suitable to measure outcome after treatment. Median sum scores dropped from 70 to 20 after surgery, which is comparable to the systematic review of Rhee et al. reviewing NOSE scores of patients with nasal airway obstruction after septo(rhino)plasty with or without turbinate surgery.[8] The authors compiled scores and found a mean pre-treatment score of 65 (standard deviation 22) and a mean post-treatment score of 23 (20). Furthermore, the authors found that that no individual study dropped less than 30 points, suggesting that a change of at least 30 may be considered a clinically meaningful measure of surgical success. Our results, with a median decrease of 40 points after surgery, therefore confirm that the NL-NOSE is able to measure clinically meaningful improvements after functional nasal surgery.

We fitted GRMs in order to explore whether a more concise version of the NL-NOSE could be constructed. These models suggest that item 3 might be nearly as informative as the overall NL-NOSE sum score. Future research pointed to this issue with larger study populations should be conducted to reach more definite conclusions regarding this matter.

A potential shortcoming of the study could be that the proportion of men is larger in the patient group compared to the control group. However, as we found no relation between the NL-NOSE and gender, we consider the influence of this difference to be minimal. Second, due to the lack of a Dutch questionnaire measuring nasal patency specific quality of life that has been validated in functional (septo)rhinoplasty patients, we had no perfect gold standard to compare results to. Instead, we chose to compare results to a nasal patency VAS score, for which our predefined hypothesis was met. Lastly this is a single-center study performed in an academic hospital, potentially causing impaired generalizability or selection bias. In the original validation study of Stewart however, the NOSE questionnaire showed good measurement properties in a multi-center study with four academic hospitals, and Larrosa et al. included both a tertiary and regional center with comparable results.[1,6]

Conclusion

This study was performed to adapt the NOSE questionnaire to the Dutch language. Satisfactory internal consistency, reliability, reproducibility, validity and responsiveness was demonstrated. We recommend the use of the NL-NOSE to quantify the subjective burden related to nasal obstruction and change herein following surgical intervention in Dutch adults.

Acknowledgements

Sarah Reuvers is thanked for her help with the inclusion of the control group.

References

1. 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):157-163.
2. Lachanas VA, Tsiouvaka S, Tsea M, Hajiioannou JK, Skoulakis CE. Validation of the nasal obstruction symptom evaluation (NOSE) scale for Greek patients. *Otolaryngol Head Neck Surg.* 2014;151(5):819-823.
3. Bezerra TF, Padua FG, Pilan RR, Stewart MG, Voegels RL. Cross-cultural adaptation and validation of a quality of life questionnaire: the Nasal Obstruction Symptom Evaluation questionnaire. *Rhinology.* 2011;49(2):227-231.
4. Marro M, Mondina M, Stoll D, de Gabory L. French validation of the NOSE and RhinoQOL questionnaires in the management of nasal obstruction. *Otolaryngol Head Neck Surg.* 2011;144(6):988-993.
5. Mozzanica F, Urbani E, Atac M, et al. Reliability and validity of the Italian nose obstruction symptom evaluation (I-NOSE) scale. *Eur Arch Otorhinolaryngol.* 2013;270(12):3087-3094.
6. Larrosa F, Roura J, Dura MJ, Guirao M, Alberti A, Alobid I. Adaptation and validation of the Spanish version of the Nasal Obstruction Symptom Evaluation (NOSE) Scale. *Rhinology.* 2015;53(2):176-180.
7. Dong D, Zhao Y, Stewart MG, et al. [Development of the Chinese nasal obstruction symptom evaluation (NOSE) questionnaire]. *Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi.* 2014;49(1):20-26.
8. Rhee JS, Sullivan CD, Frank DO, Kimbell JS, Garcia GJ. A systematic review of patient-reported nasal obstruction scores: defining normative and symptomatic ranges in surgical patients. *JAMA Facial Plast Surg.* 2014;16(3):219-225; quiz 232.
9. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976).* 2000;25(24):3186-3191.
10. 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.
11. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11(3):193-205.
12. van Belle G. *Statistical rules of thumb.* Chichester: Wiley; 2002.
13. Cicchetti D. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess.* 1994;6(4):284-290.
14. Brown T. Confirmatory factor analysis for applied research. 2006:87.
15. Reise SPYJ. Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement.* 1990;27:133-144.

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Chapter 4

Transparency in functional rhinoplasty: Benefits of routine prospective outcome measurements in a tertiary referral center

Frank R. Datema MD PhD, Floris V.W.J. van Zijl MD, Egge F. van der Poel PhD, Robert J. Baatenburg de Jong MD PhD, Peter J.F.M. Lohuis MD PhD.

Plast Reconstr Surg. 2017 Oct;140(4):691-702.

Abstract

Background: Patients, governments, health care providers, and insurance companies are increasingly interested in medical performance. Transparent outcome reporting requires a thorough methodologic design, dedicated prospective data collection process, and preferably no interference with the efficacy of daily practice. The primary aim of this article is to describe how these bottlenecks are tackled with an automated prospective rhinoplasty outcome routine. The secondary aim is to motivate others by describing practical benefits encountered during implementation.

Methods: Since April 2014, 269 consecutive patients referred for functional-aesthetic (revision) rhinoplasty were included. The Nasal Obstruction Symptom Evaluation scale, the Utrecht Questionnaire, and visual analogue scales were offered to all patients before primary consultation and follow-up to translate the subjective burden of nasal problems and change herein following surgery, into data. These data were exported for real-time automated outcome analysis supported by graphic output through a customized Web-based dashboard.

Results: One hundred seventy-one patients proved eligible for rhinoplasty, of which 121 had sufficient follow-up. The dashboard provides an overview of demographic characteristics of different populations, reasons why rhinoplasties were not performed, and real-time short- and long-term change in functional and aesthetic outcome in both primary and revision cases. Practical benefits of the instruments used are presented and discussed.

Conclusions: Routine prospective outcome monitoring provides an evidence-based response to the increasing demand for transparency in health care. The dashboard proved valuable during patient counseling, patient selection, and management of expectations and has the potential to compare rhinoplasty results between surgeons and institutions, provided that the populations share similar characteristics.

Introduction

In recent decades, governments, health care providers, insurance companies, and patients have shown an increased interest in medical performance. Motivated by the aim to improve health care quality, contain costs and provide data to facilitate an educated choice among health care providers, these parties demand transparency in medical outcome.

Rhinoplasty surgeons have always acknowledged the importance of outcome evaluations, which is reflected by numerous articles describing surgical results based on different functional and aesthetic outcome instruments. Despite all effort, a global consensus on how rhinoplasty outcome measurements should be performed remains absent. Systematic reviews on this subject conclude that an increased attention towards outcome research mainly leads to the development of new outcome measures.[1-4] This can be considered as an unwanted delay in a 'time of need', because valuable validated and practical outcome instruments are available.

Prospective outcome measuring ideally is a routine part of a daily rhinoplasty practice, but variability in surgical population characteristics, indications, and preferred maneuvers to tackle similar problems prevent rhinoplasty to be an exact science. Furthermore, outcome measurement requires a dedicated prospective data collection process, frequently anticipated to be an administrative burden interfering with the efficiency of daily practice.

This article describes relevant considerations that have led to the implementation of a short, practical rhinoplasty outcome routine, supported by automated data analysis and graphic output. Prospective results of all patients referred to our tertiary referral center in the past 3 years and additional practical benefits are presented and discussed to illustrate that with fairly limited effort it is possible to be fully transparent about overall (revision) rhinoplasty performance.

Methods

Instruments to quantify the burden of nasal form and function disturbances and change following surgery can be either objective or subjective, generic or disease specific, and patient reported or non-patient reported. The value of objective functional rhinoplasty outcome instruments remains a topic of debate and literature reports poor correlation with the subjective sensation of nasal airflow.[5-7] As the ultimate goal in rhinoplasty is to satisfy the patient's functional, aesthetic, and psychological expectations of the procedure, we considered the patient to be most qualified to quantify these endpoints. Therefore, subjective, disease specific, patient-reported outcome measures were selected as instruments for our rhinoplasty outcome routine. The patient-reported outcome measures of choice had to meet the requirements of being validated, internationally accepted, short, simple and practical. The first two characteristics are important to ensure that physicians speak the same 'meaningful language' when performing a cross-cultural comparison of outcome results. The latter three characteristics are important to minimize the influence of 'respondent burden bias' associated with incomplete and unreliable answers and to prevent an unwanted impact on the efficiency of daily practice.

Hospital setting and study population

Our rhinoplasty clinic is located at the Erasmus Medical Center, Rotterdam, The Netherlands, which is a university hospital that serves a population of approximately 3 million people in the southwestern area of the Netherlands. It acts as a tertiary referral center for approximately 30 affiliated hospitals. Every patient who was referred to the first author of this article (F.R.D.) between April of 2014 and December of 2016 to explore the indication for functional-aesthetic (revision) rhinoplasty was included in this prospective outcome routine. Candidates for septoplasty only, craniofacial patients (e.g. cleft lip, Crouzon syndrome, Muenke syndrome), and patients with a cosmetic indication only, were excluded.

Instruments

The first part of the outcome routine focuses on the quantification of subjective nasal obstruction using the Nasal Obstruction Symptom Evaluation (NOSE) scale and two visual analogue scales (VAS) to score left and right sided nasal obstruction on a scale ranging from 0 to 10 (with 0 indicating complete obstruction and 10 indicating a clear nose). The NOSE scale is a simple, short, validated and internationally accepted questionnaire containing five questions using a five-point Likert Scale related to nasal obstruction. Sum scores are multiplied by 5, resulting in a range from 0 to 100.[8]

The second part of the outcome routine focuses on the quantification of subjective body image and quality of life in relation to nasal appearance using the Utrecht Questionnaire (UQ) and one VAS to rate nasal appearance (with 0 indicating very ugly and 10 indicating very nice). The UQ is a short, validated questionnaire containing five questions using a five-point Likert Scale. Sum scores range between 5 and 25.[9-11] Of additional benefit is that questions 3 ("Does this concern affect you daily life, e.g. your work?") and 4 ("Does this concern affect your relationship with others?") are 'trick questions', where high scores hint towards a disturbance in body perception related to nasal appearance or body dysmorphic disorder. The UQ hereby complements 'gut feeling' and warning signs to identify these patients and to avoid surgery until a proper psychiatric evaluation has taken place.

Physical examination

Caudal septal luxation, loss of nasal tip support and external nasal valve insufficiency were identified in basal view (static and during inspiration). Anterior rhinoscopy was used to evaluate vestibular abnormality, aspect of the mucosa, septal deviations and perforations, and signs of internal nasal valve insufficiency. Internal nasal valve insufficiency was confirmed by a positive response to cotton ball placement in the valve apex, widening the minimal cross-sectional area. Evaluation on frontal, oblique and profile views was used to evaluate crookedness, loss of dorsal projection, loss of tip support, lateral wall collapse, and an overall evaluation of nasofacial proportions. Nasal endoscopy was performed in all patients to exclude abnormality in the posterior nose and nasopharynx. Acoustic rhinometry, rhinomanometry or radiographic studies were not used.

Timing and follow-up

Patients were included based on their referral letters. At first visit, they were asked to complete the questionnaires and VAS scores in the waiting room to minimize confounding factors related to consultation, and provided diagnostic information. Based on patient preference, questionnaires could be answered on paper or on digital tablet personal computers that run Limesurvey, an open-source survey tool (www.limesurvey.org). The physician viewed scores before consultation, providing valuable preliminary insights into the severity of nasal form and functional disturbance. This routine was repeated at regular postoperative check-ups, 3 (short term) and 12 months (long term) after surgery. Prior insights into the degree of subjective gain from surgery again proved valuable in preparation of follow-up consultation (see later under Discussion). Patients who did not show up for a planned postoperative evaluation were contacted and asked to schedule a new appointment or at least return the questionnaires by mail. The average time to complete the whole outcome routine was less than 5 minutes, not interfering with the efficacy of daily practice.

Statistical analysis

All data were collected in IBM SPSS Statistics for Windows, version 21 (IBM Corp., Armonk, N.Y.). Functional and aesthetic sum scores and change following rhinoplasty were analyzed with a *t* test for paired data. A graphical outcome dashboard was programmed with shiny: Web Application Framework for R (R package version 0.14.1[12]), allowing real-time outcome analysis when new patients were included in the data set. All data, tables, and figures presented in this paper are derived from the dashboard.

Results

Baseline population and surgical population characteristics

The first button on the left side of the dashboard gives access to the demographic characteristics of all patients that were referred with a potential indication for functional (revision) rhinoplasty between April of 2014 and December of 2016 (baseline population, *n* = 269) **(Figure 1)**. Based on our surgical indications and mandatory approval of health care insurance companies, the dashboard further shows that from the baseline population, 171 patients (63.6%) were eligible for surgery (surgical population). Ninety-three patients were male (54%) and 78 female (46%) with a mean age of 35 years (range, 15 to 74 years). Almost half of them had one or more unsatisfactory rhinoplasties performed elsewhere. Presently, 121 patients (70.4%) have had their (revision) rhinoplasty performed and 50 (29.6%) are on the waiting list. Main surgical indications were: internal nasal valve insufficiency (25.6%), crookedness (25.6%), septal reconstruction (12.8%), saddle nose deformity (9.9%), external nasal valve collapse (9.3%), combined internal and external nasal valve impairment (7.0%), drooping tip (4.7%) and tension nose deformity (2.9%) (data not provided by the dashboard).

Reasons why rhinoplasty was not performed or indicated

The second button of the dashboard shows different reasons why rhinoplasty was (or could) not be performed in 97 patients (36.1%) **(Figure 2)**. In The Netherlands, patients need to meet the review standards of the insurance carrier, which is crucial because in our hospital patients do not have the option of paying for the surgery themselves. From a total of 211 applications, 24 (11.5%) were declined. In case a patient had doubts regarding motivation for surgery after extensive explanation, if a patient was not fit for elective surgery or when a revision case was anticipated to be too complex, surgery was declined or postponed. Of further interest is that 12 patients (4.4%) had unrealistic expectations and/or showed signs of a disturbed body perception, and surgery was avoided. These patients report much higher scores on the UQ trick questions 3 and 4 (4.2 ± 1.3 and 3.8 ± 1.6, respectively) compared with the whole surgical population (2.2 ± 1.3 and 2.0 ± 1.3, respectively).

Functional outcome results

The third button on the dashboard provides access to information regarding the severity of nasal obstruction before the first consultation and 3 (short term) and 12 (long term) months after rhinoplasty **(Figure 3)**. A drop-down box located under the subheading 'selections' allows differentiation between the baseline and surgical population, primary or revision cases, and sex. The functional part of the dashboard is divided into four distinctive parts.

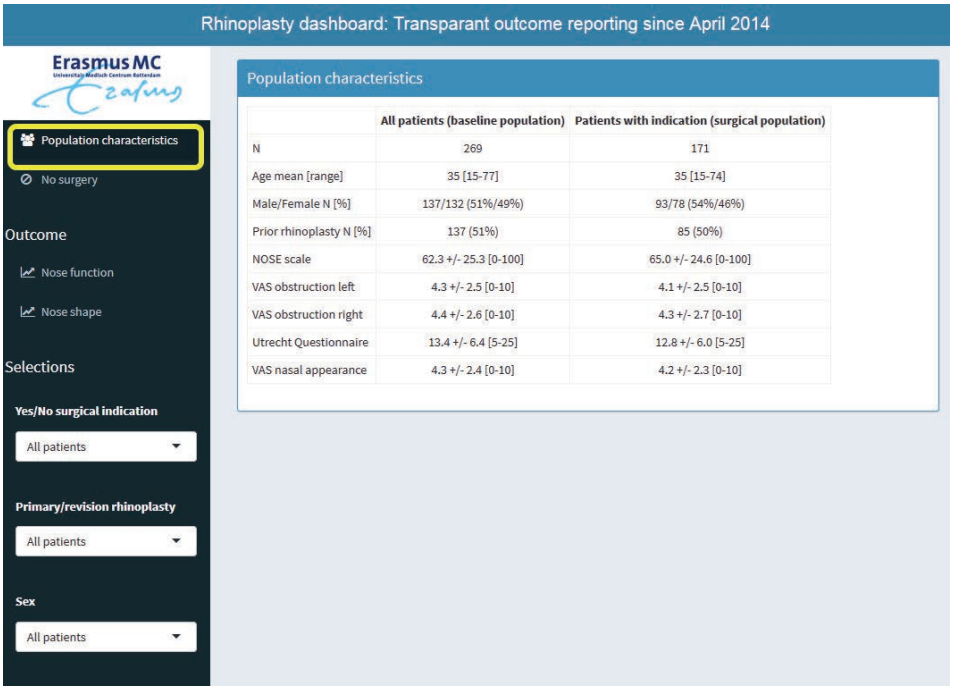


Figure 1. Dashboard snapshot showing button 1 (marked yellow) providing access to the demographic data of the baseline rhinoplasty population and the surgical population. *NOSE: Nasal Obstruction Symptom Evaluation*; *VAS: Visual Analogue Scale*

Table 1 provides mean NOSE scale scores of the individual five questions, a mean NOSE scale sum score, and two mean VAS scores for each side. The mean preoperative NOSE scale sum score was 65.0 ± 24.6. One year after surgery, the mean NOSE scale sum score decreased with 38.0 points to a mean overall sum score of 27.0 ± 23.6 ($p < 0.001$). The mean VAS scores for left and right sided nasal obstruction before surgery were 4.1 ± 2.5 and 4.3 ± 2.7, respectively; by 1 year after surgery, they improved by 2.8/2.4 points to mean VAS scores of 6.9/6.7 ± 1.9 ($p < 0.001$).

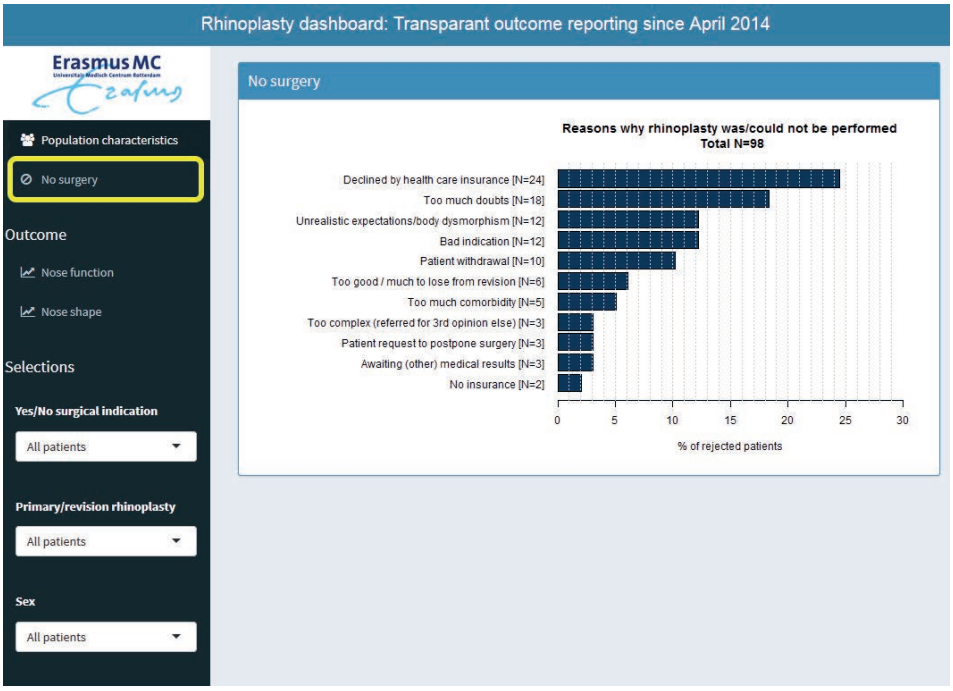


Figure 2. Dashboard snapshot showing button 2 (marked yellow) providing access to reasons why rhinoplasty could or was not performed.

A graphical distribution into five nasal obstruction severity groups, according to mean NOSE scale sum scores, is shown in **Figure 3**, *above*. In concordance with our surgical indications and review standards of insurance carriers, the majority of patients (69.6%) reported fairly bad to severe nasal obstruction. Of interest is that two patients reported no nasal obstruction. Review of their charts show that their surgical indication was correction of a severe nasal deformity following facial trauma. Presently, 97 out of 121 patients (80.2%) had sufficient follow-up data for short- and long-term effect analysis. **Figure 3,** *above,* shows that, 3 months after surgery, a majority of patients (56.7%) no longer report problems or very mild problems related to nasal obstruction. Unfortunately, 18 patients (14.9%) still experienced fairly bad to severe nasal obstruction.

Plotted VAS trend lines visualize how the improvement of nasal obstruction behaves over time (**Figure 3**, *below*). The VAS trend lines show that the mean overall short- and long-term functional benefit remained stable.

Table 1. Preoperative and postoperative mean sum scores based on complete questionnaires at 3- and 12 months follow-up.

	NOSE scale mean (SD)		
	Preoperative	*3 months postoperative*	*12 months postoperative*
All cases	65.0 (24.6)	27.8 (25.9)	27.0 (23.6)
Primary cases	63.9 (24.0)	25.6 (24.2)	21.9 (16.9)
Revision cases	66.7 (25.2)	29.9 (27.6)	30.7 (27.2)
	VAS obstruction left/right mean (SD)		
	Preoperative	*3 months postoperative*	*12 months postoperative*
All cases	4.1/4.3 (2.5)	7.0/7.0 (2.0)	6.9/6.7 (1.9)
Primary cases	4.2/4.2 (2.6)	7.1/7.2 (1.9)	7.2/7.0 (1.7)
Revision cases	4.0/4.4 (2.5)	7.0/6.8 (2.1)	6.8/6.6 (2.1)

* p<0.001 for all. *NOSE: Nasal Obstruction Symptom Evaluation; UQ: Utrecht Questionnaire; VAS: Visual Analogue Scale.*

A distribution curve of mean VAS score improvement 1 year after surgery is shown in **Figure 3**, *right*. As mentioned previously, some patients report absence of functional improvement or even increased nasal obstruction after their rhinoplasty. A critical appraisal of these patients is possible because the dashboard automatically identifies those on the far left of this distribution curve. Medical charts were reviewed to investigate possible explanations for the unwanted outcome and to see whether we can learn from potential mistakes (e.g. insufficient expectation management, influence of mucosal disorders, suboptimal surgical plan or performance; see later under Discussion).

Aesthetic outcome results

The fourth button of the dashboard provides access to information about the severity of concerns related to nasal appearance during intake and 3 and 12 months after rhinoplasty **(Figure 4)**. This part of the dashboard has an identical design and purpose as described for nasal obstruction. The mean preoperative UQ sum score was 12.8 ± 6.0 and decreased with 4.3 points to a mean overall sum score of 8.4 ± 5.1 (p <.001) (**Table 1**). The mean VAS score improved from 4.2 ± 2.3 to 6.6 ± 1.6 (p <0.001).

In contrast to nasal obstruction, the dashboard shows a more random distribution into the five aesthetic severity groups, with most patients reporting little to moderate concern (**Figure 4,** *above*). This is in concordance with a rhinoplasty population that is mainly operated on to correct nasal obstruction and/or a severe nasal deformity. The majority (59.1%) of patients that do report very much concern regarding nasal appearance were revision cases. One year after surgery, a majority of patients report none or little concern regarding nasal appearance (85.9%).

UQ mean (SD)		
Preoperative	*3 months postoperative*	*12 months postoperative*
12.8 (6.0)	7.4 (3.9)	8.4 (5.1)
12.2 (5.9)	7.1 (3.2)	7.0 (3.7)
13.8 (6.3)	7.7 (4.4)	9.4 (5.7)
VAS nasal appearance mean (SD)		
Preoperative	*3 months postoperative*	*12 months postoperative*
4.2 (2.3)	7.0 (1.5)	6.6 (1.6)
4.0 (2.3)	7.2 (1.6)	7.0 (1.4)
4.5 (2.3)	6.9 (1.4)	6.5 (1.7)

The VAS trend line shows that the mean overall short- and long-term aesthetic benefit decreases slightly over time (**Figure 4**, *below*). Because the decrease is minimal, we hypothesize that this is a habituation effect. Patients who reported an increased concern regarding nasal appearance after surgery are of specific interest. **Figure 4**, *right,* shows an aesthetic distribution curve of mean overall postoperative VAS score change, and automatically identifies the patients who are on the left side of the curve. Similar to patients with increased nasal obstruction, medical charts were reviewed to investigate possible explanations for an unwanted outcome.

Outcome comparison of primary and revision cases

As mentioned earlier under functional outcome results, the dashboard allows differentiation into outcome results of the whole cohort, primary cases only, or revision cases only. Statistical analysis in **Table 1** confirms what is generally accepted: revision cases show a slightly less favorable long-term outcome compared with primary cases. This information complements expectation management, especially in patients with severe dissatisfaction from their primary rhinoplasty.

Additional practical benefits

Although the primary aim of this study was to implement a practical prospective outcome routine as part of daily practice to provide transparency in overall rhinoplasty outcome, the information provided by the dashboard had unexpected additional benefits that upgraded our standard of care and provided evidence-based answers to questions asked by patients and third parties.

Figure 3. (*Above*) Dashboard snapshot showing button 3, providing access to severity distribution of nasal obstruction according to the mean overall NOSE scale sum scores before surgery (T1), 3 months after surgery (T2) and 12 months after surgery (T3). (*Below*) Dashboard snapshot showing a stable functional rhinoplasty result over time with median trend lines for nasal obstruction based on overall median VAS scores (blue: 25-75 percentiles; light blue 5-95 percentiles). (*Right*) Dashboard snapshot showing the distribution of VAS score improvement (mean 2.3 points ± 1.9).

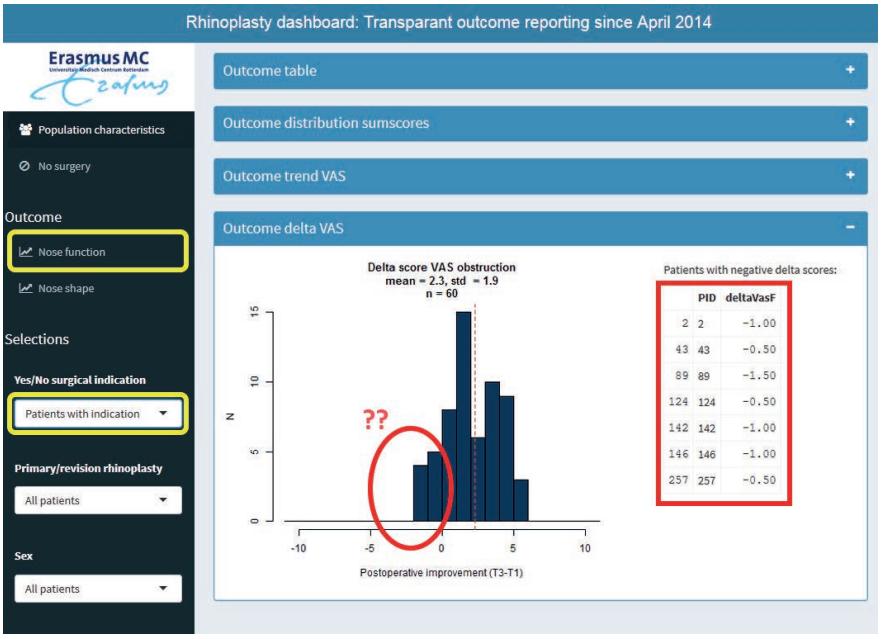


Figure 3. Continued.

1. Patient selection

Patient selection includes the availability of patient-reported outcome measures data before patient contact; in preparation of consultation, the physician has data on severity of complaints, unilateral or bilateral pathology, a forewarning of possible psychology related to nasal appearance (UQ trick questions) and a global disease severity profile that assists a targeted physical examination and differentiation between pure cosmetic candidates and functional-aesthetic candidates. **Figure 5** shows an example of a realistic rhinoplasty candidate, who reports severe nasal obstruction related to internal nasal valve insufficiency and crookedness. Despite his nasal deformities, he reports only a moderate concern regarding body image in relation to nasal appearance. An external approach was used for septal correction and extension, spreader grafts, and radix augmentation. Good patient satisfaction was reported 1 year after surgery.

Figure 4. (*Above*) Dashboard snapshot showing button 4, providing access to severity distribution of concern about nasal appearance according to the mean overall UQ sum scores before surgery (T1), 3 months after surgery (T2) and 12 months after surgery (T3). (*Below*) Dashboard snapshot showing a slight decrease of the aesthetic rhinoplasty result over time (blue: 25-75 percentiles; light blue 5-95 percentiles) based on overall median VAS score. (*Right*) Dashboard snapshot showing the distribution of aesthetic VAS score improvement (mean 2.4 points ± 2.6).

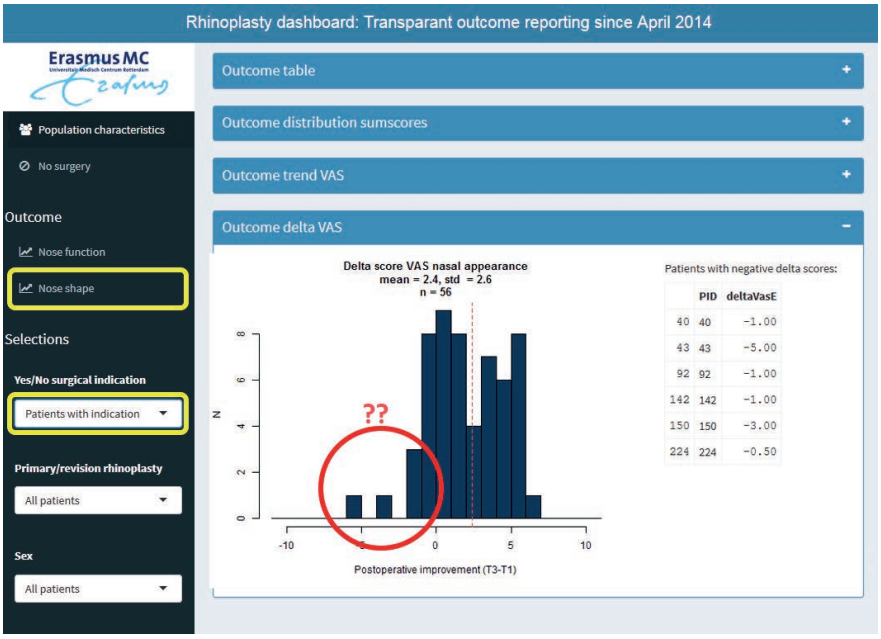


Figure 4. Continued.

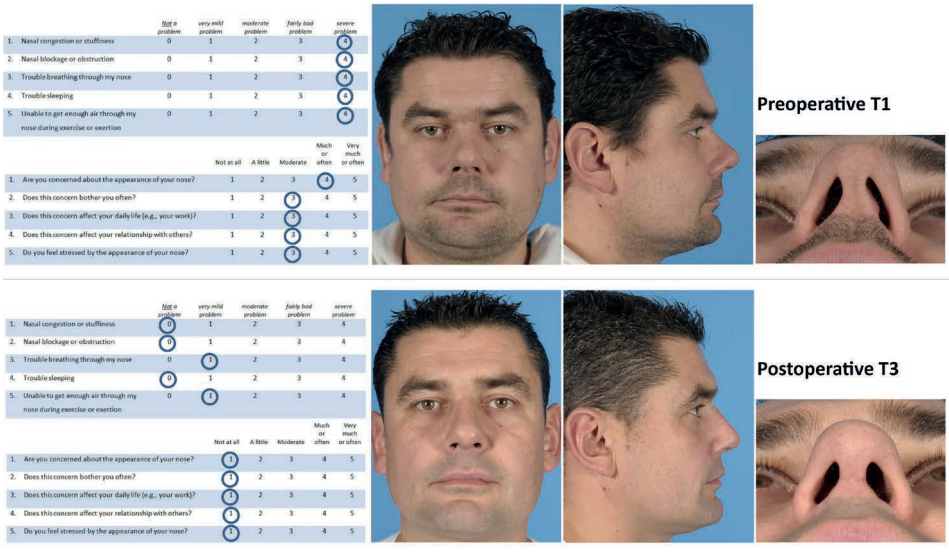


Figure 5. Crooked short nose case with preoperative and postoperative photography, NOSE and UQ scores.

2. Avoiding revisions

The availability of patient-reported outcome measures data before follow-up provides prior insights into patient satisfaction about performed surgery. **Figure 6** shows an example of a crooked nose patient. Twelve months after surgery, a noticeable subtle abnormality at the left middle third level of the nose is present and amendable for revision. However, scores on the UQ show that she is not in the least concerned about it and nasal obstruction has normalized. Revision surgery was never a topic during this postoperative visit. Unfortunately, the opposite can be true when postoperative photography hint towards a 'home run' feeling but the patient reports severe dissatisfaction.

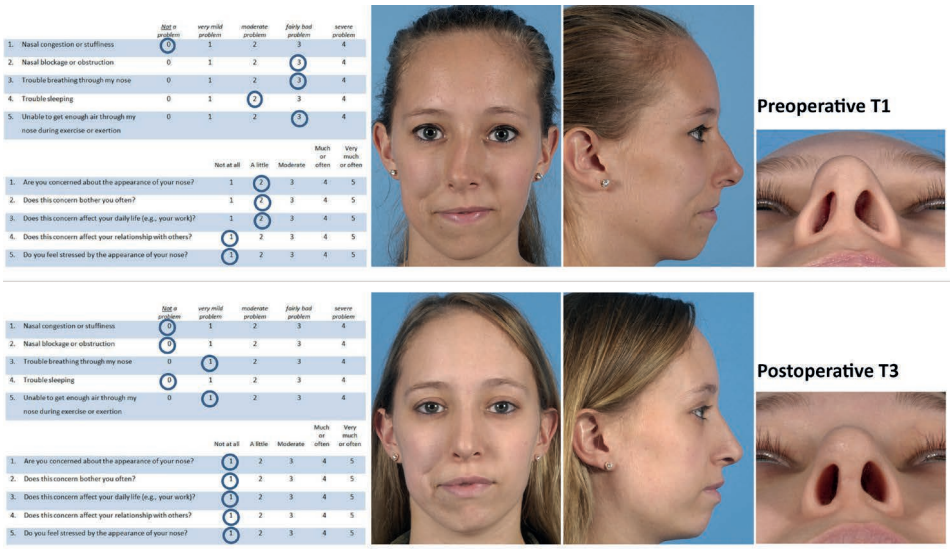


Figure 6. Preoperative and postoperative imaging and questionnaires of a crooked nose case with no concern about a small irregularity on the left middle third, 12 months after surgery. Revision was therefore avoided.

3. Patient empowerment

Patient empowerment in a candidate for rhinoplasty is a shared-decision process between doctor and patient, and is ideally based on proper informed consent leading to realistic expectations of the procedure. We noticed that the visual interpretations provided by the dashboard gave patients a better understanding of what to expect from the procedure. Furthermore, the dashboard provided immediate answers to questions such as: What average gain in nasal form and function can I expect? Do surgical results last over time? What is the chance that my health care insurance company will cover my rhinoplasty?

Results compared to literature

In 2014, Rhee et al. provided a systematic overview of studies using NOSE scale scores and functional VAS scores to measure treatment effect.[13] Normative and symptomatic ranges for both instruments in surgical patients who underwent turbinoplasty, septoplasty and/or septorhinoplasty were calculated. The functional VAS score methodologically differed from ours, with 0 indicating a clear nose and 10 indicating a complete obstructed nose, making comparison difficult. Pretreatment mean NOSE sum score in our population was 65.0, which is similar to the literature. The mean postoperative NOSE score in our patients was 28.0, which is slightly less than the mean of 23.0 reported in the literature. However, this mean includes excellent improvements reported after septoplasty only and might be higher when these patients are excluded.

Discussion

Measuring the functional and aesthetic effect of rhinoplasty is not new. However, a global consensus on which rhinoplasty outcome measures are most suited to compare results between doctors and institutions, and a protocol on how routine outcome measurements should be performed, remains absent.

In this article we describe a short and automated rhinoplasty outcome routine that easily fits a busy rhinoplasty practice and avoids the need for complex statistics. The routine translates nasal form and function disturbances into data and quantifies change following surgery. Provided that the evaluated study populations are clearly defined, the routine has the potential to facilitate a cross-cultural comparison of functional and/or aesthetic rhinoplasty results, and provide an evidence-based response toward the increasing demand for transparency in healthcare.

Despite our efforts to minimize bias and confounding factors in our outcome routine, critics might question the value of data derived from subjective patient-reported instruments. Factors that are not under control of the investigator could impact scores and might jeopardize outcome comparisons. In this context, one could ask, for example, whether preoperative functional scores of insured rhinoplasty patients might be exaggerated in order to meet insurance carrier standards. This subsequently would lead to a more dramatic postoperative improvement compared with patients who pay for the surgery themselves. We believe this effect to be minimal because the vast majority of our patients are unaware of insurance carrier criteria, and our preoperative NOSE scale scores, when compared with those reported by authors from other countries (such as the United States, where many patients pay for surgery themselves), were similar.[13] Another potential unwanted effect of insured rhinoplasty is overvaluing of the functional and aesthetic result, compared with a patient paying out of pocket who is likely to demand "high value for the money". Because our patients are not given the option of paying for surgery themselves and pure cosmetic cases were declined, we currently have no data to evaluate this potential effect. However, it is generally accepted that a pure aesthetic population is very different from an insured (revision) population, stipulating the importance of clearly defining and describing the rhinoplasty population that is being evaluated or compared. Methodologically, however, our outcome routine is suitable to evaluate and compare the overall aesthetic and/or functional benefit of rhinoplasty in both populations.[10,11]

As mentioned previously, the success of a prospective outcome routine depends on several factors. The routine needs to be quick, simple and based on validated instruments to warrant high patient compliance and generation of meaningful data.

Our outcome routine shows a compliance rate of 97.1%. Of equal importance is that the data collection process does not interfere with the efficacy of daily practice. This was avoided by completing the routine in the waiting room, which on average took less than 5 minutes. Finally, the investigating physician needs to be dedicated and motivated to include all patients in the routine and perform adequate follow-up. The shiny R dashboard proved a great motivator, because it avoids the need for complex statistics and provides real-time visual information on medical performance, which was the primary aim of this study. Unexpected additional benefits of the dashboard were found that ease daily practice and upgrade the quality of information provided during consultation.

Conclusion

We hope that the benefits of routine prospective outcome monitoring with automated data analysis appeal to other rhinoplasty surgeons and contribute to a more uniform approach. This would allow cross-cultural comparison of rhinoplasty results, creation of an evidence-based benchmark of surgical success and, when the surgical population increases with time, evaluation of specific rhinoplasty maneuvers to tackle similar problems. Of additional benefit is that the outcome routine facilitates preparation of consultation, patient selection and avoidance, and patient empowerment and that it provides evidence-based answers to questions asked by patients and third parties.

References

1. Ching S, Thoma A, McCabe RE, Antony MM. Measuring outcomes in aesthetic surgery: A comprehensive review of the literature. *Plastic and reconstructive surgery.* 2003;111(1):469-480.
2. Lee MK, Most SP. Evidence-Based Medicine. Rhinoplasty. *Facial Plastic Surgery Clinics of North America.* 2015;23(3):303-312.
3. Sharp HR, Rowe-Jones JM. Assessing outcome in aesthetic rhinoplasty. *Clinical Otolaryngology and Allied Sciences.* 2003;28(5):430-435.
4. Pawar SS, Garcia GJ, Kimbell JS, Rhee JS. Objective measures in aesthetic and functional nasal surgery: perspectives on nasal form and function. *Facial plastic surgery : FPS.* 2010;26(4):320-327.
5. Andre RF, Vuyk HD, Ahmed A, Graamans K, Nolst Trenite GJ. Correlation between subjective and objective evaluation of the nasal airway. A systematic review of the highest level of evidence. *Clin Otolaryngol.* 2009;34(6):518-525.
6. Kim DW, Rodriguez-Bruno K. Functional rhinoplasty. *Facial Plast Surg Clin North Am.* 2009;17(1):115-131, vii.
7. Lam DJ, James KT, Weaver EM. Comparison of anatomic, physiological, and subjective measures of the nasal airway. *Am J Rhinol.* 2006;20(5):463-470.
8. 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):157-163.
9. Lohuis PJ, Hakim S, Duivesteijn W, Knobbe A, Tasman AJ. Benefits of a short, practical questionnaire to measure subjective perception of nasal appearance after aesthetic rhinoplasty. *Plastic and reconstructive surgery.* 2013;132(6):913e-923e.
10. Datema FR, Lohuis P. Tongue-in-Groove Setback of the Medial Crura to Control Nasal Tip Deprojection in Open Rhinoplasty. *Aesthet. Plast. Surg.* 2015;39(1):53-62.
11. Lohuis PJFM, Datema FR. Patient satisfaction in Caucasian and Mediterranean open rhinoplasty using the tongue-in-groove technique: Prospective statistical analysis of change in subjective body image in relation to nasal appearance following aesthetic rhinoplasty. *Laryngoscope.* 2015;125(4):831-836.
12. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. shiny: Web Application Framework for R. R package version 0.14.1. *Available at:* https://CRAN.R-project.org/package=shiny. *Accessed December 22, 2016.*
13. Rhee JS, Sullivan CD, Frank DO, Kimbell JS, Garcia GJ. A systematic review of patient-reported nasal obstruction scores: defining normative and symptomatic ranges in surgical patients. *JAMA facial plastic surgery.* 2014;16(3):219-225; quiz 232.

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Chapter 5

Use of routine prospective functional and aesthetic outcome measurements in secondary cleft lip rhinoplasty

Floris V.W.J. van Zijl MD, Sarah Versnel MD PhD, Egge F. van der Poel PhD, Robert J. Baatenburg de Jong MD PhD, Frank R. Datema MD PhD.

JAMA Facial Plast Surg. 2018 Dec 1;20(6):488-494.

Abstract

Importance: Patients, governments, health care providers, and insurance companies show an increased interest in health outcomes, especially in centralized medical care, such as cleft lip nose treatment. Transparent outcome reporting requires a thorough methodological design, dedicated prospective data collection process, and, preferably, no interference with the efficacy of daily practice.

Objective: To describe the implementation of an automated and prospective secondary cleft lip rhinoplasty outcome routine.

Design, Setting, and Participants: A prospective analytic cohort pilot study was conducted among 123 consecutive patients referred for secondary cleft lip rhinoplasty from July 1, 2014, to March 31, 2018, at an academic teaching hospital.

Exposures: Secondary cleft lip rhinoplasty or revision.

Main Outcomes and Measures: Preoperative and 3- and 12-month postoperative scores on the Nasal Obstruction Symptom Evaluation scale (range 0-100, lower scores indicate better outcome), Utrecht Questionnaire (range 0-100, lower scores indicate better outcome), and visual analogue scales (range 0-10: 0, no obstruction; 10, completely blocked nose) were obtained. Data were exported for automated statistical outcome analysis that was supported by graphic output on a customized web-based dashboard.

Results: Of the 123 patients (68 male and 55 female; mean age, 23 years [range, 17-68 years]) included in the outcome routine, 103 patients (57 male and 46 female; mean age, 22 years [range, 17-50 years]) were eligible for surgery. The web-based dashboard provided demographic characteristics, reasons that surgery was not performed or indicated, and real-time, short- and long-term change in functional and aesthetic outcome after secondary cleft lip rhinoplasty. Among 66 patients with sufficient follow-up, mean (SD) Nasal Obstruction Symptom Evaluation sum scores after rhinoplasty improved from 30.8 (27.6), which is comparable to a moderate problem, to 19.2 (22.2), which is comparable to a very mild problem ($P<.001$), and mean Utrecht Questionnaire sum scores decreased from 13.1 (5.6) to 7.1 (3.3) ($P<.001$).

Conclusions and Relevance: Routine prospective outcome monitoring provides an evidence-based response to the increasing demand for transparency in health care. The web-based dashboard used during patient counseling, selection, and management of expectations has the potential to compare results of secondary cleft lip rhinoplasty between surgeons and institutions provided that the populations share similar characteristics. The administrative interference with a busy daily practice was limited.

Introduction

Surgical repair of the cleft lip nose is considered to be a difficult reconstructive procedure. Scarring from previous surgical procedures, presence of deformed and displaced cartilages, insufficient soft tissue for tension-free closure, and subsequent tendency of the nasal vestibule to contract over time are examples of factors associated with overall postoperative results. Long-term follow-up of the healing process as well as evaluation and quantification of patient satisfaction is essential to understand and communicate to our patients the potential benefits and limitations of secondary cleft lip rhinoplasty.

For years, rhinoplasty surgeons have acknowledged the importance of long-term outcome monitoring, which is reflected by numerous articles describing surgical results based on different outcome instruments. For cleft lip care specifically, participation in global collaborations, such as the International Consortium for Health Outcomes Measurement (ICHOM), emerge.[1] Despite these efforts, a global consensus on how rhinoplasty outcomes should be measured and compared remains absent. Centers that prospectively evaluate rhinoplasty outcomes as an integrated part of daily practice are scarce.

The latter is explained by the assumption that routine prospective outcome monitoring is a time-consuming administrative and statistical burden. Nevertheless, the possibility of evaluating health outcomes after elective surgical procedures based on systematic data collection is valuable.[2,3] We describe the design and implementation of a short and practical secondary cleft lip rhinoplasty outcome routine, focusing on nasal obstruction and aesthetics, that is supported by automated data collection and statistical analysis. Results are presented graphically on a customized web-based dashboard that is easy to understand by physicians and third parties who are interested in, and empowered by, health outcomes.

An unselected cohort of consecutive patients referred for secondary cleft lip and palate reconstruction, prospectively followed up until at least 1 year after surgery, was used to display the functionality and advantages of the outcome routine. Results are presented and discussed to illustrate that, with fairly limited effort, it is possible to be more transparent about surgical performance.

Methods

The ultimate goal of secondary cleft lip rhinoplasty is to satisfy the patient's functional and aesthetic expectations of the procedure. We believe the patient is the person most qualified to quantify these end points using rhinoplasty-specific, patient-reported outcome measures. The patient-reported outcome measures were chosen on validity, international acceptance, and short and simple design. The first 2 characteristics are important to ensure that physicians speak the same meaningful language when performing a cross-cultural comparison of outcome. The short and simple design minimizes the influence of respondent burden bias and avoids unwanted interference with the efficacy of daily practice.

Hospital Setting and Study Population

The Erasmus Medical Center in Rotterdam is a university teaching hospital that serves a population of approximately 3 million people in the southwestern area of the Netherlands. It acts as a referral center for 30 affiliated hospitals. Our clinic participates in a multidisciplinary cleft center, providing the whole spectrum of care from birth to final corrections in adult or adolescent patients. All patients referred to 2 of us (S.V. and F.R.D.) between July 1, 2014, and March 31, 2018, to explore the indication for secondary cleft rhinoplasty were included. Collection of data in this study was in accordance with the ethical standards of the institutional review board of the Erasmus Medical Center Rotterdam, and an oral informed consent for study participation was obtained from all patients.

Instrument

The first part of the outcome routine focuses on quantification of nasal obstruction by using the Nasal Obstruction Symptom Evaluation (NOSE) scale and 2 visual analogue scales (VAS) for left- and right-sided nasal obstruction. The VAS are scored from 0 to 10, where 0 indicates complete obstruction and 10 indicates a clear nose. The NOSE scale is a simple and validated questionnaire containing five 5-point Likert scale questions related to nasal obstruction. Sum scores are multiplied by 5, resulting in a range from 0 to 100, with a higher score indicating more patient dissatisfaction associated with nasal obstruction.[4,5]

The second part of the outcome routine focuses on quantification of body image and quality of life in relation to nasal appearance using the Utrecht Questionnaire (UQ) and 1 VAS to rate nasal appearance. This VAS is scored from 0 to 10, where 0 indicates very ugly and 10 indicates very nice. The UQ contains five 5-point Likert scale questions. Sum scores range between 5 and 25, and a higher sum score indicates more concern about

body image in relation to nasal appearance.[6,7] Of additional benefit is that question 3 and 4 of the UQ are trick questions, where high scores hint toward a disturbance in body perception related to nasal appearance or body dysmorphic disorder. The UQ complements clinicians' instinct and assists in identifying patients who may have a disturbance in body perception and for whom surgery should be avoided until proper psychological evaluation has been performed.

Physical Examination

In the basal view, the width of the alar base, nostril asymmetry, columella position, length and scarring, external nasal valve collapse, and dimensions of the hemi-tip were analyzed. Anterior rhinoscopy was used to evaluate the mucosa, volume of the inferior turbinates, septal length and deviation, vestibular floor defects, and signs of internal nasal valve insufficiency. In the frontal, oblique, and profile view, crookedness, tip asymmetries, lateral wall collapse, and overall nasofacial proportions were evaluated. Palpation was used to determine characteristics of the skin and soft-tissue envelope and composition of the osseocartilaginous dorsum.

Timing and Follow-up

At the first visit, patients completed the questionnaires in the waiting room to minimize confounding factors related to consultation and provided diagnostic information. Questionnaires could be answered on paper or a digital tablet PC that runs LimeSurvey, an open-source survey tool.[8] Prior to consultation, the physician viewed the scores providing preliminary insights into the severity of nasal aesthetic and functional disturbances. The routine was repeated at regular postoperative check-ups 3 months (short-term) and 12 months (long-term) after surgery. Patients who did not show up for postoperative evaluations were contacted to reschedule or to return completed questionnaires by mail. The percentage of patients lost to follow-up (repeated absence on check-ups 12 months after surgery) was 6.5% (5 of 77).

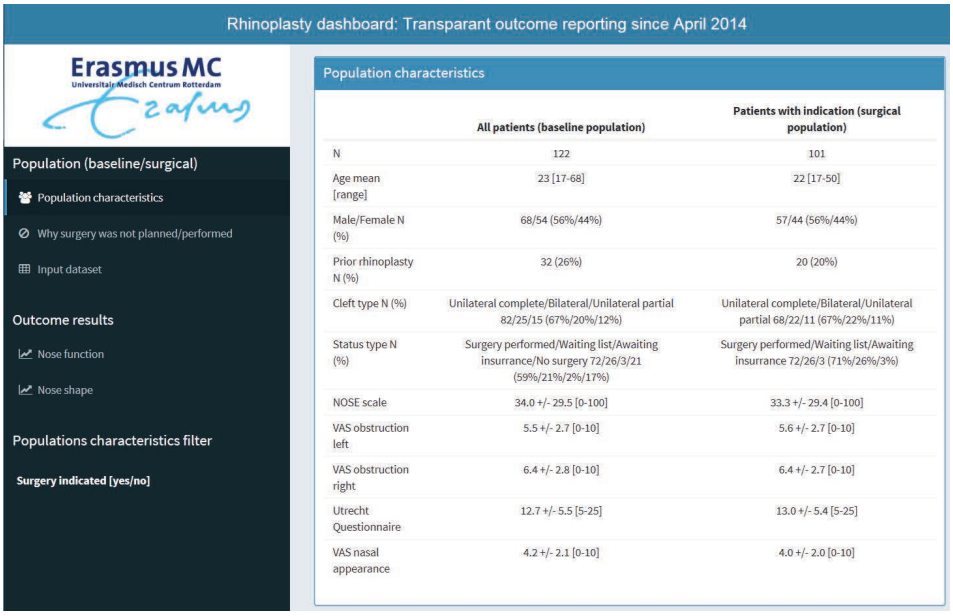
Statistical Analysis and Outcome Dashboard

All data were collected in IBM SPSS Statistics for Windows, version 21 (IBM Corp). Functional and aesthetic (sum) scores and change after surgery were analyzed with a *t* test for paired data. A web-based dashboard was programmed with shiny: Web Application Framework for R, allowing real-time outcome analysis when new patients were included in the data set.[9] All data, tables, and figures presented in this article are derived from the dashboard (available online at https://erasmusmc-kno.shinyapps.io/rhinoplasty-cleft).

Results

Characteristics of Baseline and Surgical Population

The first button on the left side of the dashboard provides access to demographic characteristics of all patients who were referred for secondary cleft lip rhinoplasty between July 1, 2014, and March 31, 2018 (baseline population; 123 patients). Sixty-eight patients were male (55.3%) and 55 were female (44.7%), with a mean age of 23 years (range, 17-68 years). Thirty-two patients (26.0%) had an unsatisfactory secondary rhinoplasty performed elsewhere. From the baseline population, 103 patients (83.7%) were eligible for surgery (surgical population; 57 male [55.3%] and 46 female [44.7%]; mean age, 22 years [range, 17-50 years]). Cleft type distribution of the surgical population was 69 patients (67.0%) with unilateral complete cleft, 12 (11.7%) with unilateral partial cleft, and 22 (21.4%) with bilateral cleft (**Figure 1**). At present, 77 of 103 patients (74.8%) have had surgery.



	All patients (baseline population)	Patients with indication (surgical population)
N	122	101
Age mean [range]	23 [17-68]	22 [17-50]
Male/Female N (%)	68/54 (56%/44%)	57/44 (56%/44%)
Prior rhinoplasty N (%)	32 (26%)	20 (20%)
Cleft type N (%)	Unilateral complete/Bilateral/Unilateral partial 82/25/15 (67%/20%/12%)	Unilateral complete/Bilateral/Unilateral partial 68/22/11 (67%/22%/11%)
Status type N (%)	Surgery performed/Waiting list/Awaiting insurrance/No surgery 72/26/3/21 (59%/21%/2%/17%)	Surgery performed/Waiting list/Awaiting insurrance 72/26/3 (71%/26%/3%)
NOSE scale	34.0 +/- 29.5 [0-100]	33.3 +/- 29.4 [0-100]
VAS obstruction left	5.5 +/- 2.7 [0-10]	5.6 +/- 2.7 [0-10]
VAS obstruction right	6.4 +/- 2.8 [0-10]	6.4 +/- 2.7 [0-10]
Utrecht Questionnaire	12.7 +/- 5.5 [5-25]	13.0 +/- 5.4 [5-25]
VAS nasal appearance	4.2 +/- 2.1 [0-10]	4.0 +/- 2.0 [0-10]

Figure 1. Dashboard snapshot showing button 1 (population characteristics), providing access to the demographic data of the baseline population and the surgical population. *NOSE: Nasal Obstruction Symptom Evaluation; VAS: Visual Analogue Scale.*

Reasons Secondary Cleft Lip Rhinoplasty Was Not Performed

Surgery was not performed in 20 of 123 patients (16.3%). In the Netherlands, all patients nominated for rhinoplasty must meet the review standards of the insurance carrier, which is crucial because we cannot offer out-of-pocket payment. One request was

permanently declined. Other reasons to discourage or postpone surgery were patients having too much to lose or too little to gain from revision rhinoplasty (n = 9), patients having significant doubts about surgery (n = 9), and patients not fit for surgery (n=1). Unrealistic expectations or signs of a disturbed body perception in relation to nasal appearance were not encountered.

Functional Outcome Results of the Surgical Population
The third button on the left side of the dashboard provides access to information about the severity of nasal obstruction before the first consultation and change after surgery. Automated outcome analysis is possible when either short-term or long-term follow-up is available. Presently, 77 of 103 eligible patients have undergone surgery, of whom 66 have at least short-term follow-up. Dropdown boxes under the subheading *population characteristics filter* allow differentiation between primary or revision cases. The dashboard is divided into 3 parts.

The first part of the dashboard is the outcome table, which provides mean scores of the individual NOSE scale questions and mean sum score and 2 mean VAS scores for left- and right-sided obstruction. The mean (SD) preoperative NOSE scale sum score of the surgical population was 30.8 (27.6), which is comparable to a moderate problem (**Table 1**). The mean postoperative NOSE scale sum score improved to 19.2 (22.2), which is comparable to a very mild problem. The mean (SD) VAS scores for left- and right-sided nasal obstruction before surgery were 5.6 (2.6) for the left side and 6.5 (2.8) for the right side; these scores significantly improved after surgery to a mean of 7.1 (2.2) for the left side and 7.4 (1.8) for the right side. A distribution into primary and revision cases shows what is generally accepted: revision cases reported a higher preoperative mean (SD) NOSE scale sum score (49.2 [26.6]) than did primary cases (26.2 [26.1]) and a less favorable postoperative end result (27.3 [24.2] vs 17.3 [21.5]). However, the overall mean sum score improvement in revision cases was higher than in primary cases.

The second part of the dashboard depicts the distribution of NOSE scale sum scores, which is a graphical distribution of all patients into 5 nasal obstruction severity groups according to mean NOSE scale sum scores (**Figure 2**).[10] Prior to surgery, a random distribution is seen, with most patients reporting very mild to moderate obstruction. Despite the usual presence of framework pathology, 20 patients reported no nasal obstruction before surgery. Severe problems were reported by 10 patients. **Figure 2** further shows that 3 and 12 months after surgery, most patients reported no problems or very mild problems. One year after surgery, 5 patients still reported fairly severe obstruction.

Table 1. Preoperative and postoperative mean sum scores based on complete questionnaires (*n* = 66)

NOSE scale	**Preop (T1) mean ± SD**	**Postop (T3 or T2) mean ± SD**	**p-value**
Individual questions			
1. Nasal congestion or stuffiness	5.4 ± 6.5	5.2 ± 5.8	0.87
2. Nasal blockage or obstruction	5.8 ± 6.7	3.3 ± 5.3	0.01
3. Trouble breathing through nose	8.5 ± 6.8	4.6 ± 5.7	< 0.01
4. Trouble sleeping	2.4 ± 4.7	1.8 ± 4.1	0.31
5. Insufficient nasal airflow during exercise/exertion	8.6 ± 7.7	4.3 ± 6.2	< 0.01
Sum scores			
Whole population	30.8 ± 27.6	19.2 ± 22.2	< 0.01
Primary cases (N = 53)	26.2 ± 26.1	17.3 ± 21.5	0.02
Revision cases (N = 13)	49.2 ± 26.6	27.3 ± 24.2	0.01
VAS nasal airway patency			
Whole population (left/right)	5.6 ± 2.6/6.5 ± 2.8	7.1 ± 2.2 / 7.4 ± 1.8	< 0.01/0.01
Primary cases (N = 53)	5.7 ± 2.6/6.9 ± 2.7	7.3 ± 2.0 / 7.5 ± 1.8	< 0.01/0.06
Revision cases (N =13)	5.2 ± 3.1/5.1 ± 2.7	6.3 ± 2.9 / 7.0 ± 2.1	0.16/0.03
Utrecht Questionnaire	**Preop (T1) mean ± SD**	**Postop (T3 or T2) mean ± SD**	**p-value**
Individual questions			
1. Concern about nasal appearance	3.4 ± 1.2	1.8 ± 1.0	< 0.01
2. Often concerned	2.9 ± 1.3	1.5 ± 0.8	< 0.01
3. Concern affecting daily life	2.2 ± 1.4	1.2 ± 0.6	< 0.01
4. Concern affecting relationship with others	2.1 ± 1.3	1.2 ± 0.6	< 0.01
5. Stressed about nasal appearance	2.5 ± 1.4	1.3 ± 0.7	< 0.01
Sum scores			
Whole surgical population	13.1 ± 5.6	7.1 ± 3.3	< 0.01
Primary cases (N = 53)	12.8 ± 5.7	7.3 ± 3.5	< 0.01
Revision cases (N = 13)	14.5 ± 4.8	6.6 ± 1.7	< 0.01
VAS nasal appearance	**Preop (T1) mean ± SD**	**Postop (T3 or T2) mean ± SD**	**p-value**
Whole surgical population	4.1 ± 2.2	7.2 ± 1.4	< 0.01
Primary cases (N = 53)	4.0 ± 2.3	7.3 ± 1.4	< 0.01
Revision cases (N = 13)	4.6 ± 2.1	7.1 ± 1.1	< 0.01

Abbreviations: NOSE: Nasal Obstruction Symptom Evaluation; VAS: Visual Analogue Scale; SD: standard deviation

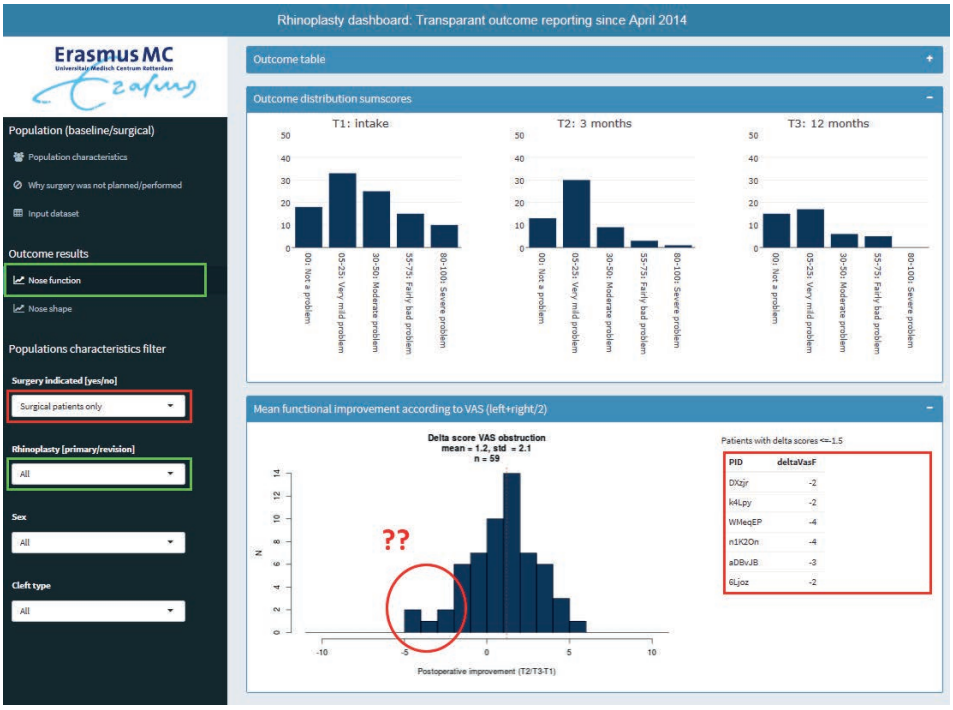


Figure 2. Dashboard snapshot showing button 4 (green: nose function). Above: access to severity distribution of nasal obstruction according to the mean overall Nasal Obstruction Symptom Evaluation scale sum scores before surgery (T1) and 3 (T2) and 12 (T3) months after surgery. The graphs show that after surgery the majority of patients experience none or very mild problems related to nasal obstruction. Below: a normal distribution of mean VAS score improvement following secondary cleft lip rhinoplasty is seen. The IDs of patients on the far left of the curve (red ellipse) are automatically presented (red rectangle) for a targeted review of their medical charts and potential explanations for this unwanted outcome. IDs shown are altered to warrant privacy. *VAS: Visual Analogue Scale.*

The third part of the dashboard depicts the mean outcome delta VAS, which is a (developing) normal distribution curve of mean VAS score improvement that shows a mean (SD) postoperative improvement of 1.3 (2.1) points (**Figure 2**). This improvement is almost similar in primary (1.2 [2.0]) and revision (1.5 [2.4]) cases. Six patients reported increased nasal obstruction after rhinoplasty. The dashboard automatically identifies these patients and allows retrospective review of the medical record to find possible explanations (e.g. insufficient management of expectations, mucosal involvement, suboptimal surgical plan/performance, or functional decrease caused by maneuvers used for aesthetic improvement.

Aesthetic Outcome Results of the Surgical Population

The fourth button on the dashboard provides access to information about the severity of concerns related to nasal appearance before intake and 3 and 12 months after surgery. This part of the dashboard has a design and purpose identical to that for nasal obstruction. The mean (SD) preoperative UQ sum score was 13.1 (5.6), which decreased to 7.1 (3.3) after surgery. Revision cases reported a slightly more beneficial overall aesthetic end result (mean [SD] UQ sum score, 6.6 [1.7]) compared with primary cases (7.3 [3.5]). The mean (SD) VAS score improved from 4.1 (2.2) to 7.2 (1.4).

The graphical distribution of sum scores shows that, before surgery, most patients reported little to moderate concern regarding nasal appearance despite the usual presence of conspicuous nasal deformities. Only 9 patients showed very much concern but no signs of a disturbed body perception in relation to nasal appearance. One year after surgery, only 2 patients reported much concern with nasal appearance, whereas most patients had no or little concern (**Figure 3**).

The aesthetic distribution curve shows a mean (SD) overall postoperative VAS score change of 3.0 (2.2) (**Figure 3**). The present number of patients is insufficient to conclude that aesthetic improvement is normally distributed. Only 2 patients reported a decrease in nasal appearance after surgery, and their patient IDs were automatically presented in order to review their medical records for explanations of these unwanted results.

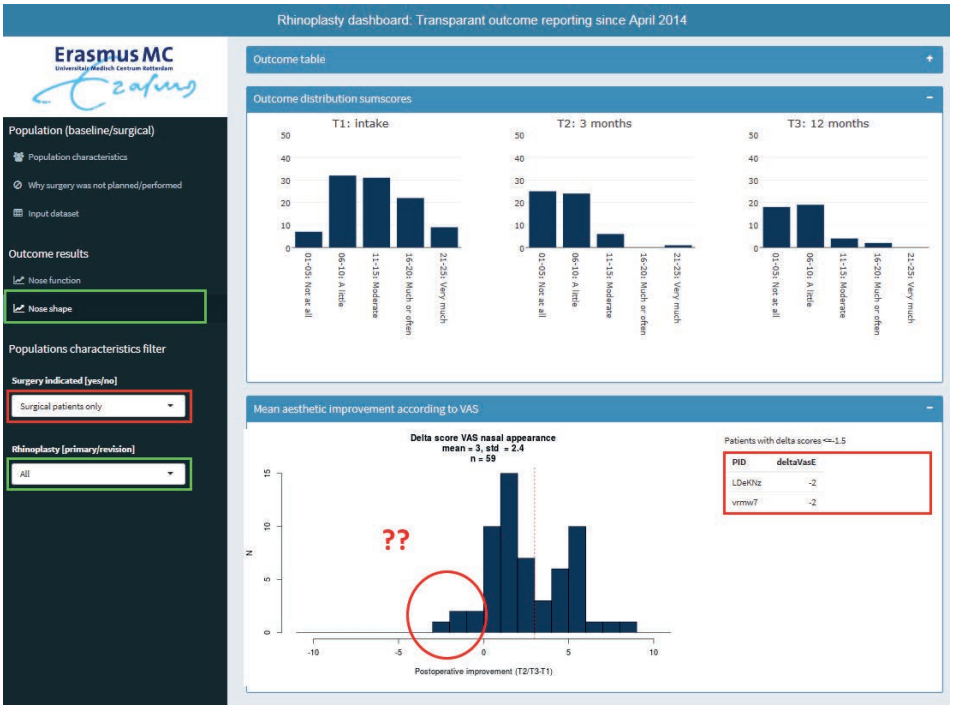


Figure 3. Dashboard snapshot showing button 5 (green: nose shape). Above: access to severity distribution of concerns related to nasal appearance according to the mean overall Utrecht Questionnaire sum scores before surgery (T1) and 3 (T2) and 12 (T3) months after surgery. Below: a distribution of mean aesthetic visual analogue scale score improvement following secondary cleft lip rhinoplasty. The IDs of patients on the far left of the curve (red ellipse) are automatically presented (red rectangle) for a targeted review of their medical charts and potential explanations for this unwanted outcome. IDs shown are altered to warrant privacy. *VAS; Visual Analogue Scale.*

Discussion

This article describes a short, automated secondary cleft rhinoplasty outcome routine that easily fits a busy rhinoplasty practice. The outcome routine translates nasal form and function disturbances as experienced by patients into data and quantifies change after surgery. The automated routine allows surgeons to be more transparent about secondary cleft lip rhinoplasty performance in real time.

The concept of measuring functional and aesthetic gain from secondary cleft lip rhinoplasty is not new.[11-13] However, a global consensus on which outcome measures are most suited to compare results between centers and institutions and a protocol on how outcome research should be performed remain absent, to our knowledge. This missed opportunity is acknowledged worldwide, resulting in the development of new instruments and emerging collaborations, such as ICHOM, which focuses on cross-cultural comparisons of medical performance by designing questionnaire sets that measure outcomes that matter to patients. For the cleft lip nose specifically, ICHOM implemented the NOSE scale to measure nasal obstruction on patients aged 8 and 12 years but not specifically before and after surgery. Our rhinoplasty outcome routine can therefore complement ICHOM's efforts.

Although the primary aim of this study was to test the functionality of a practical prospective health outcome monitor as part of daily practice to provide transparency on overall secondary cleft lip rhinoplasty performance, unexpected additional benefits that upgraded our cleft care standard were encountered. The first benefit was seen in patient selection and avoidance; prior to consultation, the surgeon has data on the severity and side of nasal obstruction and data on severity of concerns about nasal appearance. These data facilitate a targeted anamnesis and physical examination. Patients with very low scores who present for rhinoplasty are perhaps more curious about surgery than actually motivated to undergo it, and patients with high scores on trick questions give clinicians a forewarning of potential psychological issues.

A second benefit was noticed during patient counseling and expectation management. The visual display of expected functional and aesthetic improvements helps patients to keep their expectations realistic. Furthermore, it provides patients with true, surgeon-specific averages on which they can base their decision to undergo surgery.

A third benefit was seen in avoiding revision; the availability of patient-reported outcome measures data prior to follow-up consultations provides insight into the patient's perspective of the surgical result. A warning about a patient's dissatisfaction with functional or aesthetic

outcomes gives the surgeon time to prepare for a difficult consultation. Alternatively, knowing that the patient is very satisfied with the outcomes is not only a relaxing thought, but also warns the physician not to go into detail about noticeable or subtle asymmetries and especially to avoid offering revision rhinoplasty (**Figure 4**).

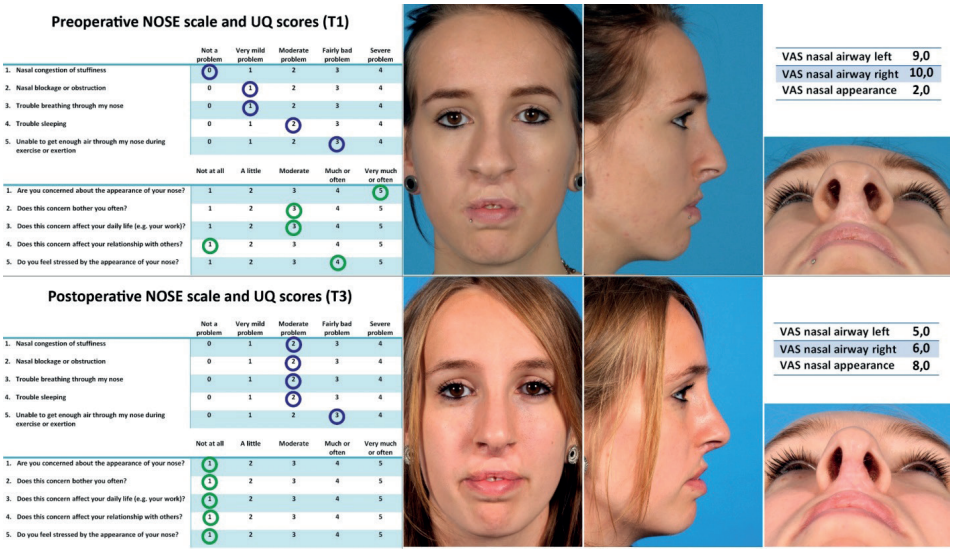


Figure 4. Front, side, and basal views of a bilateral cleft lip nose before surgery (T1). B, Front, side, and basal views 1 year after rhinoplasty (T3). A septal reconstruction, tip refinement using tongue-in-groove, suture techniques, and alar base reduction was performed. The anterior nasal spine and caudal septum could have been reduced more, and there is some asymmetry of the facets. However, the patient reported extreme aesthetic satisfaction. A revision for aesthetic purposes was therefore never part of the conversation. Her increased nasal obstruction was discussed and based on seasonal rhinitis symptoms and some narrowing of the nares caused by alar base reduction. She was not interested in revision surgery. *NOSE, Nasal Obstruction Symptom Evaluation; UQ, Utrecht Questionnaire; VAS, Visual Analogue Scale.*

A fourth benefit is the ability to show third parties (e.g. government, review committees, or insurance companies) the severity of nasal obstruction and concerns with nasal appearance in patients with cleft lip. Combined with data about the potential improvement from surgery, this benefit justifies the place for secondary cleft lip rhinoplasty in a tertiary referral center. Furthermore, specific questions (e.g. What are the characteristics of our cleft lip rhinoplasty population? How many procedures do we perform? What is the amount of revision cases?) can be answered without complex retrospective data analysis.

For most patients born with a cleft lip and palate, secondary rhinoplasty is the final surgical correction after a long history of medical care. In contrast to most other patients undergoing rhinoplasty, they are used to the hospital setting, physicians focusing on functional and aesthetic improvement, and undergoing surgical procedures. These factors influence behavior, body image in relation to facial appearance, and expectations from surgical procedures. More specifically, patients with a cleft lip and palate tend to report moderate concerns about nasal appearance, even in the presence of conspicuous deformities, and have few concerns about undergoing rhinoplasty. Functionally, they have less outspoken demands, even in the presence of severe septal deviation. This attitude is reflected by the baseline values found in our cohort; the preoperative NOSE score of 3.8 is relatively low compared with preoperative mean scores of patients undergoing rhinoplasty for general nasal airway obstruction.[14]

Limitations

The success of a prospective outcome routine depends on several factors. It must be quick, simple, and based on accepted instruments to warrant high patient cooperation and generate meaningful data. Our choice of instruments was based on a literature review focusing on previously mentioned criteria. The NOSE scale and UQ were validated for populations without a cleft lip and palate who underwent rhinoplasty; theoretically, the performance of these scales in this population is not known. However, the constructs important to patients both with and without a cleft lip and palate who are undergoing rhinoplasty (nasal breathing and nasal shape) remain the same. It has not escaped our attention that other outcome instruments are available, which emphasizes the need for global consensus on how to measure rhinoplasty outcomes. Another concern is that, despite efforts to minimize bias and confounding factors, critics might question the value of data derived from subjective patient-reported instruments. Factors that are not under investigator's control could affect scores and might jeopardize global outcome comparisons.

Conclusions

We hope that the concept of prospective, unselected data collection, follow-up, and automated real-time outcome reporting as well as the additional practical benefits appeal to others. We believe that continuation of patient inclusions will ultimately lead to a power increase, providing a higher statistical level of evidence and eventually normal distributions of functional and aesthetic gain from surgery. This gain can then be communicated more confidently to our patients and third parties interested in health outcomes.

References

1. International Consortium for Health Outcomes Measurement. Available at: https://www.ichom.org/. *Accessed April 4, 2020.*
2. Datema FR, van Zijl F, van der Poel EF, Baatenburg de Jong RJ, Lohuis P. Transparency in Functional Rhinoplasty: Benefits of Routine Prospective Outcome Measurements in a Tertiary Referral Center. *Plast Reconstr Surg.* 2017;140(4):691-702.
3. Colaianni CA, Levesque PA, Lindsay RW. Integrating Data Collection Into Office Work Flow and Electronic Health Records for Clinical Outcomes Research. *JAMA Facial Plast Surg.* 2017;19(6):528-532.
4. 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):157-163.
5. van Zijl F, Timman R, Datema FR. Adaptation and validation of the Dutch version of the nasal obstruction symptom evaluation (NOSE) scale. *Eur Arch Otorhinolaryngol.* 2017;274(6):2469-2476.
6. Lohuis PJ, Hakim S, Duivesteijn W, Knobbe A, Tasman AJ. Benefits of a short, practical questionnaire to measure subjective perception of nasal appearance after aesthetic rhinoplasty. *Plast Reconstr Surg.* 2013;132(6):913e-923e.
7. Lohuis PJFM, Datema FR. Patient satisfaction in Caucasian and Mediterranean open rhinoplasty using the tongue-in-groove technique: Prospective statistical analysis of change in subjective body image in relation to nasal appearance following aesthetic rhinoplasty. *Laryngoscope.* 2015;125(4):831-836.
8. LimeSurvey. https://www.limesurvey.org. *Accessed March 1, 2018.*
9. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. shiny: Web Application Framework for R. R package version 0.14.1. *Available at:* https://CRAN.R-project.org/package=shiny. *Accessed December 22, 2016.*
10. Lipan MJ, Most SP. Development of a severity classification system for subjective nasal obstruction. *JAMA Facial Plast Surg.* 2013;15(5):358-361.
11. Vass G, Mohos G, Bere Z, et al. Secondary correction of nasal deformities in cleft lip and palate patients: surgical technique and outcome evaluation. *Head Face Med.* 2016;12(1):34.
12. Gassling V, Koos B, Birkenfeld F, Wiltfang J, Zimmermann CE. Secondary cleft nose rhinoplasty: Subjective and objective outcome evaluation. *J Craniomaxillofac Surg.* 2015;43(9):1855-1862.
13. Byrne M, Chan JC, O'Broin E. Perceptions and satisfaction of aesthetic outcome following secondary cleft rhinoplasty: evaluation by patients versus health professionals. *J Craniomaxillofac Surg.* 2014;42(7):1062-1070.
14. Rhee JS, Sullivan CD, Frank DO, Kimbell JS, Garcia GJ. A systematic review of patient-reported nasal obstruction scores: defining normative and symptomatic ranges in surgical patients. *JAMA facial plastic surgery.* 2014;16(3):219-225; quiz 232.

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Chapter 6

The Rhinoplasty Healthcare Monitor: Using validated questionnaires and a web-based outcome dashboard to evaluate personal surgical performance

Floris V.W.J. van Zijl MD, Peter J.F.M. Lohuis MD PhD, Frank R. Datema MD PhD.

Facial Plast Surg Aesthet Med. 2021 Feb 22.

Abstract

Background: Self-assessment provides valuable feedback in the life-long process of mastering rhinoplasty. This study presents a method to measure and evaluate data-based performance of a single surgeon using a web-based dashboard.

Methods: In this prospective analytic cohort study, all patients referred to the senior author for functional–aesthetic (revision) rhinoplasty between April 2014 and September 2020 are included. Patients completed the Nasal Obstruction Symptom Evaluation scale, Utrecht Questionnaire, and visual analog scales before and after rhinoplasty. Questionnaire scores were exported to a customized web-based dashboard: the rhinoplasty health care monitor. Supported by real-time graphic output, this monitor automatically analyzes functional and aesthetic outcomes.

Results: Of 603 referred patients, 363 were eligible for rhinoplasty. Mean Nasal Obstruction Symptom Evaluation scale scores decreased from 66.6 ± 23.5 to 23.2 ± 24.0 ($p < 0.001$), and mean Utrecht Questionnaire scores decreased from 12.2 ± 6.3 to 7.1 ± 3.9 ($p < 0.001$) 1 year after surgery. The rhinoplasty health care monitor visualizes numerous outcome parameters that help the surgeon to analyze results, identify learning needs, and detect trends in performance development.

Conclusions: This automated outcome dashboard transparently measures individual surgeon performance. Gauging performance provides means to enhance surgical development and, consequently, patient satisfaction.

Introduction

Self-assessment is an important ingredient to becoming a better rhinoplasty surgeon. Evaluating personal performance is a fundamental drive of professional development and is linked closely to the quality of patient care.[1,2] Although rhinoplasty surgeons acknowledge the importance of self-reflection, this typically encompasses the ad hoc assessment of pre- and postoperative photographs.[3] Such evaluations are valuable, but also subjective and difficult to quantify. Instead, systematically measuring meaningful quantifiable outcomes among all patients encountered for a particular procedure may provide a more transparent reflection of performance. Routinely collecting standard data among patients allows surgeons to create large unselected data sets with clinical outcomes. Analyzing such data sets can help to identify learning and improvement needs or provide data-driven motivation to change concepts or techniques.

Since 2014, we routinely measure and monitor patient outcomes in our rhinoplasty practice. Each patient referred to our practice completes simple, relevant, and validated questionnaires at first consultation and, if applicable, after surgery. All patient characteristics and questionnaire scores are entered into a database that is uploaded to a web-based customized dashboard. This dashboard automatically displays graphical analyses of patient characteristics and patient outcomes of the practice, without the need for complex statistics. In 2017, we have described the implementation of this "rhinoplasty healthcare monitor" (RHM) and its advantages in the clinical environment (e.g. preparation for consultations, patient empowerment, and managing expectations) as well as the advantages on a third-party level (providing evidence-based data to insurance companies, governments, or other parties interested in health care performance).[4] Now that the RHM has been structurally embedded in our practice for more than 6 years, we aim to demonstrate the long-term results of this outcome routine and address evaluative features that specifically benefit the surgeon. We describe the value of the RHM in assessing surgical results through the eyes of the patient and demonstrate how to effectively use your own data as feedback on personal performance.

Methods

Hospital setting and study population

The Erasmus Medical Center in Rotterdam is a university teaching hospital that serves a population of ~3 million people in the southwestern area of The Netherlands. Each patient who was referred to the last author of this article (F.R.D.) between April 2014 and September 2020 to explore the indication for functional–aesthetic rhinoplasty (primary and revision) was included in this study. Candidates for septoplasty only, craniofacial patients, and patients with a cosmetic indication only were excluded. The cohort is unselected and consecutive, and all rhinoplasties were performed by one surgeon (F.R.D.). Since standard care is evaluated, this study was exempted from institutional review board approval. Informed consent for study participation was obtained from all patients.

Routine outcome measurements

In our practice, patient-reported outcomes are routinely collected on three different time points: preoperatively (at first consultation), 3 months after surgery, and 12 months after surgery. During each time point, patients are asked to complete three questionnaires: the Nasal Obstruction Symptom Evaluation (NOSE) scale, the Utrecht Questionnaire (UQ), and visual analog scales (VASs). These questionnaires are completed in person in the waiting room, before consultation. The NOSE scale measures symptoms related to nasal obstruction and contains five questions scored on a 5-point Likert scale.[5] Sum scores are multiplied by 5, resulting in a range from 0 to 100. To provide context to this score, patients can be categorized as having mild (range, 5–25), moderate (30–50), severe (55–75), or extreme (80–100) nasal obstruction.[6] Additional to the NOSE, two VASs are included measuring left- and right-sided nasal obstruction separately. The VASs range from 0 (complete obstruction) to 10 (a clear nose). To measure patient concern with nasal appearance, the UQ is administered. The UQ is validated and has a design similar to the NOSE but without a sum score multiplier: sum scores range from 5 to 25.[7] UQ sum scores can be distributed into five severity groups: no concern (range, 1–5), a little concern (6–10), moderate concern (11–15), much concern (16–20), and very much concern (21–25) about nasal appearance. The UQ is complemented with a single VAS rating patient-reported nasal appearance, ranging from 0 (very ugly) to 10 (very nice).

The rhinoplasty health care monitor

Shortly after eligible patients have been seen on the outpatient clinic (either new patients or postoperative visits), the consulted surgeon enters questionnaire scores and corresponding patient characteristics into an IBM SPSS statistics database for

Windows, Version 25 (IBM Corp., Armonk, NY). This creates a prospectively acquired data set in which each patient referred to our practice is enrolled. Patients who do not show at postoperative visits are contacted and asked to reschedule or at least return the questionnaires by e-mail.

To analyze the data, the surgeon can upload the SPSS file at any given time to a web-based graphic outcome dashboard: the rhinoplasty health care monitor (RHM). This customizable dashboard was programmed with shiny: Web Application Framework for R (R package version 0.13.112), allowing real-time automated outcome analyses based on the uploaded data set. The RHM analyzes the database as a whole cohort, but if needed results can be stratified according to several patient characteristics (e.g. gender, primary vs. revision). All outcome figures in this article are screenshots from the dashboard. To analyze differences in pre- and postoperative scores, a *t*-test for paired data was used.

Results

All results are derived directly from the RHM. By visiting www.healthcaremonitor.nl the reader can view the layout and functionality of the RHM and, since the anonymized database of our practice has been uploaded as an example, visualize all results related to this study.

Patient characteristics and overall outcome results

The tab "population characteristics" of the RHM reveals that from April 2014 to September 2020, 603 consecutive patients have been referred to explore the indication for functional-aesthetic (revision) rhinoplasty. Of these 603 included patients, 363 were eligible for surgery. The mean age of the surgical population ($n = 363$) was 36 years (range 16-75), and gender was distributed equally (184 male, 50.7% and 179 female (49.3%). More than half ($n = 187$; 51.5%) had one or more unsatisfactory rhinoplasties performed elsewhere. From the 363 eligible patients, 335 (92%) have presently had their surgery performed and 28 are on the waiting list. Preoperative questionnaire scores are available for 357 patients (98%). From the 335 operated patients, 284 (85%) have currently completed short-term follow-up (3 months after surgery) and 229 (68%) have currently completed long-term follow-up (12 months after surgery). The overall mean NOSE sum score decreased from 66.6 ± 23.5 (severe nasal obstruction) to 23.2 ± 24.0 (mild nasal obstruction) 1 year after surgery ($p < 0.001$). The mean VAS scores for left and right nasal obstruction increased from 4.2 ± 2.6 and 4.2 ± 2.7, respectively, to 7.4 ± 1.9 and 7.2 ± 2.1 ($p < 0.001$). The mean overall UQ sum score dropped from 12.2 ± 6.3 (moderate concern with nasal appearance) to 7.1 ± 3.9 (little concern with nasal appearance) 1 year after surgery ($p < 0.001$). The mean nasal appearance VAS score was 4.4 ± 2.3 before surgery and improved to 7.2 ± 1.7 ($p < 0.001$) 1 year after surgery. There were no significant differences between males and females in postoperative sum scores.

Identification of patients with poor or extremely good outcomes

In the functional and aesthetic "outcome results" tab of the RHM, the distribution of mean VAS improvement after surgery is depicted. Ranging from a decrease of three points to an increase of eight points after surgery, the bar chart displays the number of patients assigned to each score change. The distribution of mean functional improvement based on the average VAS score (left + right)/2) shows a developing Gaussian curve of normal distribution with a mean improvement score of 3.1 ± 2.0 SD after rhinoplasty (**Figure 1**). The distribution of aesthetic improvement according to VAS score shows a mean improvement score of 2.9 ± 2.6 (**Figure 2**). Patients on the left side of these curves report substantially worsened nasal obstruction or nasal

appearance, whereas patients on the right side report above-average improvement. In the RHM, clicking on a bar of choice generates a list of patient IDs of all patients reporting that specific improvement or deterioration after surgery on the right side of the screen. This allows the surgeon to critically appraise the medical charts of these patients, for instance to identify mistakes that can be learnt from.

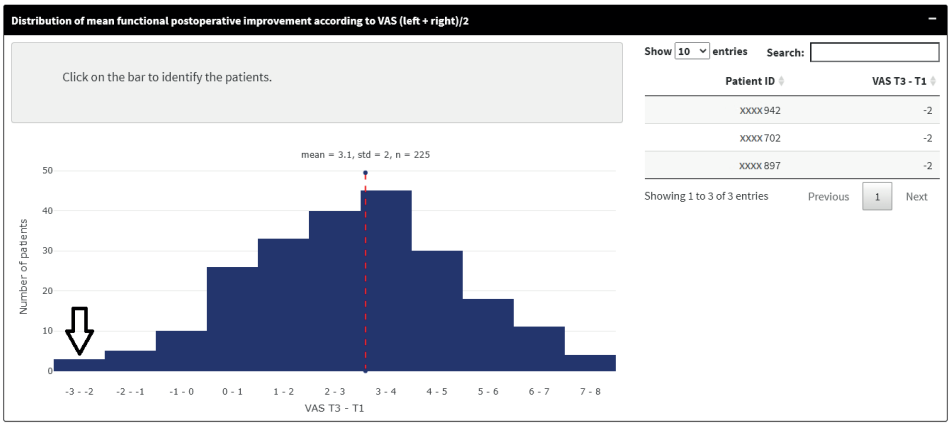


Figure 1. Screenshot of the dashboard depicting the distribution of mean functional VAS score change after surgery. A Gaussian curve with a mean improvement of 3.1 ± 2.0 is shown. By clicking on the far left bar, the ID of three patients who reported severely increased nasal obstruction (defined as ≥2 VAS points) 1 year after rhinoplasty is automatically presented in the top right corner. This facilitates a targeted appraisal of their medical charts to identify potential mistakes. *VAS: Visual Analogue Scale*.

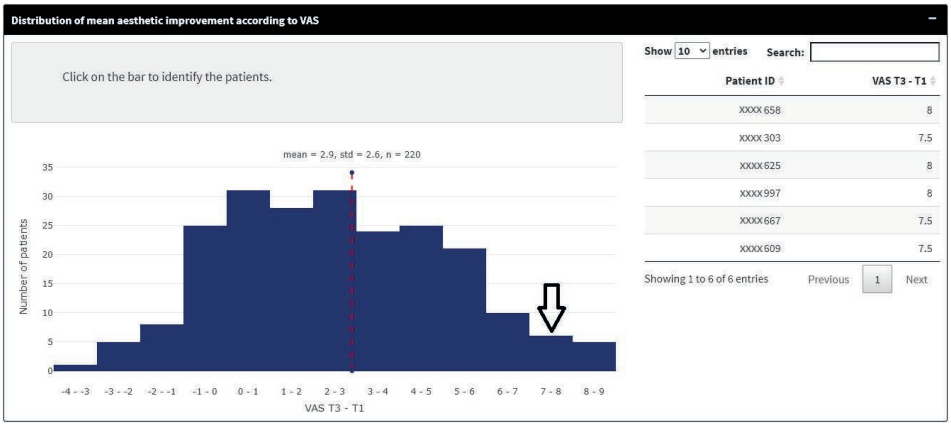


Figure 2. Screenshot of the dashboard showing the distribution of mean aesthetic VAS score changes after surgery. A Gaussian curve in development with a mean aesthetic improvement of 2.9 points ±2.6 is shown. Again, clicking on any bar in the graph reveals the patient IDs corresponding to that category of postoperative change. *VAS: Visual Analogue Scale*.

Annual performance assessments

The tab "annual self-evaluations" graphically depicts surgical performance per year. For each year, postoperative scores among all patients who have been operated on in that year have been averaged, resulting in a mean postoperative score achieved per year of surgery. This plot indicates the progression that the surgeon annually makes in terms of achieved patient outcomes (**Figures 3 and 4**). These curves inform the surgeon of annual performance, give an indication of the gradient of performance development, and may identify trends that are worth evaluating. Encountered stagnation or deterioration may motivate the surgeon to critically evaluate adverse outcomes, dedicate more time to studying rhinoplasty, visit a course, or arrange a working visit.

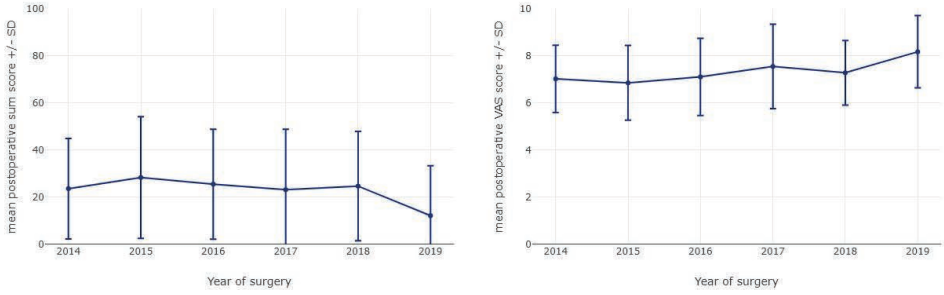


Figure 3. Screenshot of the "annual functional performance" tab demonstrating average postoperative NOSE (left) and VAS (right) scores of all patients operated in a particular year. For the NOSE, a lower score indicates an improvement in nasal function, whereas for the VAS, a higher score indicates improved nasal function. *VAS: Visual Analogue Scale; NOSE: Nasal Obstruction Symptom Evaluation.*

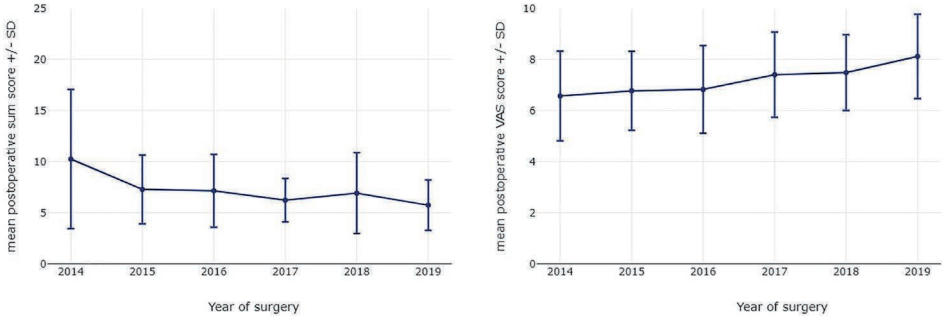


Figure 4. Screenshot of the "annual aesthetic performance" tab demonstrating average postoperative UQ (left) and VAS (right) scores of all patients operated in a particular year. For the UQ, a lower score indicates an improvement in patient-reported nasal appearance, whereas for the VAS, a higher score indicates improved patient-reported nasal appearance. *VAS: Visual Analogue Scale; UQ: Utrecht Questionnaire.*

Discussion

Near-total dedication to the study and practice of rhinoplasty is required to achieve reliable and consistently good results. An indispensable component of improving results is to actually measure results. Yet, physicians systematically monitoring their own performance are scarce and many, including rhinoplasty surgeons, are relying on gut sense to determine how well they perform.[8,9] In 2014, we have enrolled a prospective outcome routine in our rhinoplasty practice and developed the RHM: a web-based dashboard that provides real-time automated outcome analyses. Routine outcome measurements in rhinoplasty have numerous advantages, including improved patient empowerment, benchmarking surgeons or surgical techniques, measuring value in value-based health care and building big data registries.[10] In this article we have focused on the self-evaluative potential of routine outcome measurements for the individual surgeon.

The RHM contains two features that enhance self-assessment, the first of which is identifying patients with adverse (or surprisingly favorable) outcomes. Since it is argued that clinicians develop competence by learning from their mistakes, an adage that is echoed by experienced rhinoplasty surgeons, a targeted critical appraisal of these cases is highly informative.[11,12] It allows the surgeon to identify certain patterns or decisions (e.g. specific surgical technique, suboptimal surgical execution, patient selection, and expectation management) that are associated with these outcomes, providing valuable lessons for the future. Furthermore, by identifying possible causes for these outcomes, the surgeon can tailor subsequent learning activities or courses to that particular problem, instead of gravitating toward broad, general, or randomly chosen educational activities.[13] A concrete example: in our cohort we found similarities in five of eight (both functional and aesthetic) severely dissatisfied patients (deterioration of two points or more), all experiencing (partial) resorption of Tutoplast™ processed costal cartilage (TPCC). In two patients, the cartilage was used to correct a deep saddle nose deformity with severe skin retraction and subtotal septal perforation. In three other revision cases, the cartilage was unable to counter stress related to a thick and heavy skin–soft tissue envelope. Consequently, in subsequent patients with similar pathology, TPCC use was avoided and replaced by either autologous rib cartilage or iliac crest bone.

The second self-evaluative feature of the RHM is the display of annual performance, allowing the surgeon to assess longitudinal progress. By graphically representing performance as a function of time, this plot can be considered a surrogate rhinoplasty learning curve. Understanding the development of the individual learning curve and visualizing its shape provides valuable feedback: it helps to acknowledge limitations

early on in a rhinoplasty career, underlines the necessity of continuous education and focus, and provides motivation to further improve surgical outcomes.[3] It is important to note that patient factors such as complex anatomy and varying case mix can affect surgical performance.[14] It is essential to monitor such possibly confounding factors, as they could explain unexpected variations in performance. A decrease of annual performance compared with the year before may for example actually be a result of taking on more revision cases. Using the RHM, an overview of patient characteristics can be visualized per year to assess comparability of such confounding factors throughout the years. Utilizing a thorough definition of the caseload helps to put performance in perspective and may even facilitate interclinician comparison.

We did not enroll objective outcome measures to monitor surgical performance due to a lack of correlation found between these instruments and patient symptoms.[15] Ultimately, the goal of a rhinoplasty is to satisfy the patient's functional and aesthetic expectations of the procedure. Symptoms related to nasal obstruction and patient satisfaction with nasal appearance are widely considered the most important indicators of rhinoplasty success,[16-18] and, therefore, these indices are the most valid parameters of surgical performance. Indeed, patient-reported outcomes contain bias per definition, and in individual cases, a patient-reported outcome measure (PROM) score may not always be representative of surgical quality. However, the analyses performed by the RHM are based on cohort-level averages, neutralizing individual outliers. A second remark related to this study is that there is no global consensus on what PROMs should be used to measure rhinoplasty outcomes. The NOSE scale is recommended to measure functional outcome based on the quality of validation studies, interpretability, and feasibility.[19] Regarding aesthetic outcomes however, there is no clear preference for one instrument. We have implemented the validated UQ in our outcome routine because it is short, simple, and carries the specific advantage of two "trick" questions that aid in identifying patients with unrealistic expectations or body dysmorphic disorder.[7]

To demonstrate the advantages of the RHM to other practitioners, we have launched (www.healthcaremonitor.nl). Apart from visualizing the functionality of the RHM and results of this study, rhinoplasty surgeons can also analyze their own results through this website. A blank SPSS file can be downloaded as a template and used to collect patient data. The filled file can be uploaded into the RHM, allowing surgeons to view the output of their own data. We underline that uploaded data will not be stored or used for benchmarking. The goal is merely to share the advantages of the RHM and encourage participants to contribute to improving the RHM, potentially creating a common ground for routine outcome measurements and evaluations in rhinoplasty.

Conclusion

In this article, we describe the results of a 6-year prospective rhinoplasty outcome routine based on validated patient-reported outcome measures, facilitating a transparent and accurate reflection of surgical performance. The user-friendly dashboard provides clear insights into achieved results and enhances self-evaluative opportunities, with the overarching goal of improving patient outcomes.

References

1. Hays RB, Jolly BC, Caldon LJ, et al. Is insight important? measuring capacity to change performance. *Med Educ.* 2002;36(10):965-971.
2. Westberg J, Jason H. Fostering learners' reflection and self-assessment. *Fam Med.* 1994;26(5):278-282.
3. Lohuis PJFM. Advanced Caucasian and Mediterranean Rhinoplasty. Amsterdam: Kugler Publications; 2014:2-19.
4. Datema FR, van Zijl F, van der Poel EF, Baatenburg de Jong RJ, Lohuis P. Transparency in Functional Rhinoplasty: Benefits of Routine Prospective Outcome Measurements in a Tertiary Referral Center. *Plast Reconstr Surg.* 2017;140(4):691-702.
5. 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):157-163.
6. Lipan MJ, Most SP. Development of a severity classification system for subjective nasal obstruction. *JAMA Facial Plast Surg.* 2013;15(5):358-361.
7. Lohuis PJ, Hakim S, Duivesteijn W, Knobbe A, Tasman AJ. Benefits of a short, practical questionnaire to measure subjective perception of nasal appearance after aesthetic rhinoplasty. *Plast Reconstr Surg.* 2013;132(6):913e-923e.
8. Trebble TM, Carder C, Paul M, et al. Determining doctors' views on performance measurement and management of their clinical practice. *Future Hosp J.* 2015;2(3):166-170.
9. Audet AM, Doty MM, Shamasdin J, Schoenbaum SC. Measure, learn, and improve: physicians' involvement in quality improvement. *Health Aff (Millwood).* 2005;24(3):843-853.
10. Colaianni CA, Levesque PA, Lindsay RW. Integrating Data Collection Into Office Work Flow and Electronic Health Records for Clinical Outcomes Research. *JAMA Facial Plast Surg.* 2017;19(6):528-532.
11. Duffy FD, Holmboe ES. Self-assessment in lifelong learning and improving performance in practice: physician know thyself. *JAMA.* 2006;296(9):1137-1139.
12. Toriumi DM. Nasal Tip Contouring: Anatomic Basis for Management. *Facial Plast Surg Aesthet Med.* 2020;22(1):10-24.
13. Galbraith RM, Hawkins RE, Holmboe ES. Making self-assessment more effective. *J Contin Educ Health Prof.* 2008;28(1):20-24.
14. Subramonian K, Muir G. The 'learning curve' in surgery: what is it, how do we measure it and can we influence it? *BJU Int.* 2004;93(9):1173-1174.
15. Spataro E, Most SP. Measuring Nasal Obstruction Outcomes. *Otolaryngol Clin North Am.* 2018;51(5):883-895.
16. Rhee JS, Weaver EM, Park SS, et al. Clinical consensus statement: Diagnosis and management of nasal valve compromise. *Otolaryngol Head Neck Surg.* 2010;143(1):48-59.
17. Ching S, Thoma A, McCabe RE, Antony MM. Measuring outcomes in aesthetic surgery: A comprehensive review of the literature. *Plastic and reconstructive surgery.* 2003;111(1):469-480.
18. Cannon DE, Rhee JS. Evidence-based practice: functional rhinoplasty. *Otolaryngol Clin North Am.* 2012;45(5):1033-1043.
19. van Zijl F, Mokkink LB, Haagsma JA, Datema FR. Evaluation of Measurement Properties of Patient-Reported Outcome Measures After Rhinoplasty: A Systematic Review. *JAMA Facial Plast Surg.* 2019.

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Chapter 7

The value of averageness in aesthetic rhinoplasty: Humans like average noses

Floris V.W.J. van Zijl MD, David I. Perrett PhD, Peter J.F.M. Lohuis MD PhD,
Carolina E. Touw MD, Dengke Xiao PhD, Frank R. Datema MD PhD

Aesthet Surg J. 2020 Nov 19;40(12):1280-1287.

Abstract

Background: The aesthetic ideal of the nose eludes clear definition. Averageness may be an important determinant of ideal nasal shape: research has shown that averageness plays an important role in the human perception of facial attractiveness.

Objectives: The aim of this study was to test whether an averaged nasal shape is attractive, and whether deviation away from average is associated with decreased attractiveness.

Methods: Photographic series of the face were obtained from 80 Caucasian female volunteers aged 25-40 years. A mathematically averaged composite image was created from the first 40 volunteers. Forty-one panel members were recruited to judge the attractiveness of the nose of each original image and the composite, based on a 5-point Likert scale ranging from 1 (very ugly) to 5 (very pretty). Deviation of nasal shape from average was calculated by principal components analysis of standardized nasal landmarks.

Results: Twenty-one respondents were male (51%). The mean age of the respondents was 35.3 [15.6] years. The rating of the composite was significantly higher than the distribution of ratings for the 80 original nose images (4.2 vs 2.8, t = 31.24, $P < 0.001$). The rating of the original nose images correlated negatively with deviation from average shape ($r = -0.40$, n = 80, $P < 0.001$).

Conclusions: In Caucasian females, a mathematically averaged nose is an attractive nose. Furthermore, the more an individual nose shape resembles the average shape, the more attractive it is rated. Calculating deviation from average before and after rhinoplasty may aid in objectively measuring aesthetic rhinoplasty outcome.

Introduction

A comprehensive understanding of facial aesthetics is of paramount importance to rhinoplasty surgeons. In pursuit of high patient satisfaction, every rhinoplasty surgeon deals with the dilemma of what constitutes the ideal nose. Traditionally, neoclassical canons that focus on optimal alignment, angularity, and proportionality of the face serve as guidelines. However, large cross-cultural anthropometric studies show that only a minor percentage of the population exhibits these aesthetic ideals, and attractive faces do not conform to them any more than do less attractive faces.[1-3]

Evolutionary psychologists have studied facial attractiveness extensively, linking the importance of attractiveness to human interactions and sexual selection. They identified possible cues to what humans find universally attractive in others, such as averageness, symmetry, youthfulness, and sexual dimorphism.[4] Of these 4 factors, averageness has been proposed as the most important factor in perceiving facial attractiveness.[5] Averageness in this context does not refer to typical mean in the sense of "common" or "frequently occurring in the population," but more to a mathematically averaged, computer-manipulated composite of whole faces that is based on mean numerical values across the face. In 1990, Langlois and Roggman[6] produced such an "average composite," and found that this composite was perceived as more attractive than the individual grayscale images of the faces used to create the composite. Later, in 2003, Halberstadt and Rhodes[7] reported that our attraction to averageness was also prevalent in birds, fish, and even automobiles.

To date, the relation between averageness and attractiveness has not been tested for the nose specifically. We hypothesized that a digitally averaged composite nose of a specific population is rated as more attractive than the individual noses used to create it. If this proves to be so, the composite nose might serve as an aesthetic surgical reference for that specific population seeking rhinoplasty. Furthermore, taking the composite nose as a reference, we attempt to calculate the degree to which an individual nasal shape deviates from average. We hypothesized that deviation away from average correlates with decreased attractiveness. If this hypothesis can be confirmed, this "shape deviation score" could be applied to objectively quantify aesthetic rhinoplasty results.

Methods

This pilot study was approved by the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam, The Netherlands, documented by study number MEC-2015-238, and was conducted between June 2016 and January 2019. Informed consent for study participation was obtained from all volunteers. Because a large number of aesthetic rhinoplasty candidates in the Netherlands are Caucasian females, aged 25 to 40, we adhered to this specific population to test our hypotheses.

Eighty Caucasian female volunteers, who never had a (septo)rhinoplasty or extensive facial trauma, were recruited among hospital staff and patients visiting the otolaryngology outpatient clinic for complaints unrelated to the nose. For each subject, standardized frontal, oblique, and profile photographs of the face in a Frankfurt horizontal plane were taken by a medical photographer. Background and lighting were similar across subjects, hair was not allowed to cover any part of the face, and heavy make-up, glasses, or piercings were removed. For every face image, a cropped image of the nose in isolation was created as well.

Creating the Composite Average

The first subset of 40 photographic series was used to compute an averaged composite for each view (**Figure 1**). A set of 40 subjects is sufficient to create a stable composite that does not significantly alter when more subjects are added. Java Psychomorph version 6 was used to create the composite (Dr B. Tiddeman, Department of Computer Science, Aberystwyth University, UK; software publically available at http://users.aber.ac.uk/bpt/#software).[8] Standardized points were manually positioned on each face to delineate the position of facial feature landmarks of each individual image, and the average shape was calculated as the mean positions of corresponding delineated points across the face set. For each view, the images were reshaped to this average shape, and this set was then merged to a composite. A comprehensive description of this methodology, albeit for different purposes, is described elsewhere.[9]

Attractiveness Rating

A panel was organized to rate the attractiveness of the individual images of the first subset of 40 photographic series, the composite, and a second subset of 40 images. The second subset of 40 images was recruited to increase sample size, and to assess whether preference for the composite would also be evident in a comparable subpopulation that was different to the population from which the composite was created. For each subject, including the composites, panoramically oriented images of frontal, oblique, and profile views were created (**Figure 2**) and presented on an online interface. Each panel member

was asked to provide sex, age, ethnicity, level of education, and city of residence. Next, the panoramically oriented images were presented in random order in 2 blocks, comprising a set of 81 full faces (including the composite full face) and a set of 81 isolated noses (including the composite nose). Block order was counterbalanced across panel members. For both blocks, the panel was asked to rate the attractiveness of the nose specifically. Attractiveness ratings were assessed on a Likert scale ranging from 1 (very ugly) to 5 (very pretty).

Figure 1. Digitally averaged composite of 40 images of randomly selected Caucasian females aged 25 to 40 years (frontal view example).

Calculating Deviation from Average

We hypothesized that nasal shape deviation away from average correlates with lower attractiveness ratings. For each nose, including the composite nose, 18 landmarks were systematically placed to delineate the shape of the nose in the profile view (similar to the delineation process used to create the composite). To calculate shape deviation of an original nose from the composite, the *x* and *y* coordinates of the 18 landmarks for each of the 80 noses were subjected to principal components analysis. This procedure expresses the variation in landmark positions in a small number of components. In line with the Kaiser-Guttman criterion, only the components explaining the majority of variance in nose shape across the 80 faces were selected. The square root of the sum of the absolute deviations from average of these components can be used to create a "shape deviation score" for each nose. Subsequently the correlation between this deviation score and the attractiveness score was calculated.

Statistical Analysis

Statistical analyses were performed with IBM SPSS Statistics for Windows, version 22.0 (IBM Corp, Armonk, NY). An independent-sample t test was used to compare mean age between volunteer subsets and a 1-sample t test was used to determine

whether the composite ratings were statistically different from the sample mean. A Pearson correlation coefficient was calculated to indicate the correlation between shape deviation and attractiveness score.

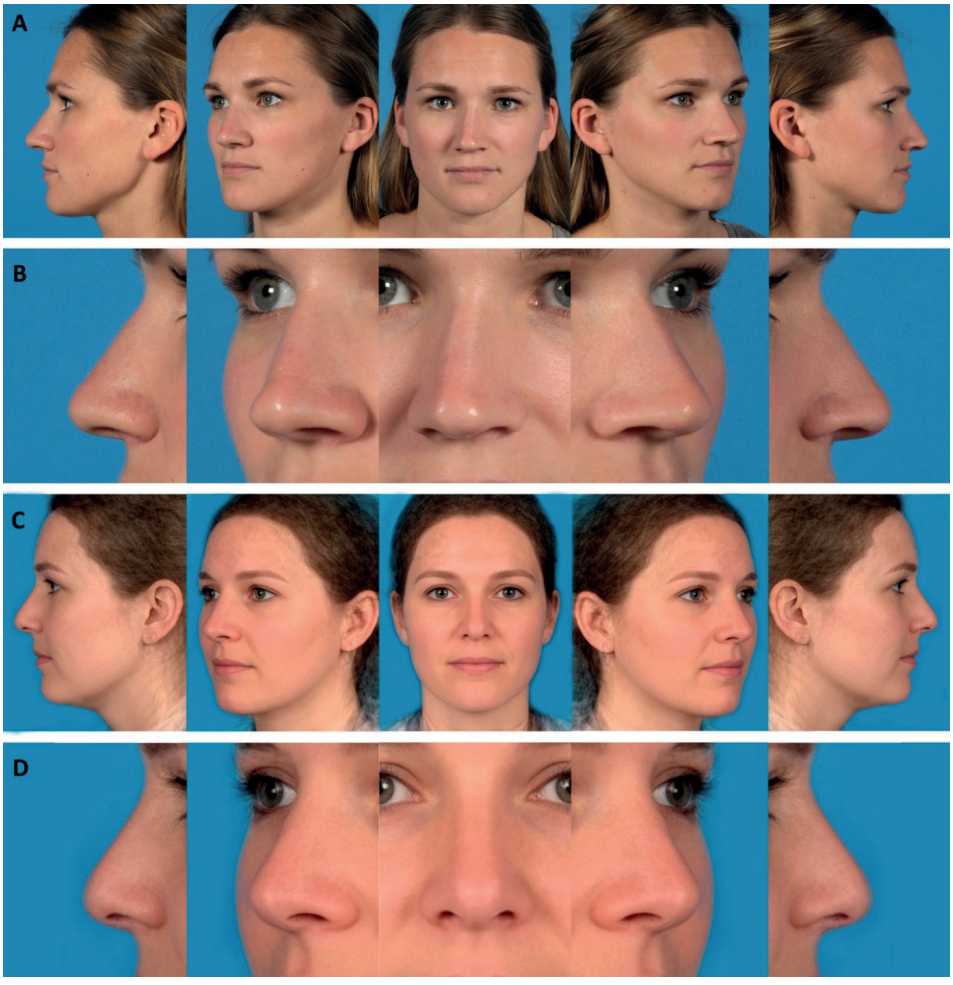


Figure 2. Examples of panoramically oriented images as presented to the panel of raters. Example of original full face of a 28-year-old female subject (A), example of original isolated nose of same subject (B), composite full face (C), and composite isolated nose (D).

Results

The mean age of the 80 Caucasian female volunteers was 28.8 [3.8] years (range, 25-40 years). There was no difference in mean age between the first and second subset of volunteers (29.1 vs 28.6 years, t = 0.52, *P* = 0.30). The attractiveness rating panel consisted of 41 volunteers from the general public. Twenty-one were male (51%), 20 were female (49%) and the mean age of the respondents was 35.3 [15.6] years (range, 18-64 years). Thirty were Caucasian (73%), 3 Asian (7%), and 8 of mixed origin (20%). The level of education was fairly high, with 27 volunteers possessing a bachelor's or master's degree (66%).

For each full face or isolated nose image, all raters' scores were averaged to produce a mean score. The mean score of the 81 full face images was 2.8 (range, 1.8-4.2). The nose of the composite face was rated highest, with a mean score of 4.2. The mean score of the 81 isolated nose images was 2.6 (range, 1.5-3.9). Again, the composite nose was rated highest with a score of 3.9. Both the nose in the composite face and the isolated composite nose were rated significantly higher than the distribution of ratings of the original images (t = 31.24 and t = 27.76, respectively; df = 79, $P < 0.001$). Hence the hypothesis that a digitally averaged composite nose of a specific population is rated as more attractive than the individual noses used to create it cannot be rejected.

Based on principal component analysis of the *x* and *y* coordinates of the nasal landmarks across the profile view of the 80 original noses, the first 4 components explaining 42%, 27%, 9%, and 6% (a total of 84%) of the variance in profile nose shape were used to calculate the difference of each nose from the averaged composite. The rating of the isolated nose correlated negatively with difference from average shape ($r = -0.40$, n = 80, $P = 0.0002$), implying that the larger the difference from average shape, the lower the aesthetic judgment of the nose (**Figure 3**). Similarly, the rating of the nose within the whole face correlated negatively with difference from average shape ($r = -0.36$, n = 80, $P = 0.00151$). Therefore, the hypothesis that deviation away from average correlates with decreased attractiveness cannot be rejected either.

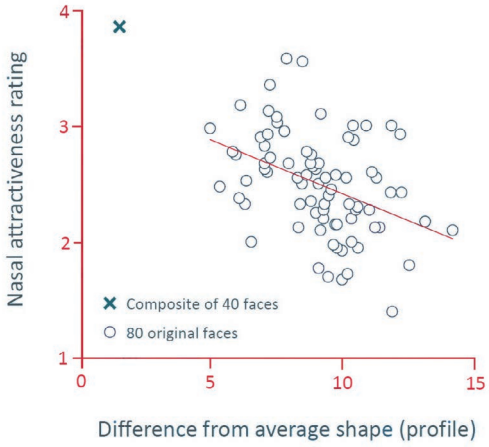


Figure 3. The rating of the nose across the faces (blue circles) correlated negatively with difference from average estimate (r = –0.40, n = 80, P = 0.0002); the larger the difference score, the lower the aesthetic judgment of the nose shape. The thin red line represents the best-fit regression line relating nose rating to the difference from average score. X denotes the rating of the composite nose from the blend of 40 faces. The shape difference of the composite nose is not zero because the composite was computed from the first set of 40 faces, not all 80 faces.

Discussion

This study shows that a composite nose, mathematically created from standardized photographs of 40 Caucasian females, is rated as more attractive than the individual noses used to create the composite. Therefore, for this specific population, an average nose is an attractive nose.

The composite nose in this study could serve as a surgical template for female Caucasians seeking aesthetic rhinoplasty. The template might aid in defining a surgeon's artistic sense of aesthetics next to the traditional rules and guidelines that are frequently used as the basis for nasofacial shape analysis and surgical planning. In this context we were interested in comparing the characteristics of the average composite "attractive" nose to a selection of important neoclassical canons and traditional aesthetic guidelines. We found some deviations (**Figure 4**): the evaluation of transverse facial proportions (rule of fifths, indicated by vertical black lines) revealed an alar base that is slightly wider than the inner canthal distance. The rule of thirds in profile view (horizontal black lines; the trichion-nasion, nasion-subnasale, and subnasale-menton distances should be equal) does not fit as the composite has a shorter nose and longer lower face, which is similar to the findings of the anthropologic studies of Farkas et al.[1] Nasal tip projection, calculated by the Goode method, and nasolabial angle are in line with common standards.[10,11] One must be aware that the composite found in this study only serves as a template for Caucasian female faces aged 25 to 40 years. Because the ideal nasal shape is likely not universal, the development of average composites to serve as prototypes for different rhinoplasty populations (e.g. Mediterranean, Asian, African-American, male) is necessary.[12]

Furthermore, we have shown that a principal component analysis of nasal landmarks in profile view makes it possible to calculate a shape deviation score. The more a nose deviated away from the average composite (high shape deviation score), the lower the attractiveness score of that nose was. The change of this shape deviation score following aesthetic rhinoplasty could have potential as an objective measure for the success of the procedure (**Figures 5 and 6**). This purpose would first necessitate further study, investigating whether shape deviation scores correlate with patient satisfaction (quantified via patient-reported outcome measures). Although patient satisfaction in the end remains the major determinant of the success of cosmetic surgery, the availability of an objective aesthetic rhinoplasty outcome measure could complement patient-reported outcome measurements by eliminating the impact of confounding factors that influence subjective ratings.

Figure 4. A selection of ratios, angles, and proportions of the composite nose. Horizontal black lines: rule of thirds. Vertical black lines: rule of fifths. Blue: nasolabial angle, 104°. Red: nasal tip projection according to Goode, 0.57. Yellow: nasofrontal angle, 139°.

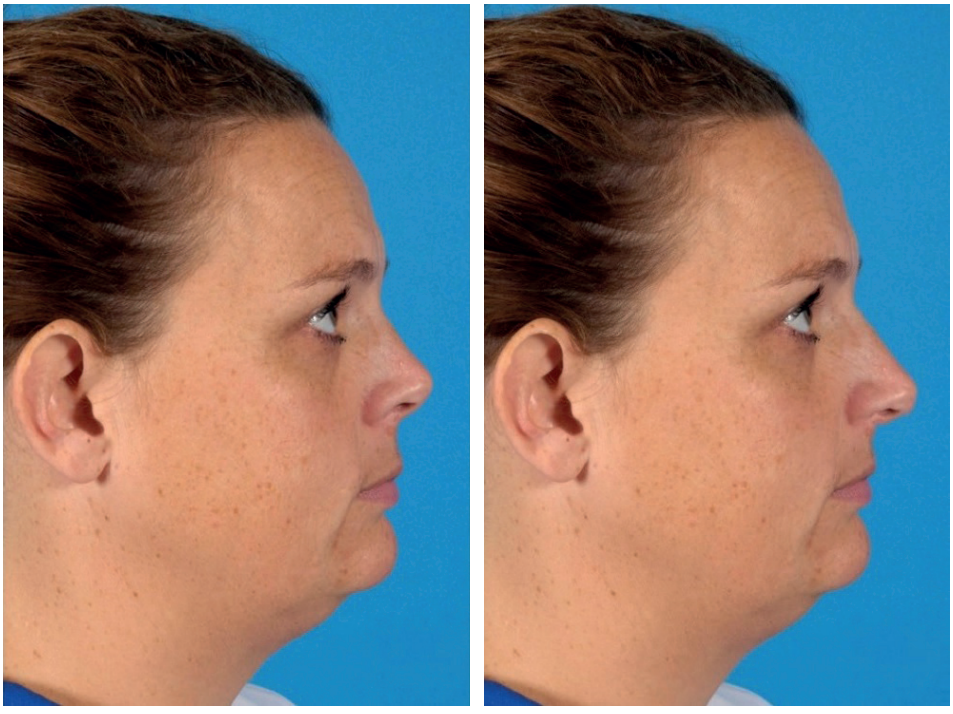


Figure 5. (A) Example of original nose (32-year-old female subject); (B) nose reshaped by adjusting the landmark position to that of the averaged composite.

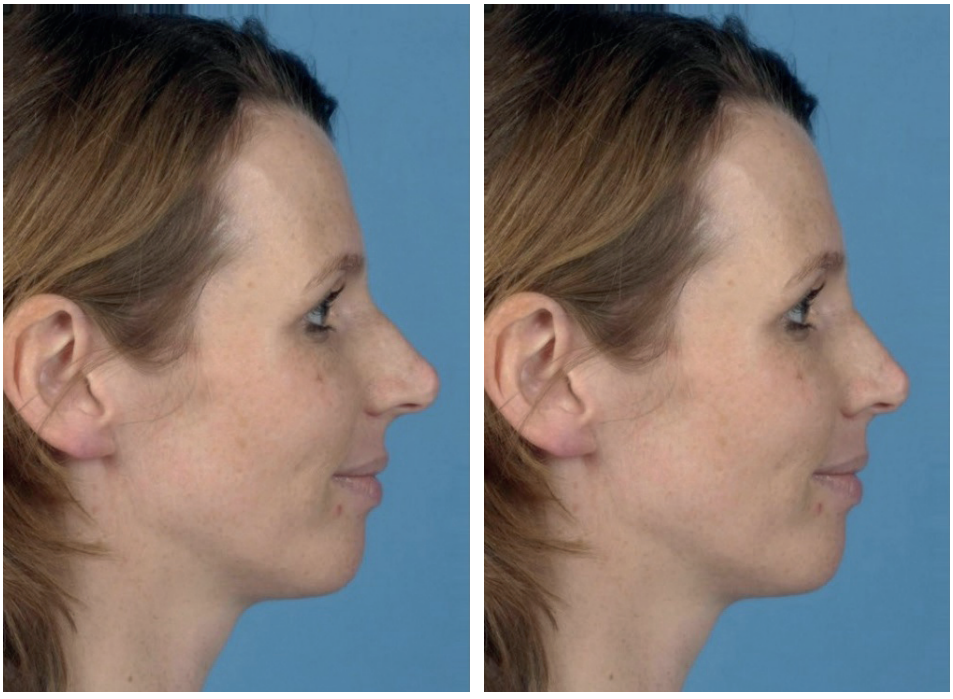


Figure 6. (A) Example of original nose (34-year-old female subject); (B) nose reshaped by adjusting the landmark position to that of the averaged composite.

This conceptual pilot study has several limitations. Creating a composite does not only result in a nose that is average in shape, but also introduces smoothed skin texture and more symmetry. Therefore, symmetry and skin tone could have been confounding factors that contributed to the preference of the composite over the original faces. However, these confounders do not affect the demonstrated correlation between shape deviation from average and attractiveness. We showed that within the original images (n = 80, excluding the composite), a deviation from average score purely based on nose shape metrics and not on symmetry or skin tone is inversely related to attractiveness. These images all have original skin tone and, in our opinion, there is no reason to assume that faces with a more averaged nose shape have fewer skin blemishes. Furthermore, the shape metric is calculated in profile view, which is independent of nasal symmetry. As such, the experiment performed supports the influence of averageness on attractiveness, independently of symmetry and skin tone. Additionally, previous research has shown that for the face as a whole the importance of averageness in the judgment of attractiveness is indeed quite robust: several authors have demonstrated that preference for averageness remains when corrected for symmetry or skin tone.[9,13-19] A second limitation is the need for manual positioning of the delineation landmarks. Despite our standardized and meticulous placement, there is a chance of a small

degree of inconsistency or error. There has been marked progress in the automated delineation of facial landmarks in the past 2 decades.[20-23] Although similar automation could be applied to the nose, automation reduces but does not eliminate all error. Third, in our panel of raters, the number of Caucasians was high, as was their level of education. Nevertheless, the volunteers with a lower level of education still preferred the average composite nose. Fourth, although average faces and noses are perceived as attractive, averageness is not a substitute for the aesthetic ideal. When creating a "high average composite" of the 25% most attractive faces, this prototype is rated as even more attractive.[24] This implies that averageness alone does not completely explain attractiveness.

Conclusions

This conceptual study shows that in Caucasian females, mathematically averaged noses are attractive. Furthermore, we were able to calculate a score expressing deviation from average nasal shape, which correlated negatively with attractiveness. Change in this nasal shape deviation score following rhinoplasty has potential as an objective aesthetic rhinoplasty outcome measure, although the correlation between averageness, attractiveness, and patient satisfaction needs further research.

References

1. Farkas LG, Hreczko TA, Kolar JC, Munro IR. Vertical and horizontal proportions of the face in young adult North American Caucasians: revision of neoclassical canons. *Plast Reconstr Surg.* 1985;75(3):328-338.
2. Farkas LG, Kolar JC. Anthropometrics and art in the aesthetics of women's faces. *Clin Plast Surg.* 1987;14(4):599-616.
3. Torsello F, Mirigliani L, D'Alessio R, Deli R. Do the neoclassical canons still describe the beauty of faces? An anthropometric study on 50 Caucasian models. *Prog Orthod.* 2010;11(1):13-19.
4. Bashour M. History and current concepts in the analysis of facial attractiveness. *Plast Reconstr Surg.* 2006;118(3):741-756.
5. Rubenstein AJ, Langlois JH, Roggman LA. What makes a face attractive and why: The role of averageness in defining facial beauty. In: Rhodes G, Zebrowitz LA, eds. *Advances in Visual Cognition-Facial Attractiveness: Evolutionary, Cognitive, Cultural, and Motivational Perspectives.* Westport, Conn.: Ablex; 2001:1-33.
6. Langlois JH, Roggman LA. Attractive Faces Are Only Average. *Psychol Sci.* 1990;1(2):115-121.
7. Halberstadt J, Rhodes G. It's not just average faces that are attractive: computer-manipulated averageness makes birds, fish, and automobiles attractive. *Psychon Bull Rev.* 2003;10(1):149-156.
8. Tiddeman B, Burt DM, Perrett D. Computer Graphics in Facial Perception Research. *IEEE Computer Graphics and Applications.* 2001;21(5):42-50.
9. Valentine T, Darling S, Donnelly M. Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces. *Psychon Bull Rev.* 2004;11(3):482-487.
10. Goode RL. A method of tip projection measurement. In: Powell N, Humphreys B, eds. *Proportions of the aesthetic face*: New York: Thieme-Stratton; 1984:15-39.
11. Boahene KDO, Orten SS, Hilger PA. Facial analysis of the rhinoplasty patient. In: Papel ID, Frodel JL, Holt GR, et al., eds. *Facial plastic and reconstructive surgery*. New York: Thieme; 3rd edition; 2009:477-487.
12. Broer PN, Buonocore S, Morillas A, et al. Nasal aesthetics: a cross-cultural analysis. *Plast Reconstr Surg.* 2012;130(6):843e-850e.
13. Rhodes G, Sumich A, Byatt G. Are average facial configurations attractive only because of their symmetry? *Psychol Sci.* 1999;10:52-58.
14. Rhodes G, Yoshikawa S, Clark A, Lee K, McKay R, Akamatsu S. Attractiveness of facial averageness and symmetry in non-western cultures: in search of biologically based standards of beauty. *Perception.* 2001;30(5):611-625.
15. Langlois JH, Roggman LA, Musselman L. What is average and what is not average about attractive faces? *Psychol Sci.* 1994;5:214.
16. Jones BC, DeBruine LM, Little AC. The role of symmetry in attraction to average faces. *Percept Psychophys.* 2007;69(8):1273-1277.
17. Rhodes G, Tremewan T. Averageness, exaggeration, and facial attractiveness. *Psychol Sci.* 1996;7:105-110.
18. Little AC, Hancock PJ. The role of masculinity and distinctiveness in judgments of human male facial attractiveness. *Br J Psychol.* 2002;93(Pt 4):451-464.
19. O'Toole AJ, Price T, Vetter T, Bartlett JC, Blanz V. 3D shape and 2D surface textures of human faces: The role of "averages" in attractiveness and age. *Image & Vision Computing.* 1999;18:9-19.

20. Haghighat M, Abdel-Mottaleb M, Alhalabi W. Fully automatic face normalization and single sample face recognition in unconstrained environments. *Expert Systems with Applications*. 2016;47:23-34.
21. Lanitis A, Taylor CJ, Cootes TF. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and machine intelligence*. 1997;19(7):743-756.
22. Fabian Benitez-Quiroz C, Srinivasan R, Martinez AM. Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016:5562-5570.
23. Zulgarnian Gilani S, Shafait F, Mian A. Shape-based automatic detection of a large number of 3D facial landmarks. *In Proceedings of the IEEE conference on computer vision and pattern recognition* 2015:4639-4648.
24. Perrett DI, May KA, Yoshikawa S. Facial shape and judgements of female attractiveness. *Nature*. 1994;368(6468):239-242.

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Chapter 8

General discussion

Rationale

"If you cannot measure it, you cannot improve it". This quote by Irish physicist and engineer William Thomson (1883) underpins the essence of this thesis. As a business manager, doctor or even a golf player it is impossible to determine whether you are successful or not if you fail to keep track of scores. Sound measures of performance are indispensable to quantify progress, identify underperformance and make the adjustments that are necessary to achieve the desired result. This thought was the primary reason to implement routine outcome measurements in our practice in 2014. Remarkably, the importance and relevance of this implementation was underlined one year later. The Dutch government labeled 2015 as the 'year of transparency', aiming to increase the availability of valid information on quality of delivered healthcare. In her letter, the public health minister stressed that insights in both costs and relevant outcomes are essential to improve healthcare quality and control rising expenditures.[1] This government call motivated us to further explore and improve outcome measurements in rhinoplasty and served as catalyst for the work presented in this thesis.

Effectuating transparency in delivered care is however not merely a Dutch initiative. Insights in what patient benefits are delivered against what costs is precisely the essence of value-based healthcare (VBHC).[2] Since the concept of VBHC was launched, many providers across the globe have embraced this new idea of organizing healthcare, resulting in numerous initiatives that intend to maximize patient-relevant outcomes. Our own center, the Erasmus Medical Center in Rotterdam, can be considered an early adopter as it has encouraged the start of various in-house pilot projects aiming to operationalize new VBHC strategies. Therefore, apart from the letter of our former public health minister, the evolution of VBHC worldwide and 'at home' was another trigger to initiate the work presented in this thesis.

Creating transparency and increasing value both start with measuring relevant outcomes. The aim of this thesis was to improve the process of measuring outcomes in rhinoplasty, explore the potential of measuring outcomes and tackle several barriers that prevent rhinoplasty surgeons from measuring outcomes on a day-to-day basis. We attempted to answer *what* outcome instruments to use by presenting an evidence-based recommendation on the best patient-reported outcome measure (PROM) to evaluate the effects of rhinoplasty. Second, a prospective rhinoplasty outcome routine is presented, demonstrating *how* to systematically measure outcomes. Third, we outlined the advantages of routine outcome measurements in both functional-aesthetic and secondary cleft-lip rhinoplasty patients, that emphasize *why* to measure outcomes in

rhinoplasty. This is complemented with a methodology that allows surgeons to use their own data to assess personal surgical performance and identify needs for improvement. The set of studies is concluded by an experimental study, investigating a method to objectively quantify aesthetic outcomes in rhinoplasty. The current chapter provides an interpretation of the main findings of the studies included in this thesis, discusses methodological and practical issues and raises suggestions for future research.

1. Selecting the best PROM to measure rhinoplasty outcome

1.1 Reflection on results

In **chapter 2,** we have used the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guideline to select the most appropriate PROM to measure functional and aesthetic outcomes in rhinoplasty. The COSMIN initiative, founded in 2005, consists of an international multidisciplinary team with expertise in the development and evaluation of outcome measurement instruments.[3] In 2018, an updated methodological guideline for systematic reviews of PROMs was developed.[4,5] This step-by-step manual focuses specifically on the methodology of systematic reviews of existing PROMs that are used for evaluative purposes (e.g. to measure effects of treatment). Using this guideline we have evaluated 12 different evaluative rhinoplasty PROMs to determine what instruments carry adequate validity, reliability and responsiveness. Based on such measurement properties, as well as on interpretability and feasibility aspects, we have narrowed these 12 different measurement instruments down to 1 functional instrument (the Nasal Obstruction Symptom Evaluation [NOSE] scale) and 2 aesthetic instruments (FACE-Q and Standardized Cosmesis and Health Nasal Outcomes Survey [SCHNOS] aesthetic subscale) that carry the most potential for further use.

The attentive reader of this thesis will have noted that the suggestions in **chapter 2** do not stroke with the aesthetic PROM used in the Rhinoplasty Healthcare Monitor (RHM) in **chapters 4, 5 and 6.** Instead of the FACE-Q or SCHNOS aesthetic subscale, we have utilized the Utrecht Questionnaire (UQ) in our own practice. This is primarily explained by the fact that the RHM was initiated in 2014. With the rhinoplasty validation of the FACE-Q performed in 2016 and the development of the SCHNOS in 2018, these PROMs were not available to us at the time.[6,7] The decision to use the UQ was based on available validation studies in rhinoplasty populations, its practicability, and the additional advantage of including 'trick' questions that aid in identifying patients with a possible body dysmorphic disorder.[8] In hindsight, using the elaborate methodology of the COSMIN, the validation studies of the UQ did not reach the same level of quality compared to the FACE-Q and SCHNOS in terms of structural validity and internal consistency. However, as mentioned in **chapter 2**, this does not imply that the quality of the UQ itself is inadequate. With high-quality validation studies, the quality of the UQ can still be established. The findings of the systematic review pose a challenging dilemma regarding the choice of instruments in the RHM. Generally speaking there are two options; we could decide to further validate the UQ using high-quality validation studies in order to justify its use, and continue using the UQ in the RHM. However, this will sustain heterogeneity in outcome instruments used in rhinoplasty, as other rhinoplasty

surgeons or practices might decide to continue or start using the SCHNOS or FACE-Q. The other option is to replace the UQ with the SCHNOS or FACE-Q in the RHM. A major disadvantage of this is that this will 'reset' the dataset, rendering the acquired UQ data over the past years useless as it cannot be compared to scores of the new instrument. If, in the future, with more data on measurement properties, consensus is reached that either the SCHNOS or FACE-Q should be used in outcome reporting in rhinoplasty, this switch is inevitable. Yet, a third solution may be contemplated, in which different instruments of adequate quality continue to be used but *with* the possibility of inter-instrument comparison. We will elaborate on this theory in chapter 1.4.

1.2 Lack of content validity

In order to label an instrument as having 'adequate quality', the most important measurement property that must have been established is content validity. Content validity refers to the extent to which the items of the instrument adequately reflect the construct that is measured, i.e. that the instrument indeed measures what it claims to measure. Content validity is considered the most important measurement property, and lack of content validity can affect all other measurement properties.[9] Therefore, when making an evidence-based decision on the best available PROM, information on content validity is essential. Content validity can be ensured either by the quality of the development study or by separate content validity studies. However, none of the development studies of the identified PROMs in **chapter 2** were of sufficient quality and no separate content validity studies were found. In the development study of the NOSE, a quantitative instead of qualitative data collection method was used to identify relevant items, analysis of data in the concept elicitation phase was doubtful, and cognitive interview studies were not performed.[10] The development studies of the FACE-Q and SCHNOS were also considered inadequate by the COSMIN standards. Furthermore, no other content validity studies for these PROMs are available. Consequently, high-quality content validity studies of the identified PROMs are necessary. Such studies should encompass interview or focus group studies with patients and professionals, using qualitative methods to obtain data on the relevance, comprehensiveness and comprehensibility of items.

On the other hand, we must take into consideration the stringency of the COSMIN standards. The development of the FACE-Q for example is elaborate, including a modern psychometric approach and cognitive interview studies with patients. According to the COSMIN standards, however, the development study is rated inadequate because during concept elicitation, qualitative data was coded by one member of the research team instead of independent coding. This leaves the FACE-Q development study with an inadequate score while compared to other available aesthetic rhinoplasty PROMs, the design has been elaborate, thorough and well-documented.

1.3 Towards standardization in rhinoplasty outcome measurements

To counteract heterogeneity in outcome reporting and simultaneously ensure that all relevant outcomes are captured, it is imperative to establish national and international consensus on what outcomes to measure and how to measure them. Although we have focused on determining how to measure a specific outcome in **chapter 2**, i.e. choosing the best instrument, agreement on what outcomes to measure in rhinoplasty patients is of equal importance. We focused on the outcome domain 'symptoms', defined as symptoms related to nasal obstruction or concerns with nasal appearance, because these are considered the most important determinants of the functional and aesthetic success of rhinoplasty.[11-13] However, these are likely not the only outcomes that are relevant to rhinoplasty patients. Examples of other possible outcome domains that may be relevant are health-related quality of life (physical, emotional and social), psychological functioning or adverse events.

Progress has recently been made in various healthcare areas on the development of core outcome sets (COS). A COS is a standardized set of outcomes that ought to be measured and reported as a minimum in effectiveness studies. This set is agreed upon by all relevant stakeholders (patients, physicians, policy makers, outcome experts), ensuring that the included outcomes are relevant and important. The availability of a COS counteracts inconsistency in outcome measurement and facilitates synthesis of published outcome data from different studies (meta-analysis). Furthermore, it reduces outcome-reporting bias, which occurs when only a subset of the original recorded outcomes is presented in the article based on statistically significant results.[14] Guidelines to develop a COS are provided by initiatives such as COMET (Core Outcome Measures in Effectiveness Trials).[15] To date, a COS has not yet been established for rhinoplasty. Another initiative that guides the development of standardized sets of outcomes is the International Consortium for Health Outcomes Measurement (ICHOM).[16] Using a methodology comparable to COMET, including consensus meetings with various stakeholders and systematic evaluations of available evidence, ICHOM focuses on standardized outcome reporting in clinical practice. The aim lies on developing a globally applicable and freely available set of outcomes that matter to patients, encouraging adoption and implementation of value-based healthcare in daily practice. The standard sets provide a framework of what outcomes to measure, how to measure them and when to measure them throughout the care process. Numerous standard sets have been published covering different conditions and populations, including coronary artery disease, diabetes, Parkinson's disease and cleft lip and palate. ICHOM initiates the development of a standard set itself, and appoints a working group once funding for the development has been acquired. No standard set has yet been developed for rhinoplasty.

1.4 An alternative route to standardization in rhinoplasty outcome measurements?
One of the challenges of developing a COS in rhinoplasty could be a political one. Most rhinoplasty PROMs have been developed by large research groups with significant support from geographically or socially affiliated surgeons and practices. It will be difficult to persuade such groups to abandon the PROM they developed themselves, which has often been a time-consuming process, and convert to using another. This is particularly relevant if future validation studies establish good measurement properties including good content validity for multiple PROMs, and the PROM suggested by the COS is not intrinsically better than the one already used. Given that many functional and aesthetic PROMs are rather similar to each other, with only slightly different structure and wording, this is a realistic possibility. This would create the problem of multiple PROMs being utilized with each of adequate quality, but given the different (sub)score systems impossible to compare. Second, as we have pointed out earlier with our use of the UQ in the RHM, surgeons who have established prospective registries using a particular instrument might be reluctant to switch to a different instrument. If historical PROM data cannot be bridged with a new PROM, this leads to a loss of valuable data. We consider this a substantial drawback of a global COS rollout in which one PROM per outcome is advocated.

A solution to this problem could be to develop so-called crosswalks. Crosswalks allow scores from one instrument to be converted to another instrument's score. Each PROM (sum)score has a scale, and one cannot simply assume that different PROMs are scaled equally. It is not clear, for instance, whether a score of 50 out of 100 on the NOSE is equivalent to a score of 10 out of 20 on the SCHNOS functional subscale. Thus, to compare scores between different PROMs, bidirectional crosswalks have to be created and validated using data of both PROMs that are completed by the same participants. Examples of methods to assess equivalence are the equipercentile equating method or Rasch analysis-based equating.[17,18] The advantage of such conversion tables is that it facilitates comparison and pooling of data acquired using different PROMs. This means that surgeons can continue using their PROM of choice, as long as it has evidence of adequate quality, while preserving the value of previously collected data. In our case, if future studies establish adequate content validity of the UQ, that implies that we can continue using the UQ and use validated crosswalks to compare our results to practices using the SCHNOS or FACE-Q. We are currently exploring the possibility of creating such crosswalks.

2. The Rhinoplasty Healthcare Monitor

The core of this thesis focuses on practical aspects of outcome measurements in rhinoplasty. **Chapters 4, 5 and 6** concern the implementation and evaluation of a standardized and prospective rhinoplasty outcome routine that includes simple and relevant questionnaires. Results from this routine are monitored using a web-based dashboard that provides real-time automated data analyses with the click of a button. The RHM demonstrates that with limited effort outcome data can be gathered, monitored and analyzed in a transparent fashion, resulting in significant benefits on a patient- and cohort level.

2.1 Optimizing the rhinoplasty consultation

Utilizing the RHM provides several advantages to both the patient and the surgeon during consultation. Since questionnaires are filled out in the waiting room, both patients and surgeon can prepare for the consultation. Patients are encouraged to think about their complaints beforehand, e.g. the impact of nasal appearance on daily functioning or laterality of nasal obstruction, promoting an efficient anamnesis. To the surgeon, beforehand insights into complaints support relevant history-taking and physical examination. In case of postoperative consultations, the availability of PROM data prior to patient contact avoids unexpected discordance regarding the result. A surgeon can anticipate on an encounter with a dissatisfied patient by analyzing the operative report and preparing possible solutions beforehand. Conversely, a result that may appear suboptimal to the surgeon but PROM data indicating the patient is not the least concerned about it, refrain the surgeon from unnecessarily discussing revision surgery. Furthermore, the RHM is valuable in patient counseling and expectation management. The visualization of average expected functional and aesthetic improvement aids in creating realistic expectations and supports the shared decision-making process. Using the dashboard, pending questions (e.g. how much gain can I expect on average, does my previous surgery influence the expected postoperative result, do surgical results last over time, is it likely that my insurance company will cover expenses) are replied with data-driven answers. Hence, the RHM supports informed clinical decision-making that is based on patient preferences and severity of disease. It is important to remark that this advantage of the RHM has been based on our own experience and direct feedback from patients. To confirm that the RHM improves patient counseling and shared decision making we would have to test this empirically. Future qualitative research using structured interviews among patient users is therefore necessary to verify these benefits.

2.2 Learning from personal performance as a surgeon

As demonstrated in **Chapter 6,** the RHM helps to visualize and evaluate personal surgical performance. The RHM contains two features that allow the surgeon to detect trends in performance and identify learning needs. One of these features is the depiction of a surrogate learning curve. By measuring surgical performance (patient outcomes, Y-axis) as a function of time (X-axis), a surgical learning curve can be plotted.[19] Understanding the development of the individual learning curve and visualizing its shape helps to identify areas for improvement and can provide motivation to increase the gradient of the learning process. We hypothesize that the rhinoplasty learning curve is unique for each rhinoplasty surgeon and that its course is sculpted by interpersonal features (e.g. dedication, surgical talent), surgical exposure, training (e.g. fellowship, courses, writing articles on rhinoplasty), and retroflection. Increasing focus on rhinoplasty, observing experts, attending courses or reflecting on your own results may all accelerate the learning curve, whereas factors such as reduced focus on rhinoplasty, ageing or health issues may have a negative impact on surgical outcomes.[20] Using the learning curve, it would be interesting to visualize the effect of potential accelerators on surgical development. For the dedicated surgeon, the challenge is to increase the gradient of the individual learning curve in order to achieve superior results in a shorter period of time.

One may argue whether PROMs represent an adequate reflection of surgical performance. In the literature, two variables are distinguished to measure surgical performance: measures of the clinical process and task efficiency, or measures of patient outcome.[19,21] Whether process measures (e.g. time to complete an operation or blood loss) are valid indicators of surgical performance is under debate, since these measures are not necessarily related to proficiency.[22] Furthermore, rhinoplasty as a procedure can be substantively variable, making it difficult to evaluate process measures of one specific surgical procedure. The other option to measure performance is using measures of patient outcome. Traditional patient outcomes such as the incidence of intraoperative complications or survival may be used to measure surgical performance, but are inadequate indices to measure performance in rhinoplasty due to the rarity of these events. In rhinoplasty it is the patient, and the patient only, who dictates success. Hence, we theorize that PROMs are the most relevant parameter of performance.

2.3 Benchmarking surgeons using the RHM

Standardized data collection facilitates comparisons of rhinoplasty results among surgeons, practices or countries. Benchmarking providers could contribute to improving the quality of rhinoplasty care by establishing a (inter)national average, identifying potential outliers and investigating differences in practice. This possibly

creates new and improved standards of care. Interestingly, controversy exists regarding benchmarking results between doctors. Openly comparing performance may by some be anticipated to be something frightening, and prone to creating a 'naming and shaming' culture. However, we believe outcome transparency between doctors should be embraced. Openly comparing results can create a learning environment in which the goal is to collectively improve, by analyzing differences in outcomes. Instead of blaming underperformers, the goal is to learn from high performers. Comparisons of day-to-day performance may provide a feasible and pragmatic method to identify certain surgical techniques or perioperative care regimens that are associated with superior outcomes. A Dutch case study of VBHC implementation compared outcomes between breast cancer surgeons in different hospitals, and found significant differences in the number of reoperations (an outcome that was labeled important by patients) due to postoperative bleedings.[23] After an open discussion between the concerned surgeons, this variance was thought to be explained by prolonged intraoperative wound irrigation: this was a standard-of-care for the surgeon with the lowest postoperative hemorrhage rate. Adoption of this strategy by the other surgeons led to a considerable decrease of reoperations among most of the collaborated hospitals. This is perhaps a rather idealistic example, but it does illustrate the potential strength of benchmarking.

Of note is that in order to facilitate accurate benchmarking, it must be ensured that surgeons are indeed comparable. Apart from collecting outcomes, this requires the collection of patient characteristics and conditions, also known as case-mix variables or risk-adjustment variables. Collecting relevant risk-adjustment variables is crucial to ensure valid comparisons of outcomes between providers that serve different rhinoplasty populations. A rhinoplasty practice mainly operating on revision cases, for example, is expected to attain lower postoperative outcome scores compared to a practice exclusively performing primary rhinoplasties. Similarly, whether the intention of the surgery is functional or aesthetic is a very relevant risk-adjustment variable. In its current state the RHM keeps record of age, sex, primary or revision surgery, and whether the focus was functional, cosmetic or both. Hypothetically, additional potentially relevant risk-adjustment variables may be race/ethnicity, history of severe nasal trauma, educational level (as indirect measure of social economic status) and whether expenses for the procedure were covered or paid out-of-pocket.

2.4 Should surgeon performance parameters be publicly available?

Apart from exposing personal performance measures to colleagues, an even more controversial issue is whether surgical outcomes should be available to patients. In the Netherlands, the free choice of physician that patients enjoy is highly valuated, but at the same time, there is not much to choose from. Little to no information about

physician performance is currently available to patients. Although standard clinical quality indicators collected by hospitals, such as readmissions or surgical site infections, are applicable to rhinoplasty, the incidence of such events is low and therefore these indicators lack the ability to discriminate between functional rhinoplasty providers. To decide where and from whom to receive a cosmetic rhinoplasty, patients rely on examples of successful cases provided by the surgeon's website. Although these may certainly be of value to showcase a sense of style of the surgeon, they are by no means a true reflection of overall performance. In striving for transparency in healthcare, one particular initiative that is endorsed by the ministry of health is Zorgkaart Nederland, an online review platform that allows patients to assess and rate registered physicians.[24] The reviews and scores on this website are publicly available. Although this source provides some idea of treatment outcomes in rhinoplasty (i.e. patient satisfaction), it is primarily a patient-reported experience measure and not an outcome measure. More importantly, the reliability of this platform is questionable: patients tend to rate in extremes resulting in skewed scores and enterprising physicians have been accused of using this platform for marketing purposes, pumping up scores by selectively recruiting patients.[25]

Given the consecutive nature of patient enrollment in the RHM, the data generated by the RHM reflects performance of an unselected cohort, including both satisfied and dissatisfied patients. As opposed to the selective nature of patient reviews on Zorgkaart Nederland, the cohort-level outcomes provided by the RHM offer patients a fair, meaningful and transparent indicator of rhinoplasty results. In the commercial market of aesthetic rhinoplasty, publicizing provider-specific outcomes may create competition among surgeons to strive for high patient outcomes and could reveal malpractices. Furthermore, insights in outcomes may justify differences in pricing between clinics, offering patients the option of making an educated choice based on price/performance ratio. On the other hand, implementing reliable and valid public performance measures is challenging. Not only should the acquired outcomes be relevant and valid, data should also be risk-adjusted in order to prevent surgeons from manipulating scores by avoiding difficult or revision cases. Second, one may argue what timepoint in the follow-up dictates an accurate performance score. The RHM measures outcomes at 3 months and 12 months after surgery, but long-term sequelae (e.g. alar retraction that may reveal itself after several years) exist and should ideally be incorporated in the performance score. Third, as a public quality ranking may have far-stretched financial consequences, especially in the commercial aesthetic market, public performance measures are potentially vulnerable to fraud. All public information should be fully verified, preferably by an independent centralized facilitator, to ensure accuracy and reliability.

2.5 Improving evidence levels in rhinoplasty

Perhaps the most significant advantage of initiatives such as the RHM is the potential to elevate the level of evidence in rhinoplasty. Evidence levels rank the methodological quality and relative strength of results obtained from scientific studies. The highest level of evidence is level I, representing high-quality randomized controlled trials (RCTs); the lowest level of evidence is level V, representing expert opinion (**Table 1**).[26] In facial plastic surgery, including rhinoplasty, evidence levels are relatively low compared to other healthcare areas. Most studies are anecdotal, based on expert opinion or retrospective case series; level I or level II studies are exceedingly rare.[27] As scientific evidence is becoming tied to reimbursements, the absence of high-level evidence can have far stretched consequences. In the Netherlands, for example, health insurance carriers are collectively starting to decline expense coverages for functional rhinoplasty as there are no randomized controlled trials (RCTs) or meta-analyses to support the effectiveness of nasal valve surgery.[28] As a consequence, an increasing percentage of Dutch patients with moderate to severe nasal obstruction are withheld from treatment; treatment that every experienced rhinoplasty surgeon will label as effective.

Table 1. Evidence levels.

Level of evidence	Qualifying Studies
I	High-quality, multicenter or single-center, randomized controlled trial with adequate power; or systematic review of these studies
II	Lesser-quality, randomized controlled trial; prospective cohort or comparative study; or systematic review of these studies
III	Retrospective cohort or comparative study; case-control study; or systematic review of these studies
IV	Case series with pre-/post test or only post test
V	Expert opinion developed via consensus process; case report or clinical example; or evidence based on physiology, bench research, or "first principles"

As the impact of bias is minimized by strictly controlling the study population, setting and methods, RCTs are considered the gold standard to study the effect of an intervention.[29] However, attaining such evidence in surgical interventions poses significant challenges. Apart from obstacles arising from sample size recruitment, blinding, ethical considerations and financial barriers, most surgical interventions are subject to substantial procedural variability. Due to a tremendous heterogeneity of surgeon-specific factors (e.g. preference for techniques, experience, learning curve) and patient factors (varying individual anatomy and pathology), this is particularly evident in rhinoplasty. Primarily, it is nearly impossible to conduct a trial in which such

surgeon and patient factors are completely homogeneous, a necessary requirement to reliably study an isolated effect of a specific rhinoplasty technique. Second, the procedural variability of rhinoplasty could well imply that a technique studied in a well-controlled experimental setting in center A, may behave significantly different when executed in center B. Therefore, even if RCTs were logistically, technically and financially feasible, generalizability of trial results is questionable.

Consequently, rhinoplasty surgeons must consider different routes to raise the level of evidence. High-level evidence that is relatively feasible to acquire could arise from large outcome registries. Using big data analyses from prospectively acquired outcome data is relatively affordable and provides level 2c evidence; evidence that is currently scarce in facial plastic surgery. Although causality will remain difficult and bias is never eliminated, large datasets could produce solid prediction models that reliably indicate what effect may be expected from rhinoplasty following specific anatomical, procedural and patient characteristics. Such registries combining data on facial plastic procedures are already available, but these usually lack the most important parameter of effectiveness: patient-reported outcomes.[30] Initiatives such as the RHM can facilitate standardized and multi-center acquisition of PROM data, producing large sample sizes that lead to valuable (inter)national datasets. Such datasets may not only help to improve surgical decision-making and establish clinical practice guidelines, but also provide a data-driven reply to third-party skepticism.[31]

2.6 Future improvement of the RHM

Despite all good intentions, the past years have learned us that virtually all attempts to improve quality in healthcare have come at the cost of increased administrative burden. Clinical documentation, procedural checklists and quality accreditations have all emerged as 'guards' of quality in healthcare, but have also accumulated to the number one frustration of healthcare professionals nowadays.[32] Physicians worldwide have reported that administrative duties negatively affect the ability to deliver high-quality care, result in lower levels of career satisfaction and contribute to the incidence of burnouts.[33-35] As much as we have tried to design the RHM to fit a busy practice with minimal administrative load, the design currently dictates that users manually enter PROM scores into their practice' database. This takes time and especially for non-academic surgeons, that additional time could be a significant barrier to implementing routine outcome measurements in practice. Therefore, one of the most important improvements that we intend to achieve in the near future is a fully digital, automated PROM administration and acquisition method. Early 2021 we have replaced the paper-based PROM collection with an electronic PROM system that allows patients to either fill in the questionnaires at home or in the waiting room using a tablet computer. This

contributes to automated data acquisition and is also in line with a recent qualitative study that suggests that patients prefer electronic platforms over paper.[36] Still, the current electronic PROM administration produces a raw data file that mandates manual input into the RHM database. Apart from creating administrative load, manually entering PROM scores into a database also potentially introduces data entry errors.[37] We are therefore working on a method to automatically transfer the acquired raw data into the RHM database, for which a few technological hurdles will have to be cleared.

3. The role of averageness in aesthetic rhinoplasty outcome

3.1 Reflection on results

Patient-reported outcome measures cover a wide range of outcomes relevant to rhinoplasty patients, including symptom severity, symptom burden, satisfaction, and health-related quality of life. Since these domains necessitate the patient perspective, PROMs take a vital and indispensable position in the measurement of rhinoplasty outcomes. Still, PROMs have its disadvantages. Given the subjective nature of responses, PROMs inherently introduce bias. PROM answers may be affected by factors such as mood, expectations, interaction with the healthcare provider and socioeconomic situation.[38] For rhinoplasty specifically, rhinoplasty patients may be concerned that the severity of symptoms affects eligibility for surgery and adjust their responses accordingly. In several countries such as the Netherlands and the UK, this bias is further fueled by the role of the insurance carrier or commissioner: the decision to cover the procedure is made on an individual basis in which the severity of functional disturbances plays an important role.[39] Although currently few patients appear to be aware of coverage criteria, this potentially inflates baseline PROM scores. Furthermore, PROMs may also overestimate treatment effects. A large meta-analysis of over 1000 trials found that when patients are aware of their allocation, patients allocated to treatment arms tend to be overly optimistic in reporting outcomes, whereas control patients tend to undervalue outcomes.[40] Moreover, it has been suggested that in cosmetic surgery, which is elective in nature, patients overvalue postoperative outcomes in order to justify their decision to undergo the procedure.[41] On the other side of the spectrum, facial plastic surgeons frequently encounter patients being overly critical about relatively small disturbances. Consequently, arguments to include objective measures in core outcome sets for rhinoplasty exist.

We have attempted to produce a basis for an objective measure of aesthetic rhinoplasty in **chapter 7**. Using averageness of nasal shape, we identified an objectively measurable trait of attractive noses. The more an individual Caucasian female nose resembles the average shape of Caucasian female noses, the more attractive it is perceived. We performed a mathematical analysis of nasal landmarks, producing a numerical value that indicates how much an individual nose deviates from average. This deviation score could potentially serve as an objective measure of aesthetic rhinoplasty: the more a nose approaches the average, the more successful the rhinoplasty. To our knowledge, no other authors have attempted to quantify nasal shape altogether as a metric that correlates with a favorable outcome. Several authors managed to objectively measure changes after rhinoplasty using two- or three-dimensional surface analysis, but did not superimpose these objective changes onto an outcome scale.[42-45] Szychta et al. objectively measured symmetry of the nose using a three-

dimensional scanner, and reported an increase in symmetry after posttraumatic rhinoplasty.[46] Although symmetry is important, its role in determining the success of a rhinoplasty is limited especially given its irrelevance in profile analysis. Prior to the publication of **chapter 7**, no objective aesthetic outcome instruments assessing the nose as a whole have been identified.[47]

The scarce literature indicates that creating a valid and reliable objective outcome measure is challenging, and consequently several remarks can be made regarding our findings. Whether averageness is a complete substitute for either attractiveness or the success of aesthetic rhinoplasty is questionable. Primarily, attractiveness has proven to be a difficult concept to capture objectively. Evolutionary and social psychologists have studied attractiveness for decades, but the concept still eludes a satisfactory understanding.[48] Although averageness has been shown to be a robust factor, it is likely not the only contributor to attractiveness.[49] Similarly, whether the degree of averageness is a valid parameter to dictate the aesthetic success of an individual rhinoplasty is debatable. Aesthetic rhinoplasty patients may have particular individual desires that substantially differ from average nasal shape. Some patients may have specific requests in line with their personal sense of beauty, others may desire a nose that deliberately diverts away from their ethnical origin. A rhinoplasty result that the patient considers successful is therefore not necessarily equal to an improvement in averageness deviation score. One could argue that the objective measure of averageness is of particular value on a cohort-level assessment of outcome, and less suitable in determining outcomes in individual cases. On the other hand, it is postulated that rhinoplasty candidates mainly desire a less conspicuous nose.[50] Instead of an operated, stylized look, patients predominantly prefer a nose that restores facial harmony in which not the nose, but aesthetically more pleasing features such as the eyes draw attention.[20] Based on that theory, the degree of averageness may well be a valid indicator of success as averageness irrevocably reduces conspicuities. Furthermore, many surgeons agree that the majority of non-Caucasian rhinoplasty candidates wish to maintain rather than eliminate their ethnical traits, suggesting that an averaged nose of corresponding ethnicity sets an adequate surgical goal.[51-53] We therefore assume that for the majority of patients, averaged nasal shape is indeed indicative of successful outcome. Naturally, future research is needed to substantiate this assumption.

3.2 Directions for future research

Given the experimental nature of the study presented in **chapter 7**, we propose several areas for further study before qualifying the averageness deviation score as an objective outcome measure for aesthetic rhinoplasty. Primarily, the association between nasal averageness and attractiveness will need further investigation, as we

are the first to report on this relation. A future study could involve changing the original shape of study participant's noses into average shape, as we demonstrate in **Figure 1**, and assess whether this leads to an increase in attractiveness. Further, the value of the averageness deviation score will have to be investigated in rhinoplasty patients. Studying the correlation between change in averageness deviation score and change in PROM scores after rhinoplasty is the most obvious method to establish validity for this objective outcome measure. If indeed the averageness deviation score proves to be reliable and valid to assess aesthetic outcome in Caucasian females, averaged prototypes for patients of different gender and ethnical origin may be developed and tested.

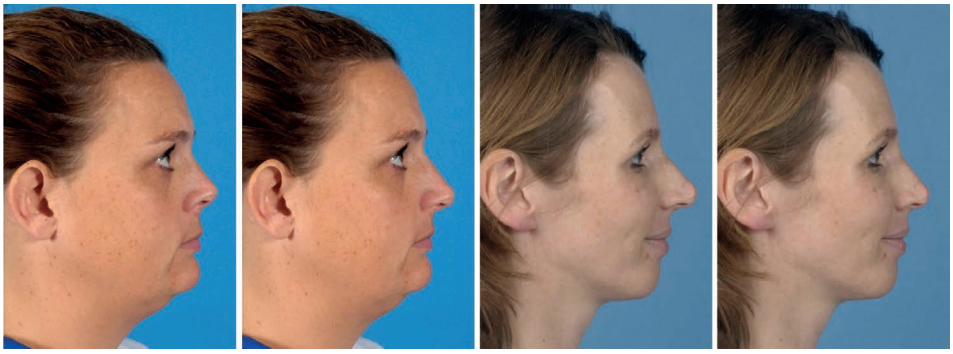


Figure 1. Two examples of the original nose (*left,* female, age 32; *center right,* female, age 34) reshaped by adjusting the landmark position to that of the averaged composite (*center left and right, respectively).*

4. Concluding remarks

As outcome information is increasingly considered essential in medical decision-making, outcome-based approaches are experiencing a boost in healthcare. In the realm of VBHC, the patient perspective has taken a central role in defining relevant outcomes. With this thesis, we have attempted to adapt to this paradigm change and capture the patient perspective in rhinoplasty. This thesis has demonstrated that routine outcome measurements can improve preparation of consultation, patient selection, patient empowerment and management of expectations, and visualize personal surgical performance. Above all, we hope that these advantages encourage other rhinoplasty surgeons to participate in routinely measuring outcomes. As new opportunities arise from digitization and information technologies, platforms such as the RHM potentially provide infrastructures to collect study data on large scales and combine data to build big data registries. To achieve this, finding a common ground for routine outcome measurements is key. This thesis makes an appeal to join forces and collectively invest time and resources in enhancing routine outcome measurements in rhinoplasty. By collaborating we can advance from expert opinion to cohort-level evidence and provide data-driven responses to third parties, our patients and ourselves.

References

1. Schippers EI. Ministry of Health Welfare and Sport. Letter to House of Representatives: '2015 The year of transparency'. November 10, 2014. *(letter in Dutch)*. https://www.rijksoverheid.nl/documenten/kamerstukken/2015/03/02/kamerbrief-het-jaar-van-de-transparantie. *Accessed September 8, 2020.*
2. Porter ME. What is value in health care? *N Engl J Med.* 2010;363(26):2477-2481.
3. Amsterdam Public Health. COSMIN helps you select the most suitable outcome measurement instrument. Available at: https://www.cosmin.nl. *Accessed March 24, 2020.*
4. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27(5):1147-1157.
5. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res.* 2018;27(5):1171-1179.
6. Klassen AF, Cano SJ, East CA, et al. Development and Psychometric Evaluation of the FACE-Q Scales for Patients Undergoing Rhinoplasty. *JAMA Facial Plast Surg.* 2016;18(1):27-35.
7. Moubayed SP, Ioannidis JPA, Saltychev M, Most SP. The 10-Item Standardized Cosmesis and Health Nasal Outcomes Survey (SCHNOS) for Functional and Cosmetic Rhinoplasty. *JAMA Facial Plast Surg.* 2017.
8. Lohuis PJ, Hakim S, Duivesteijn W, Knobbe A, Tasman AJ. Benefits of a short, practical questionnaire to measure subjective perception of nasal appearance after aesthetic rhinoplasty. *Plast Reconstr Surg.* 2013;132(6):913e-923e.
9. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res.* 2018;27(5):1159-1170.
10. 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):157-163.
11. Rhee JS, Weaver EM, Park SS, et al. Clinical consensus statement: Diagnosis and management of nasal valve compromise. *Otolaryngol Head Neck Surg.* 2010;143(1):48-59.
12. Cannon DE, Rhee JS. Evidence-based practice: functional rhinoplasty. *Otolaryngol Clin North Am.* 2012;45(5):1033-1043.
13. Ching S, Thoma A, McCabe RE, Antony MM. Measuring outcomes in aesthetic surgery: A comprehensive review of the literature. *Plastic and reconstructive surgery.* 2003;111(1):469-480.
14. Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials.* 2012;13:132.
15. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. *Trials.* 2017;18(Suppl 3):280.
16. International Consortium for Health Outcomes Measurement. Available at: https://www.ichom.org/. *Accessed April 4, 2020.*
17. Vos L, Whiteneck GG, Ngan E, Leon-Novelo L, Sherer M. Rasch Analysis of Postconcussive Symptoms: Development of Crosswalks and the Brain Injury Symptom Scale. *Arch Phys Med Rehabil.* 2019;100(10):1844-1852.
18. Polascik BA, Hidaka C, Thompson MC, et al. Crosswalks Between Knee and Hip Arthroplasty Short Forms: HOOS/KOOS JR and Oxford. *J Bone Joint Surg Am.* 2020;102(11):983-990.
19. Hopper AN, Jamison MH, Lewis WG. Learning curves in surgical practice. *Postgrad Med J.* 2007;83(986):777-779.

20. Lohuis PJFM. Advanced Caucasian and Mediterranean Rhinoplasty. Amsterdam: Kugler Publications; 2014:2-19.
21. Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Assessment of the learning curve in health technologies. A systematic review. *Int J Technol Assess Health Care*. 2000;16(4):1095-1108.
22. Darzi A, Smith S, Taffinder N. Assessing operative skill. Needs to become more objective. *BMJ*. 1999;318(7188):887-888.
23. Santeon Report 2017. Better Breast Cancer Care Through Collaboration Santeon: Rotterdam.
24. Zorgkaart Nederland. Available at: https://www.zorgkaartnederland.nl/. *Accessed Februari 26, 2021*.
25. Verheul H. Hoge notering Zorgkaart zo gefikst. *Medisch Contact, 2016*. Available at: https://www.medischcontact.nl/nieuws/laatste-nieuws/artikel/hoge-notering-zorgkaart-zo-gefikst.htm. *Accessed Februari 26, 2021*.
26. Sullivan D, Chung KC, Eaves FF, 3rd, Rohrich RJ. The level of evidence pyramid: indicating levels of evidence in Plastic and Reconstructive Surgery articles. *Plast Reconstr Surg*. 2011;128(1):311-314.
27. Eggerstedt M, Brown HJ, Shay AD, et al. Level of Evidence in Facial Plastic Surgery Research: A Procedure-Level Analysis. *Aesthetic Plast Surg*. 2020;44(5):1531-1536.
28. Rhee JS, Arganbright JM, McMullin BT, Hannley M. Evidence supporting functional rhinoplasty or nasal valve repair: A 25-year systematic review. *Otolaryngol Head Neck Surg*. 2008;139(1):10-20.
29. Jones DS, Podolsky SH. The history and fate of the gold standard. *Lancet*. 2015;385(9977):1502-1503.
30. Smith AM, Chaiet SR. Big data in facial plastic and reconstructive surgery: from large databases to registries. *Curr Opin Otolaryngol Head Neck Surg*. 2017;25(4):273-279.
31. Ishii LE, Rhee JS. Harnessing the Power of Data in Facial Plastic and Reconstructive Surgery-From Refuse to Riches. *JAMA Facial Plast Surg*. 2017;19(6):532-533.
32. Medical Economics Staff (2021). Top Challenges 2021: #1 Administrative burdens and paperwork. *Medical Economics Journal. 2021;*98(01).
33. Administratiedruk medisch specialisten: Enquête (Ont)Regel de Zorg VvAA en FMS (2017). Available at: https://www.demedischspecialist.nl/sites/default/files/20171117_DEF%20Rapport-administratiedruk-specialisten.pdf. *Accessed Februari 26, 2021*.
34. Rao SK, Kimball AB, Lehrhoff SR, et al. The Impact of Administrative Burden on Academic Physicians: Results of a Hospital-Wide Physician Survey. *Acad Med*. 2017;92(2):237-243.
35. Physician Administrative Burden Survey - Final Report. Office of Regulatory Affairs and Service Effectiveness September 2020. Available at: https://doctorsns.com/sites/default/files/2020-11/admin-burden-survey-results.pdf. *Accessed Februari 26, 2021*.
36. Kaur M, Pusic A, Gibbons C, Klassen AF. Implementing Electronic Patient-Reported Outcome Measures in Outpatient Cosmetic Surgery Clinics: An Exploratory Qualitative Study. *Aesthet Surg J*. 2019;39(6):687-695.
37. Cella D, Hahn EA, Jensen SE, et al. *Patient-Reported Outcomes in Performance Measurement*. 2017/02/18 ed: Research Triangle Park (NC): RTI Press; 2015.
38. Kluzek S, Dean B, Wartolowska KA. Patient-reported outcome measures (PROMs) as proof of treatment efficacy. *BMJ Evid Based Med*. 2021 Jun 4;bmjebm-2020-111573.
39. Ross EE, Anari S. Clinical commissioning group criteria for rhinoplasty and septorhinoplasty in England: survey of current guidance. *J Laryngol Otol*. 2018;132(7):591-595.
40. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*. 2008;336(7644):601-605.

41. Byrne PJ. The Role of Objective Outcomes in Surgery in Overcoming Cognitive Dissonance. *JAMA Facial Plast Surg.* 2016;18(3):163-164.
42. Chau H, Dasgupta R, Sauret V, Kenyon G. Use of an optical surface scanner in assessment of outcome following rhinoplasty surgery. *J Laryngol Otol.* 2008;122(9):972-977.
43. Largo RD, Wettstein R, Fulco I, et al. Three-dimensional laser surface scanning in rhinosurgery. *Facial Plast Surg.* 2013;29(2):116-120.
44. Kortbus MJ, Ham J, Fechner F, Constantinides M. Quantitative analysis of lateral osteotomies in rhinoplasty. *Arch Facial Plast Surg.* 2006;8(6):369-373.
45. Van Heerbeek N, Ingels KJAO, Van Loon B, Plooij JM, Bergé SJ. Three dimensional measurement of rhinoplasty results. *Rhinology.* 2009;47(2):121-125.
46. Szychta P, Witmanowski H, Rykala J. Assessment of the usefulness of three-dimensional scanner in aesthetic evaluation of post-traumatic rhinoplasty. *Journal of Plastic Surgery and Hand Surgery.* 2013;47(2):106-112.
47. Hassan Y, Leveille CF, Gallo L, et al. Reporting Outcomes and Outcome Measures in Open Rhinoplasty: A Systematic Review. *Aesthet Surg J.* 2020;40(2):135-146.
48. Little AC. Facial attractiveness. *Wiley Interdiscip Rev Cogn Sci.* 2014;5(6):621-634.
49. Perrett DI, May KA, Yoshikawa S. Facial shape and judgements of female attractiveness. *Nature.* 1994;368(6468):239-242.
50. Chait LA, Widgerow AD. In search of the ideal nose. *Plast Reconstr Surg.* 2000;105(7):2561-2567; discussion 2568-2572.
51. Slupchynskyj O, Gieniusz M. Rhinoplasty for African American patients: a retrospective review of 75 cases. *Arch Facial Plast Surg.* 2008;10(4):232-236.
52. Patel PN, Most SP. Concepts of Facial Aesthetics When Considering Ethnic Rhinoplasty. *Otolaryngol Clin North Am.* 2020;53(2):195-208.
53. Toriumi DM, Pero CD. Asian rhinoplasty. *Clin Plast Surg.* 2010;37(2):335-352.

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Chapter 9

English summary

The current era of healthcare is characterized by a gradual paradigm shift. Formerly fueled by volume of services, healthcare is now transiting to a system based on outcomes of delivered services. Outcomes reflect the health benefits that have been achieved with a particular treatment. Information on achieved outcomes is essential to determine the effectiveness of delivered care. Patients rely on outcome information to make educated decisions about their health, providers to choose the most suitable treatment and monitor their performance, and governments and insurance carriers to improve quality and cost-effectiveness of healthcare.

In functional and aesthetic nasal surgery, or rhinoplasty, the patient perspective is crucial to determine the outcome. The aim of this procedure is to improve nasal breathing or the appearance of the nose, as experienced by the patient. Hence, rhinoplasty largely depends on outcomes that reflect patient's views. Currently, however, patient-reported outcomes of rhinoplasty are scarce because they are not routinely measured. This lack of outcome data hinders the establishment of high-level evidence on the effectiveness of rhinoplasty, and compromises transparency about rhinoplasty results towards patients and third parties. The aim of this thesis is to tackle several challenges that refrain rhinoplasty surgeons from collecting and analyzing outcomes in rhinoplasty. Specifically, this thesis provides an evidence-based recommendation on which outcome instruments to use, demonstrates a methodology to implement routine outcome measurements into daily practice, and explores additional clinical advantages resulting from routine outcome measurements. Additionally, nasal attractiveness is quantified, in an attempt to lay a basis for an objective outcome instrument in aesthetic rhinoplasty.

Currently, numerous patient-reported outcome measures (PROMs) of different quality measuring different rhinoplasty outcomes exist, which are used arbitrarily. This causes difficulty in pooling and comparing outcome data and hinders the transparent communication of outcomes to patients and third parties. To make an evidence-based decision on which PROM to use, a systematic evaluation of the quality of each PROM is necessary. A high-quality PROM is interpretable, feasible and carries adequate measurement properties such as validity, reliability and responsiveness. **Chapter 2** describes a systematic review using a consensus-approved guideline to evaluate the measurement properties of all PROMs that measure functional and aesthetic outcomes in rhinoplasty. Among instruments measuring functional outcome, the NOSE scale was found to have the most evidence of good quality. The NOSE scale demonstrated high-quality evidence for sufficient structural validity, internal consistency, reliability, construct validity, and responsiveness, along with favorable interpretability and feasibility aspects, and was therefore selected as the most suitable instrument to measure functional outcome. The recommendation on the best instrument to measure aesthetic outcome was not as outspoken; the FACE-Q

Rhinoplasty Module and the SCHNOS aesthetic subscale were the only instruments that carried evidence for sufficient structural validity and internal consistency. Both PROMs are relatively new compared to the NOSE, which could explain the lack of studies investigating the measurement properties of these scales. Future studies investigating validity, reliability and responsiveness of the FACE-Q Rhinoplasty Module and SCHNOS will be necessary to confirm the quality of these PROMs.

In **chapter 3**, we describe the translation and validation of the NOSE scale into the Dutch language. Apart from the available evidence for adequate quality, the NOSE scale is practical and popular: it has been translated and validated in many different languages. However, no Dutch language version exists. A valid Dutch version that is linguistically and culturally adapted to the original version is mandatory in order to administer the NOSE scale in Dutch patients. Chapter 3 describes the translation process of the NL-NOSE, followed by an assessment of measurement properties of this translated questionnaire. The NL-NOSE demonstrated satisfactory reliability and validity, and is therefore recommended to measure subjective severity of nasal obstruction in Dutch adult patients.

Although the theoretical value of outcome data is evident for many, the anticipated difficulty of integrating routine outcome measurements in the clinical environment has been a common obstacle for healthcare providers, including rhinoplasty surgeons. Routinely collecting outcomes is often considered an administrative or statistical burden, and means to easily aggregate and assess outcome data are scarce. We describe the development and implementation of a prospective rhinoplasty outcome routine in **chapter 4**. Using an automated and real-time data dashboard, an overview of functional and aesthetic results within a rhinoplasty practice and insights into the characteristics of the served population are presented. Additional clinical advantages of using this Rhinoplasty Healthcare Monitor (RHM) related to patient selection, patient empowerment and third-party requests are demonstrated. In **chapter 5**, we demonstrate the benefits of using the RHM in secondary cleft lip and palate rhinoplasty patients. This population differs substantially from general rhinoplasty patients in terms of body image and expectations from surgery, on which the surgeon should anticipate. Apart from patient expectations, expectations of the surgeon regarding functional and aesthetic gain after rhinoplasty should likewise be adjusted since secondary cleft lip rhinoplasty is a particularly difficult procedure. The outcome dashboard exposes these differences and similarly provides additional clinical advantages of using an outcome routine in this specific population.

Routine outcome measurements also benefit the individual surgeon. The rhinoplasty learning curve lasts a lifetime, and its course is sculpted by learning curve accelerators (e.g. number of rhinoplasties performed, attendance to courses) and decelerators (e.g. less focus on rhinoplasty, ageing). One of the most important learning curve accelerators is reflection on achieved results. **Chapter 6** presents two features of the RHM that enhance self-assessment. The first is identifying patients with adverse outcomes, allowing the surgeon to critically appraise the medical charts of these patients and possibly identify mistakes that can be learnt from. The second self-evaluative feature is the display of annual performance, allowing the surgeon to assess longitudinal progress. By graphically representing performance as a function of time, this plot informs the surgeon of annual performance, gives an indication of the gradient of performance development, and may identify trends that are worth evaluating.

Although patient satisfaction remains the major determinant of aesthetic rhinoplasty success, patient-reported outcome is subjective in nature. Scores may be influenced by confounding factors not under the control of the surgeon. An objective measure ignores these influences and could therefore complement patient-reported outcome. Developing an objective aesthetic outcome instrument, however, requires a definition and objective quantification of ideal nasal shape. Based on previous research on facial attractiveness, **chapter 7** explores the role of averageness on nasal attractiveness. In this experimental study the correlation between mathematical averageness of nasal shape and subjective attractiveness is investigated. Photographic series of the face were obtained from Caucasian female volunteers, and converted into a mathematically averaged composite image. A rating panel rated this composite as significantly more attractive than the distribution of ratings for the original images. Furthermore, we were able to calculate a score expressing deviation from average nasal shape, which correlated negatively with attractiveness. These findings imply that an average nose is an attractive nose. Change in nasal shape deviation score following rhinoplasty has potential as an objective aesthetic rhinoplasty outcome measure, although the correlation between averageness, attractiveness, and patient satisfaction needs further research.

Chapter 8 contains a general discussion in which the findings of this thesis are reviewed in a broader perspective and suggestions for future directions are provided. We describe possible routes to better standardization in rhinoplasty outcome measurements, discuss the merits of surgeon-specific performance parameters and consider the role of routinely acquired outcome data in improving evidence levels in rhinoplasty. We conclude with an appeal to collaborate: only by joining forces can surgeons advance from expert opinion to cohort-level evidence and improve data-driven decision-making in rhinoplasty.

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Chapter 10

Nederlandse samenvatting

De afgelopen decennia is er in de zorgsector steeds meer nadruk komen te liggen op resultaten van zorg. Waar evaluatie van zorg in het verleden vooral berustte op kwantiteit, zijn het nu met name de uitkomsten van geleverde zorg die centraal staan. Uitkomsten geven weer welke gezondheidswinst is geboekt met een bepaalde behandeling. Informatie over de daadwerkelijk geboekte gezondheidswinst van geleverde zorg is zowel voor patiënten, zorgverleners, zorgverzekeraars als overheid bijzonder waardevol. Uitkomstinformatie is namelijk essentieel om de kwaliteit van zorg te verbeteren, de kosten van zorg betaalbaar te houden en om een goed onderbouwde keuze te kunnen maken tussen behandelingen of zorgaanbieders.

Bij neuscorrecties, ook wel rhinoplastieken, is het perspectief van de patiënt leidend in het bepalen van de uitkomst. Het doel van deze operatie is namelijk het verbeteren van de neusademhaling of het uiterlijk van de neus, zoals dat door de patiënt zelf wordt ervaren. Patiënt-gerapporteerde uitkomstmetingen, wat wil zeggen dat de patiënt deze uitkomsten zelf beoordeelt en scoort, zijn dus cruciaal om het resultaat van een neuscorrectie te kunnen evalueren. Momenteel zijn patiënt-gerapporteerde uitkomsten van rhinoplastieken echter nauwelijks beschikbaar omdat ze óf niet routinematig worden gemeten, óf omdat ze op teveel uiteenlopende manieren worden gemeten. Dit gebrek aan uniformiteit belemmert de totstandkoming van goed wetenschappelijk bewijs over de effectiviteit van een neuscorrectie. Bovendien is een gestandaardiseerde, breed gedragen uitkomstregistratie noodzakelijk om transparant te kunnen zijn over behandelresultaten richting patiënten en derde partijen. Het doel van dit proefschrift is om neuschirurgen handvatten te bieden die het systematisch verzamelen en analyseren van uitkomsten bevorderen. Concreet gaat het hierbij om een wetenschappelijk onderbouwde aanbeveling welke uitkomstinstrumenten gebruikt zouden moeten worden, een methode om systematische uitkomstregistraties in de dagelijkse praktijk te implementeren, en een demonstratie van additionele voordelen van gestandaardiseerd uitkomsten meten. Daarnaast wordt onderzocht of aantrekkelijkheid van de neus gekwantificeerd kan worden, als potentiële basis voor een objectief esthetisch uitkomstinstrument.

Momenteel is er voor neuscorrecties een breed scala aan verschillende patiënt-gerapporteerde uitkomstmaten (*patient-reported outcome measures*, PROMs) beschikbaar, die willekeurig worden gebruikt. Deze wildgroei bemoeilijkt het samenvoegen en vergelijken van uitkomstgegevens, en belemmert transparantie over uitkomsten richting patiënten en andere geïnteresseerde partijen. Om een wetenschappelijk onderbouwde beslissing te kunnen nemen over welke PROM gebruikt zou moeten worden, is een systematische evaluatie van de kwaliteit van alle beschikbare PROMs noodzakelijk. Een PROM van hoge kwaliteit is niet alleen praktisch

en goed interpreteerbaar, maar beschikt ook over adequate meeteigenschappen zoals validiteit, betrouwbaarheid en responsiviteit. **Hoofdstuk 2** beschrijft een systematische review waarbij de meeteigenschappen van alle functionele en esthetische PROMs voor rhinoplastieken worden geëvalueerd. Van de instrumenten die functioneel resultaat meten, bleek de NOSE schaal over het meeste bewijs voor voldoende kwaliteit te beschikken. Dit geldt met name voor structurele validiteit, interne consistentie, betrouwbaarheid, construct validiteit en responsiviteit. Dit in combinatie met het feit dat deze schaal praktisch en interpretabel is, maakt dat deze schaal het meest geschikt werd geacht om het functionele resultaat van een rhinoplastiek te meten. De aanbeveling ten aanzien van het instrument om het esthetische resultaat in kaart te brengen was niet zo uitgesproken; de FACE-Q Rhinoplastiek Module en de SCHNOS esthetische subschaal waren de enige instrumenten met bewijs voor voldoende structurele validiteit en interne consistentie. Beide PROMs zijn relatief nieuw in vergelijking met de NOSE, wat het gebrek aan studies naar de meeteigenschappen van deze schalen zou kunnen verklaren. Om de kwaliteit van de FACE-Q en de SCHNOS te bevestigen, zullen toekomstige studies nodig zijn die de validiteit, betrouwbaarheid en responsiviteit van deze schalen verder onderzoeken.

In **hoofdstuk 3** beschrijven we de vertaling en validatie van de NOSE-schaal in de Nederlandse taal. De NOSE-schaal is inmiddels in veel verschillende talen beschikbaar, maar een Nederlandstalige versie ontbreekt nog. Om de NOSE-schaal bij Nederlandstalige patiënten af te kunnen nemen moet de vertaalde schaal taalkundig en cultureel gelijkwaardig zijn aan de originele versie. Hoofdstuk 3 beschrijft het vertaalproces van de NL-NOSE, gevolg door een beoordeling van de meeteigenschappen van deze Nederlandse variant. De NL-NOSE toonde voldoende betrouwbaarheid en validiteit en wordt daarom aanbevolen om de subjectieve ernst van neusobstructie te meten bij volwassen Nederlandse patiënten.

Hoewel de meeste zorgverleners het er wel over eens zijn dat uitkomstinformatie waardevol is, worden routinematige uitkomstregistraties in de praktijk nog maar beperkt geïmplementeerd. Belangrijke obstakels hierbij zijn de opvatting dat uitkomstregistraties interfereren met de dagelijkse praktijkvoering, alsmede het gebrek aan een methodiek om uitkomsten snel en gemakkelijk te kunnen analyseren. In **hoofdstuk 4** beschrijven we de ontwikkeling en implementatie van een prospectieve uitkomstregistratie voor rhinoplastiek patiënten. Met behulp van een geautomatiseerd dashboard waarin data real-time gepresenteerd wordt, worden patiëntkarakteristieken en functionele en esthetische resultaten overzichtelijk weergegeven. Tevens worden aanvullende klinische voordelen van het gebruik van deze *Rhinoplasty Healthcare Monitor* (RHM) met betrekking tot patiëntselectie, gedeelde besluitvorming en verantwoording naar

derde partijen beschreven. In **hoofdstuk 5** demonstreren we de voordelen van de RHM bij patiënten met een lip- en/of gehemeltespleet (schisis) die in aanmerking komen voor een secundaire neuscorrectie. Deze populatie verschilt aanzienlijk van algemene neuscorrectiepatiënten als het gaat om zelfbeeld en verwachtingen van chirurgie, waar de chirurg op dient te anticiperen. Omdat een secundaire rhinoplastiek bij schisis een lastige operatie is, moeten de verwachtingen van zowel de patiënt als de chirurg worden bijgesteld. Het dashboard legt deze verschillen bloot en brengt op eenzelfde manier aanvullende klinische voordelen met zich mee in deze specifieke populatie.

Het gestandaardiseerd meten van uitkomsten biedt niet alleen voordelen voor patiënten en andere partijen die geïnteresseerd zijn in resultaten van zorg, maar ook voor de individuele chirurg. Het verrichten van rhinoplastieken kent een levenslange leercurve, en deze curve wordt zowel beïnvloed door factoren van progressie (bijvoorbeeld het aantal uitgevoerde rhinoplastieken, deelname aan cursus) als factoren van regressie (zoals verminderde expositie of veroudering). Eén van de belangrijkste factoren die de hellingshoek van de rhinoplastiek leercurve vergroot is het reflecteren op behaalde resultaten. **Hoofdstuk 6** presenteert twee kenmerken van de RHM die helpen bij het evalueren van eigen presteren. De eerste is het automatisch identificeren van patiënten met ongunstige uitkomsten. Analyse van deze patiënten kan bepaalde patronen of beslissingen boven tafel krijgen die ten grondslag lagen aan de teleurstellende resultaten. Dit zou waardevolle lessen voor de toekomst op kunnen leveren. De tweede functie die zelf-evaluatie bevordert is de weergave van jaarlijkse prestaties, waarmee de chirurg zijn eigen voortgang kan beoordelen. Door een gemiddelde van behaalde uitkomsten per jaar uit te zetten tegen de tijd, krijgt de chirurg een idee van jaarlijkse prestaties, prestatieontwikkeling en trends in de prestatiecurve die het evalueren waard zijn.

Hoewel bij esthetische rhinoplastieken de tevredenheid van de patiënt de belangrijkste determinant blijft van een goed resultaat, heeft de subjectieve aard van een PROM ook nadelen. Scores kunnen worden beïnvloed door factoren die los staan van het technisch resultaat en die de chirurg niet altijd onder controle heeft. Een objectieve uitkomstmaat negeert dit soort invloeden en zou daarom een goede aanvulling kunnen zijn op patiënt-gerapporteerde uitkomsten. Het ontwikkelen van een objectief esthetisch uitkomstinstrument vereist echter een definitie en objectieve kwantificering van het best mogelijke, of 'ideale' resultaat. Gebaseerd op bestaand onderzoek naar aantrekkelijk van het hele gezicht, onderzoekt **hoofdstuk 7** de rol van gemiddeldheid bij het bepalen van de aantrekkelijkheid van de neus. In deze experimentele studie wordt de correlatie tussen mathematische gemiddeldheid van de vorm van de neus en subjectieve aantrekkelijkheid van de neus onderzocht. Hierbij werden fotoseries

van vrouwelijke Kaukasische vrijwilligers mathematisch samengesmolten tot een gemiddelde neusvorm. Een jury beoordeelde deze samengestelde neus als aanzienlijk aantrekkelijker dan de neuzen op de originele foto's. Daarnaast wordt in dit hoofdstuk een score berekend die de deviatie ten opzichte van de gemiddelde neusvorm uitdrukt, wat negatief bleek te correleren met aantrekkelijkheid. Deze bevinding impliceert dat een gemiddelde neus een aantrekkelijke neus is. Een vergelijking tussen preoperatieve deviatie van gemiddeld en postoperatieve deviatie van gemiddeld zou in potentie de esthetische uitkomst weer kunnen geven in objectieve zin. Wel wordt hierbij opgemerkt dat de correlatie tussen gemiddeldheid, aantrekkelijkheid en de correlatie met PROMs verder onderzoek behoeft.

Hoofdstuk 8 bevat een algemene discussie waarin de bevindingen van dit proefschrift in een breder perspectief worden bekeken en suggesties voor toekomstige studies worden gedaan. We beschrijven mogelijke routes naar een betere standaardisatie van uitkomstmetingen van rhinoplastieken, bespreken de waarde van chirurg-gebonden prestatieparameters en benoemen de rol van uitkomstregistraties bij het verhogen van bewijsniveaus in de rhinoplastiek-gerelateerde literatuur. We sluiten af met een oproep om de handen ineen te slaan: alleen door krachten te bundelen kunnen we wetenschappelijke bewijsniveaus naar een hoger niveau tillen, en data-gestuurde besluitvorming rondom rhinoplastieken maximaliseren.

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Appendix

List of abbreviations

List of abbreviations

RCT	Randomized Controlled Trial
PROM	Patient-Reported Outcome Measure
VBHC	Value-Based Healthcare
CT	Computed Tomography
PNIF	Peak Nasal Inspiratory Flow
NOSE	Nasal Obstruction Symptom Evaluation
NASION	Non-Invasive Assessment and Symptomatic Improvement of the Obstructed Nose
SCHNOS	Standardized Cosmesis and Health Nasal Outcomes Survey
RhinoQoL	Rhinologic Quality of Life
SNOT	Sino-Nasal Outcome Test
SF	Short Form
GHQ	General Health Questionnaire
ROE	Rhinoplasty Outcome Evaluation
UQ	Utrecht Questionnaire
DAS	Derriford Appearance Scale
MBSRQ	Multidimensional Body-Self Relations Questionnaire
RSE	Rosenberg Self Esteem
COSMIN	Consensus-based Standards for the selection of health Measurement Instruments
MIC	Minimal Important Change
VAS	Visual Analogue Scale
FROI	Functional Rhinoplasty Outcome Inventory
NSQ	Nasal Surgical Questionnaire
HR-PRO	Health-Related Patient-Reported Outcomes
WHO	World Health Organization
ICC	Intraclass Correlation Coefficient
CFI	Comparative Fix Index
TLI	Tucker-Lewis Index
SD	Standard Deviation
SRM	Standardized Response Mean
CFA	Confirmatory Factor Analysis
GRM	Graded Response Models
SRMR	Standardized Root Mean squared Residual
RMSEA	Root Mean Squared Error of Approximation
pclose	Probability of RMSEA ≤ 0.05
rho	Spearman correlation

M-W	Mann-Whitney U test
ICHOM	International Consortium for Health Outcomes Measurement
RHM	Rhinoplasty Healthcare Monitor
TPCC	Tutoplast™ Processed Costal Cartilage

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Appendix

Affiliations of contributing authors

Affiliations of contributing authors

Department of Otorhinolaryngology and Head and Neck Surgery, Erasmus Medical Center, Rotterdam, The Netherlands

Frank R. Datema, Egge F. van der Poel, Robert J. Baatenburg de Jong, Carolina E. Touw.

Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands

Juanita A. Haagsma

Department of Medical Psychology and Psychotherapy, Erasmus University Medical Center, Rotterdam, The Netherlands

Reinier Timman

Department of Plastic Surgery, Erasmus Medical Center, Rotterdam, The Netherlands

Sarah Versnel

Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, The Netherlands

L. (Wieneke) B. Mokkink

Department of Otorhinolaryngology and Head and Neck Surgery, University Hospital Sveti Duh, Zagreb, Croatia

Peter J.F.M. Lohuis

School of Psychology and Neuroscience, University of St Andrews, St Andrews, UK

David I. Perrett, Dengke Xiao

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Appendix

List of publications

List of publications

D.A. Monserez, S. Keereweer, **F.V.W.J. van Zijl**, S. Koljenovic, R.P. Takes, J.A.U. Hardillo, R.J. Baatenburg de Jong. Adjuvant topical 5-fluorouracil versus adjuvant radiotherapy in macroscopically resected sinonasal intestinal-type adenocarcinomas. *Under review*

F.V.W.J. van Zijl, P.C.J. de Laat, A.P. Nagtegaal, F.R. Datema. Aggressive vascular tumor mimicking posttraumatic hematoma: a case report of kaposiform hemangioendothelioma on the nose. *Under review*

F.V.W.J. van Zijl, M.L. Gerdes, J. Kusmierczyk, H.C. Hafkamp. Intranasale corticosteroïden geven geen verhoogd risico op oftalmologische complicaties: een systematische review. *Nederlands Tijdschrift voor Allergie, Astma en Klinische Immunologie*. 2021;21(4):134-41

F.V.W.J. van Zijl, P.J.F.M. Lohuis, F.R. Datema. The rhinoplasty health care monitor: using validated questionnaires and a web-based outcome dashboard to evaluate personal surgical performance. *Facial Plastic Surgery and Aesthetic Medicine***.** 2021 Feb 22; online ahead of print

F.V.W.J. van Zijl, D.I. Perrett, P.J.F.M. Lohuis, C. Touw, D. Xiao, F.R. Datema. The value of averageness in aesthetic rhinoplasty: humans like average noses. *Aesthetic Surgery Journal*. 2020 Nov 19;40(12):1280-1287

F.V.W.J. van Zijl, M.L. Gerdes, J. Kusmierczyk, H.C. Hafkamp. Het gebruik van intranasale corticosteroïden geeft geen verhoogd risico op oftalmologische complicaties zoals glaucoom en cataract: een systematische review. *Nederlands Tijdschrift voor Keel-Neus-Oorheelkunde*. 2020, 26e jaargang, nr 2.

F.V.W.J. van Zijl, W.M. Mokkink, J.H. Haagsma, F.R. Datema. Evaluation of measurement properties of symptom measurement instruments for functional and aesthetic rhinoplasty: a systematic review. *JAMA Facial Plastic Surgery*. 2019 Mar 1;21(2):152-162

F.V.W.J. van Zijl, S. Versnel, E.F. van der Poel, R.J. Baatenburg de Jong, F.R. Datema. Benefits of routine prospective functional and aesthetic outcome measurements in secondary cleft lip rhinoplasty. *JAMA Facial Plastic Surgery*. 2018 Dec 1;20(6):488-494

O. Bugter, D.A. Monserez, **F.V.W.J. van Zijl**, R.J. Baatenburg de Jong, J.A.U. Hardillo. Surgical management of inverted papilloma; a single-center analysis of 247 patients with long follow-up. *Journal of Otolaryngology and Head and Neck Surgery*. 2017 Dec 20;46(1):67

F.R. Datema, **F.V.W.J. van Zijl**, E.F. van der Poel, R.J. Baatenburg de Jong, P.J.F.M. Lohuis. Transparency in functional rhinoplasty: benefits of routine prospective outcome measurements in a tertiary referral center. *Plastic and Reconstructive Surgery*. 2017 Oct;140(4):691-702

F.V.W.J. van Zijl, R. Timman, F.R. Datema. Adaptation and validation of the Dutch version of the nasal obstruction symptom evaluation (NOSE) scale. *European Archives of Otorhinolaryngology*. 2017 Jun;274(6):2469-2476

F.V.W.J. van Zijl, D.A. Monserez, T.I.M. Korevaar, O. Bugter, M.H. Wieringa, R.J. Baatenburg de Jong, J.A.U. Hardillo. Postoperative value of serum squamous cell carcinoma antigen as a predictor of recurrence in sinonasal inverted papilloma. *Clinical Otolaryngology*. 2017 Jun;42(3):528-535

F.V.W.J. van Zijl, P.A. van der Eerden. De subcutaan gesteelde wekedelenlap: een effectieve behandeling voor chondrodermatitis nodularis van de antihelix. *Nederlands Tijdschrift voor Keel-Neus-Oorheelkunde*. 2015, 21e jaargang, nr 1

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Appendix

PhD portfolio

PhD portfolio

Summary of PhD training and teaching activities	
Name of PhD student:	Floris V.W.J. van Zijl
Erasmus MC dept. of:	Otolaryngology and Head & Neck Surgery
Promotor:	Prof. Dr. R.J. Baatenburg de Jong
Co-promotor:	Dr. F.R. Datema

	Year(s)	ECTS
General courses		
Research integrity	2018	0.3
Basic surgical exam	2015	0.5
SPSS basic introduction course	2014	1.0
Advanced Life Support	2014	1.0
Specific courses		
International reconstructive and aesthetic facial surgery course, Nijmegen	2021	1.0
27th Stuttgart Advanced Course for Rhinoplasty, Stuttgart	2019	1.0
ENT-training 'ENTER' courses	2015-2019	3.0
Desiderius Education 'discipline overstijgend onderwijs'	2015-2019	2.0
PhD-related presentations		
PROMs in facial plastic surgery		
Virtual Pan Scotland Registrar Training Day Facial Plastic Surgery	2022	0.5
Self-assessment using the Rhinoplasty Healthcare Monitor		
Annual meeting of EAFPS, Nice	2021	0.5
Humans like average noses		
Journal Club Aesthetic Surgery Journal	2021	0.5
Learning and improving using rhinoplasty outcome routine		
Scientific Event Dept. of Otorhinolaryngology Erasmus MC, Rotterdam	2020	0.5
EAFPS Fellowship: experience and practical tips		
Annual meeting of EAFPS, virtual	2020	0.5

	Year(s)	ECTS
EAFPS juniors		
Netherlands Society of Otorhinolaryngology meeting, Nieuwegein	2019	0.5
The role of averageness in aesthetic rhinoplasty		
Accredited course Erasmus MC, Rotterdam	2019	0.5
Research as EAFPS junior: benefits and opportunities		
Annual meeting of EAFPS, Amsterdam	2019	0.5
How to select an appropriate rhinoplasty outcome instrument		
Netherlands Society of Otorhinolaryngology meeting, Nieuwegein	2018	0.5

Conferences		
Annual Meeting of European Academy for Facial Plastic Surgery, Nice	2021	1.0
Attendance to biannual Netherlands Society of Otorhinolaryngology meeting	2015-2021	4.0
Attendance to annual research day otorhinolaryngology, Erasmus MC	2015-2021	2.0
Annual Meeting of European Academy for Facial Plastic Surgery, Virtual	2020	1.0
Annual Meeting of European Academy for Facial Plastic Surgery, Amsterdam	2019	1.0
4th Congress of European Otorhinolaryngology Head Neck Surgery, Barcelona	2017	1.0

Teaching		
Direct supervision of graduation research of MSc student	2020-2021	1.0
Supervising various workgroups for 1st, 3rd and 5th year medical students	2015-2019	1.0
ENT-education for medical students, theatre nurses and oncology nurses	2015-2019	2.0

	Year(s)	ECTS
Other academic activities		
Boardmember Junior Board of European Academy for Facial Plastic Surgery	2018-2021	4.0
Organizer 'Symposium Experimenteel Onderzoek Heelkundige Specialismen'	2018	2.0
Research visit prof. D.I. Perrett's 'Perception Lab', St. Andrews	2016	1.0
Total		35.3

1 ECTS (European Credit Transfer System) is equal to a workload of 28 hours

9
10
11
12
13
14
15
33
34
35
36
37
38
39

Appendix

About the author

About the author

Floris Vincent Willem Joseph van Zijl was born on August 22nd 1988 in Maastricht. Following numerous visits to the operating theatre with his father, a gynecologist, Floris' interest in medicine was aroused at an early stage. In 2006 he graduated from the Alfrink College, Zoetermeer, and started his medical degree at Erasmus University in Rotterdam. Apart from Rotterdam, his medical studies brought him to clinical internships at the Hospital Calderón Guardia in San José, Costa Rica (dr. A. Fonseca) and Academisch Ziekenhuis Paramaribo in Suriname (dr. A.R. van Kanten). During medical school he developed a specific interest in the head and neck area, as evidenced by his position as course director Head and Neck at the Erasmus Anatomy Research Programme and an additional internship in facial plastic and reconstructive surgery (dr. P.A. van der Eerden).

Floris graduated in 2014 and started his medical career as a senior house officer at the departments of Surgical Oncology and Head and Neck Oncology at the former Erasmus MC location Daniel den Hoed. He commenced specialty training in otolaryngology and head and neck surgery in 2015 under auspices of prof.dr. R.J. Baatenburg de Jong, dr. R.M. Metselaar, dr. H.M. Blom and dr. F.A.W. Peek. Owing to his interest in facial plastic surgery and specifically rhinoplasty, Floris started a PhD project on rhinoplasty parallel to his residency. As a PhD candidate, he participated in several courses and meetings in Barcelona, Amsterdam, Stuttgart, London and Nice, and visited St. Andrews University in Scotland to work with the group of prof.dr. D.I. Perrett.

Floris qualified as an ENT-surgeon in 2019 and pursued his education at Gelderse Vallei Hospital in Ede and Radboud University Medical Center in Nijmegen. Under supervision of dr. W.M. Boek and dr. K.J.A.O. Ingels he completed a 1-year fellowship in facial plastic and reconstructive surgery, for which he was awarded a European Academy for Facial Plastic Surgery grant. In pursue of board certification, Floris successfully sat the oral and written examination of the International Board for Certification in Facial Plastic and Reconstructive Surgery in London. In 2020, Floris acquired a position as consultant ENT-surgeon / Rhinologist at Erasmus MC, with rhinoplasty as his specific field of interest.

Floris lives in Rotterdam with his wife Tessa, who also specializes in otorhinolaryngology. Together they have two children, Louis and Rosa.