Evaluation of nasal septal deviation in patients with chronic rhinosinusitis – an interrater agreement study *

Anne Stockmann¹,², Kenneth Lenhardt Larsen³, Bibi Lange³, Peter Darling³, Gita Jørgensen³, Lisbeth Høgedal³, Anette Drøhse Kjeldsen³

¹ University of Southern Denmark, Odense, Denmark
² Research Unit for Otorhinolaryngology Head and Neck Surgery, Odense University Hospital, Odense, Denmark
³ Department of Radiology, Odense University Hospital, Odense, Denmark

Abstract

Background: The significance of nasal septal deviations may be hard to evaluate. Patient history, clinical examination, nasal endoscopy and sinus CT scans contribute in the evaluation. We aimed to investigate the interrater agreement in the evaluation of nasal septal deviations in patients with chronic rhinosinusitis (CRS).

Methodology: A total of 30 patients were included in the study. Three rhinologists using nasal endoscopy evaluated the presence and degree of septal deviation. A rhinologist and a radiologist also evaluated the presence and degree of septal deviation on sinus CT scans. Interrater agreement was measured using unweighted Fleiss’ kappa ($K_f$).

Results: In the endoscopic evaluation of septal deviation, the raters attained a $K_f$ of 0.31 (SE 0.12), 0.33 (SE 0.11) and 0.37 (SE 0.11) for the assessment of anterior deviations, inferior/posterior deviations and deviations by the perpendicular plate, respectively. In the radiologic evaluation of septal deviation, the raters attained a $K_f$ of 0.52 (SE 0.13), 0.63 (SE 0.16) and 0.38 (SE 0.16) for the assessment of anterior deviations, inferior/posterior deviations and deviations by the perpendicular plate, respectively.

Conclusion: Our study showed a limited agreement in the endoscopic evaluation of septal deviation. Conversely, a higher agreement was achieved in the radiologic evaluation. Sinus CT scans might be a helpful objective examination technique, but it cannot be an alternative to clinical evaluation using acoustic rhinometry, rhinomanometry or Peak Nasal Inspiratory flow.

Key words: nasal septum, nasal obstruction, nasal surgical procedures, observer variation, endoscopy

Introduction

Clinical examination, consisting of anterior rhinoscopy, nasal endoscopy, and sinus CT scan, are useful examination techniques for the classification of nasal septal deviation and preoperative planning prior surgery⁹-¹⁰. Although CT scan is normally only a part of the evaluation if the patients also have sinus disease, it may be a helpful examination if the nasal stenosis is not fully explained by nasal endoscopy.

Nasal septal deviations are anatomically often described as C- or S-shaped deviations in the vertical or horizontal plane⁶. However, previous research has shown that the shape and size of septal deviation not always interconnects with the symptoms and the need for surgery – thus the clinical significance of the septal deviation⁹¹. The majority of rhinologists would agree that total occlusion of one nasal cavity represents a significant deviation in any part of the nasal cavity, but more discussion occur regarding significance of partly occlusion where the location of deviation inside the nasal cavity may also play a role. The location of deviations inside the nasal cavity are not always noted; although anterior deviations may significantly reduce the nasal airway; deviations by the perpendicular plate may cause obstruction primarily of the middle meatus; and inferior/posterior obstruction of the nasal cavity at the site of Vomer may cause both types of obstructions. The lack of a standardized scoring system and difficulty in addressing significant septal deviations might lead to some degree of subjectivity in otherwise objective
Evaluation of nasal septal deviation

examination techniques\textsuperscript{6, 8, 9}. Although objective examination techniques are important, research concerning the interrater variability of nasal endoscopy and sinus CT scan for the evaluation of septal deviation is limited. The aim of our study was to determine the interrater agreement in the endoscopic and radiologic evaluation of nasal septal deviations in patients with CRS. Additionally, we sought to evaluate clinicians’ ability to predict clinically significant deviations when blinded to patients’ symptoms.

Materials and Methods
Study design and setting
An observational, single-institutional study was conducted to evaluate the interrater agreement in the assessment of septal deviation in patients with CRS. Patients were included between September 2015 and March 2016 at the Department of Otorhinolaryngology, Odense University Hospital, Odense, Denmark.

Participants
A total of 31 patients were referred to the study by ENT practitioners and were diagnosed with CRS according to the criteria outlined in the European Position paper on Rhinosinusitis and Nasal polypos (EPOS 2012). To evaluate the severity of CRS the patients completed the Sino-Nasal Outcome Test (SNOT-22) and a Visual Analogue Scale (VAS) on overall sino-nasal symptoms. Inclusion criteria have been described previously\textsuperscript{10}. Exclusion criteria were: previous surgery in the nose and/or paranasal sinuses, failure to complete the full examination, large septal perforation, pregnancy and insufficient language skills.

Nasal endoscopy
Three rhinologists, all blinded to patients’ medical history and symptoms, performed bilateral anterior rhinoscopy after which the last rater administered topical decongestion (lidocaine 5% phenylephrine hydrochloride 1%) in the nasal cavity. Then, the raters performed nasal endoscopy (0-degree 4 mm rigid endoscope). Each rater noted the presence and degree of deviations at the anterior part of the septum, the inferior/posterior part of the septum (vomer), and the superior part of the septum (perpendicular plate). This subdivision of the nasal septum was used throughout this study and it is illustrated in Figure 1. Each nasal cavity was rated as 0 (no deviation), 1 (mild non-significant deviation) or 2 (severe significant deviation) for each of the three types of septal deviation as shown in Figure 2.

After examination, blinding was removed, and symptoms and medical history were revealed to the raters. Clinical significant septal deviation was defined by objective significant septal deviation and sino-nasal symptoms considered related to the septal deviation. If clinical significant septal deviation was concluded the patient was offered and underwent septal surgery. Thus, we chose to perceive the offer of septoplasty as an indicator of a clinical significant septal deviation.

Sinus CT scans
Following clinical examination, patients underwent sinus CT scans. An experienced rhinologist and an experienced radiologist evaluated the presence and degree of septal deviations. The findings were quantified as shown in Figure 3. The sinus CT scans were evaluated using coronal and axial projections and the raters were blinded to symptoms, medical history and objective findings.

Figure 1. Subdivision of the nasal septum: anterior part, inferior/posterior part (vomer) and superior part (perpendicular plate).

Figure 2. Nasal septal deviation scoring system for nasal endoscopy.

<table>
<thead>
<tr>
<th>Rating</th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior deviation of septal nasal cartilage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior/posterior deviation (0, 1, 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviation by the perpendicular plate (0, 1, 2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Deformities or similar factors leading to parameters impossible to assess was marked with X.

Figure 3. Nasal septal deviation scoring system for sinus CT scans.

<table>
<thead>
<tr>
<th>Rating</th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior deviation of septal nasal cartilage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior/posterior deviation (0, 1, 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviation by the perpendicular plate (0, 1, 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a: Conflated posteriorly by aperture piriformis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b: Vomer, including septal spur</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c: Just posterior to cr nares — or where the perpendicular plate may be visible in the nasal septum — towards the level where bulbus oscli has its greatest diameter.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Deformities or similar factors leading to parameters impossible to assess was marked with X.
Statistical analysis
The statistical analysis was based on data from 30 observations obtained from the initial study by Larsen et al. (10). The agreement between raters was addressed using unweighted Fleiss’ kappa ($K_f$) where a $K_f$ of -1 indicated complete disagreement, a $K_f$ of 0 indicated agreement expected by chance and a $K_f$ of 1 indicated complete agreement. P values below 0.05 were considered statistically significant. A bootstrap method (1000 bootstrap replications) was used to calculate a bias corrected 95% confidence interval (95% CI).

Statistical analysis was performed using STATA version 15.1 (StataCorp LLC, Lakeway Drive College Station, TX, USA).

Results
Thirty-one patients were included in the study. One patient was excluded due to a large septal defect resulting in a study population of thirty patients consisting of 20 males (66.7%) and 10 females (33.3%). Patients had an average age of 50 years (range: 28-77), an average SNOT-22 score of 43.17 (range: 9-75, SD 18.47) and an average VAS score of 6.64 (range: 2.2-9.8, SD 2.12). SNOT-22, VAS and the prospective indication for Functional Endoscopic Sinus Surgery are only descriptive for the population, and are not used in the evaluation of septal deviation. Fifteen patients (50%) were offered and underwent septoplasty.

Nasal endoscopy
The evaluation of the nasal cavities of each patient resulted in 30 observations for each category from each rater. In the assessment of anterior deviations, 72.2%, 7.8% and 20.0% of the cases were rated as 0, 1 and 2, respectively. In the assessment of inferior/posterior deviations, 35.6%, 22.2% and 35.6% of the cases were rated as 0, 1 and 2, respectively. In the assessment of deviations by the perpendicular plate, 62.2%, 14.4% and 18.9% of the cases were rated as 0, 1 and 2, respectively. Patients, where one or more raters marked any of the three parameters as unable to assess, were excluded from analysis. Hence, observations from 30 patients, 26 patients and 28 patients were used to calculate a $K_f$ for the anterior deviations, inferior/posterior deviations and deviations by the perpendicular plate, respectively. The $K_f$ values are shown in Table 1.

<table>
<thead>
<tr>
<th>Deviation</th>
<th>$K_f$</th>
<th>SE</th>
<th>95% CI</th>
<th>$z$ value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior deviation</td>
<td>0.31</td>
<td>0.12</td>
<td>0.09 to 0.55</td>
<td>3.63</td>
<td>0.001</td>
</tr>
<tr>
<td>Inferior/posterior deviation</td>
<td>0.33</td>
<td>0.11</td>
<td>0.15 to 0.55</td>
<td>4.11</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Deviation by the perpendicular plate</td>
<td>0.37</td>
<td>0.11</td>
<td>0.19 to 0.61</td>
<td>4.56</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

The p values and 95% CIs suggest the agreement between the raters was not due to chance as the evaluation of all parameters obtained a $K_f$ above 0. The best interrater agreement was achieved in the assessment of deviation by the perpendicular plate.

Furthermore, we evaluated the raters’ ability to predict clinically significant deviations by comparing deviations marked with the rating 2 and the number of patients offered septoplasty. Table 2 shows the distribution of assessments for each rater and whether patients were offered septoplasty or not.

Rater 1 achieved a sensitivity of 66.67% and a specificity of 93.33%. Rater 2 achieved a sensitivity of 73.33% and specificity of 80%. Rater 3 achieved a sensitivity of 100% but a specificity of 60%.

Sinus CT scans
Thirty observations were used to calculate a $K_f$ for radiologic evaluation of septal deviation. In the rating of all three types of deviations, the rhinologist used the rating 2, 11 times whereas the radiologist used the rating 2, 9 times. Table 3 shows the $K_f$ values for the right and left side of the nasal septum along with an average $K_f$ value for each deviation.

The p values and 95% CIs suggest the agreement between our raters was not achieved just by chance as the evaluation of all parameters obtained a $K_f$ above 0. The greatest interrater agreement was achieved in the assessment of inferior/posterior deviations. However, wide 95% CIs are seen in all parameters, which indicates a substantial uncertainty concerning the reported $K_f$ values.
Evaluation of nasal septal deviation

Discussion

Synopsis of key findings

In the endoscopic evaluation of septal deviations, the raters attained a level of agreement significantly greater than what should be expected by chance for all three parameters. However, specifying the null-hypothesis as $K_f = 0$, has a questionable value when measuring agreement between clinicians in a diagnostic examination\(^1\). Sim et al. suggest a clinical cut off value of 0.40 and a value below this should be deemed unacceptable for use in clinical health care\(^1\). None of the three parameters achieved a $K_f$ value above 0.40 comparing endoscopic evaluation of the nasal cavity. This illustrates the possible limitation of nasal endoscopy as a sole standing (objective) examination technique and it is reasonable to assume that some degree of subjectivity in the evaluation of septal deviation is inevitable. In the assessment of a clinical significant septal deviation our raters were able to achieve a sensitivity of 66.67%, 73.33% and 100% as well as a specificity of 93.33%, 80% and 60%.

Using a clinical cut off value of 0.40 in the radiologic evaluation of septal deviations shows that anterior deviations and inferior/posterior deviations can be objectively evaluated using sinus CT scans. However, the agreement in the evaluation of deviation by the perpendicular plate falls below 0.40 indicating a limited agreement. This indicates that a clinician - in this case a rhinologist - and a radiologist in some cases disagree in the assessment of the significance of deviations by the perpendicular plate. We have speculated that a reason for this disagreement could be differences in clinical experience, as the rhinologist has seen and treated several patients with septal deviations, while the radiologist evaluates the x-ray images and describes extend of pathology, but without having experience of treating the patients.

Strength and weakness of the study

The most obvious strength of our study is the contribution to field of classification of septal deviation by including an inter-rater analysis based on nasal endoscopy and sinus CT scans. To the best of our knowledge, only one study has investigated the interrater agreement in the endoscopic evaluation of septal deviation and no studies have investigated the interrater agreement in the radiologic evaluation of septal deviation. However, potential limitations should be considered. All participants in the present study were diagnosed with CRS, which might suggest that our raters encountered a greater number of cases with septal deviation than if we had used a sample from the general population\(^1\). Using patients diagnosed with CRS also made it difficult to distinguish between symptoms caused by CRS and septal deviation, as the inflammatory state in the nasal epithelium in patients with CRS can cause a sensation of nasal obstruction without necessarily an objective obstruction of nasal flow. Unfortunately, as our patients were diagnosed with CRS, we were not able to use SNOT-22 to address symptoms caused by septal deviation. Therefore, CRS and SNOT-22 were only descriptive for our population and the evaluation of septal deviation and indication for septoplasty was independent of this. Thus, the study could, with advantage, have been performed on a study population without CRS. This would, however, imply CT scans in patients that do not normally have CT scans performed.

The anatomical classification of the septum, as shown in Figure 1, was chosen by the study group. Other classifications could be used e.g. subdivision into nasal valve, maxillary crest, cartilaginous septum and osseous septum or the well-known and widely used Mladina classification\(^1\). A relatively large number of observations was rated as 0 in all categories. A heavily one-sided distribution of ratings will

Table 3. Agreement analysis for radiologic evaluation of nasal septal deviation.

<table>
<thead>
<tr>
<th>Category</th>
<th>$K_f$</th>
<th>SE</th>
<th>95% CI</th>
<th>z value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior deviation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>0.45</td>
<td>0.13</td>
<td>0.08 to 0.82</td>
<td>3.39</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Right</td>
<td>0.59</td>
<td>0.13</td>
<td>0.18 to 1.00</td>
<td>4.67</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Total</td>
<td>0.52</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior/posterior deviation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>0.66</td>
<td>0.14</td>
<td>0.40 to 0.89</td>
<td>4.77</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Right</td>
<td>0.60</td>
<td>0.17</td>
<td>0.27 to 0.93</td>
<td>3.49</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Total</td>
<td>0.63</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviation by the perpendicular plate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>0.27</td>
<td>0.14</td>
<td>0.00 to 0.57</td>
<td>1.87</td>
<td>0.031</td>
</tr>
<tr>
<td>Right</td>
<td>0.49</td>
<td>0.17</td>
<td>0.06 to 0.83</td>
<td>2.95</td>
<td>0.0016</td>
</tr>
<tr>
<td>Total</td>
<td>0.38</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
lead to a large expected agreement and impact the size of $K_f$. Hoehler argues that the effect of rating distribution should not be omitted by adjusting it, but a sample size consisting of equal amount of participants with and without the presence of septal deviations would keep the expected agreement low without artificially adjusting for it. If nasal endoscopy is a truly useful tool for objective examination it would not matter how large the expected agreement is for the raters to agree, but if a degree of subjectivity is present, a large expected agreement could hamper the interrater agreement. Our choice to include a third rating category "2" was chosen to assess whether the raters were capable of agreement not only in determining the presence of a deviation but also in assessing whether a given deviation was expected to be of clinical significance for the patient. The addition of a third rating category might reduce the $K_f$ value, but we chose to include it nonetheless. Lastly, the interrater agreement analysis of the radiologic evaluation of septal deviation showed wide 95% CIs indicating uncertainty in the results. An increased sample size and/or number of raters would probably result in more precise results.

Comparisons with other studies
Our results have several similarities with Raithatha et al. investigating the agreement of nasal endoscopy in evaluation of septal deviation. Fourteen patients with CRS underwent digital video nasal endoscopy, which were reviewed by five blinded rhinologists. In the assessment of septal deviation, they achieved a $K_f$ value of 0.240 showing an even lower interrater agreement compared to our results. Regarding differences in study design, Raithatha and colleagues used video recordings instead of real-time nasal endoscopy. Using video recordings makes it impossible for the rhinologist to adjust the nasal endoscope and this might explain that a higher agreement is obtained when using real-time endoscopy. Compared to our study, Raithatha et al. included fewer patients, but more raters, resulting in narrower 95% CI and thereby precise results. However, they have simplified the assessment of septal deviation as their raters consider the entire nasal septum without subdividing it into regions. A greater precision in the assessment of septal deviation must be preferred wherefore the optimal study design might include our level of detail in the assessment of septal deviation combined with the number of raters used in the study by Raithatha et al.

A study by Sedaghat et al. investigated the ability of two otolaryngologists to predict the clinical significance of nasal septal deviations. This evaluation was based on medical history and clinical examination consisting of anterior rhinoscopy and nasal endoscopy. They achieved a sensitivity and specificity of 86.9% and 91.8%, respectively, which is higher compared to the calculated average sensitivity and specificity of our raters of 80% and 77.78%, respectively. This strongly suggests that medical history also needs to be considered when clinicians assess whether a septal deviation is of clinical significance or not. If the raters in our study had been allowed such insight it is most likely that the interrater agreement would have been higher.

The interrater agreement in the radiologic evaluation of septal deviation showed a significant agreement between a rhinologist and radiologist in the evaluation of anterior deviations and inferior/posterior deviations suggesting that our scoring system for classification of septal deviations is reproducible and useful in a clinical setting. However, applicability of sinus CT scans in the assessment of septal deviation might be questionable. Sedaghat et al. investigated the assessment of nasal septal deviation when evaluated by clinical examination (anterior rhinoscopy and nasal endoscopy) and sinus CT scans. The clinical examinations were performed by an otolaryngologist whereas the sinus CT scans were reviewed by a neuroradiologist. They showed that clinical examination was superior to sinus CT scans and the usefulness of sinus CT scans was limited. Nevertheless, in this present study we have shown that the interrater agreement in the evaluation of septal deviation by clinical examination is limited and below clinical acceptable, while the agreement in the radiologic evaluation of septal deviation is significant. Even though Sedagath et al. found that clinical examination is superior to sinus CT scans, the validity of nasal endoscopy is questionable, as a low interrater agreement incontrovertible will result in uncertain examinations. The difference between interrater agreement between nasal endoscopy and sinus CT scans in our study is notable. Thus, one could argue that if the clinician is in doubt whether septoplasty is indicated or not, sinus CT scan may be a helpful additional examination technique. Other helpful examination techniques to identify septal deviation are acoustic rhinometry (AR), rhinomanometry (RMM) and Peak Nasal inspiratory Flow (PNIF). Especially, PNIF is a promising tool, as it is quick to use, cheap and do not expose patients to radiation. Furthermore, studies have shown that PNIF is a sensitive test, which correlates well with patients' sino-nasal symptoms. Despite the immediate advantages, PNIF is not routinely used in Denmark. The evaluation of septal deviation in this study was performed in line with evaluation of patients in the outpatient clinic where for AR, RMM, and PNIF were not included. Furthermore, one could argue that PNIF is not capable of discriminating between nasal obstruction caused by septal deviation, polyps, edema etc. Thus, CT scans is an excellent additional examination technique in the assessment of a significant septal deviation, especially in patients with CRS. However, the effectiveness of septoplasty in general has been questioned, but a recent study by van Egmond et al. found that if patients have a septal deviation, septoplasty is more effective than non-surgical interventions and is also potentially cost-effective from a healthcare perspective. Thus, objective assessment of nasal septal deviation is essential in the preoperative planning prior septoplasty.
Evaluation of nasal septal deviation

Conclusions
Limited agreement was achieved in the endoscopic evaluation of septal deviation suggesting that nasal endoscopy is not a reliable tool as a purely objective examination technique in an interrater context. Conversely, significant agreement was achieved in the radiologic evaluation of anterior and inferior/posterior septal deviations suggesting that sinus CT scan can be a reliable objective examination technique. No significant agreement was attained in the evaluation of deviation by the perpendicular plate. Knowledge of symptoms and medical history will presumably improve the reliability of both measures. If the clinician is in doubt whether septoplasty is indicated or not, sinus CT scan may be a helpful additional examination technique. In any case, it cannot be an alternative to nasal obstruction assessment methods such as AR, RMM or PNIF.

Conflict of interest
No conflicts of interests to declare.

Ethics approval and consent to participate
The study was approved by The Regional Committees on Health Research Ethics for Southern Denmark (project-ID: 20150055) and by the Danish Data Protection Agency. All participants gave informed consent before clinical examination.

Consent for publication
Not applicable.

Availability of data and materials
The datasets used during the current study are available from the corresponding author on reasonable request.

Funding
The study was funded by The Undergraduate Research Pool, Odense University Hospital, Odense, Denmark.

Abbreviations

Acknowledgments
We thank local ENT doctors Danial Shirazi, Lotte Jung, Lotte Rosenvinge, Michael Leetmaa, Pia Juul Skov and Lars Gosvig for referring patients for the study and the secretaries in the Department of Otorhinolaryngology, Odense University Hospital, Odense, Denmark for management of referred patients.

Authorship contribution
AS performed analysis and manuscript writing. KLL performed study design and clinical examination. BL performed study design and clinical examination. PD performed clinical examination. GI performed study design. LH performed radiologic evaluation. ADK performed study design, clinical examination and radiologic evaluation.

References
16. Ottaviano G, Fokkens WJ. Measurements of nasal airflow and patency: a critical review with emphasis on the use of peak nasal


