Comparison of soft-tissue profiles in Le Fort I osteotomy patients with Dolphin and Maxilim softwares

Nasser Nadjmi,a Azita Tehranchi,b Niloufar Azami,c Bahram Saedi,d and Wouter Mollemans,e Antwerp and Leuven, Belgium, and Tehran and Qazvin, Iran

Introduction: To correct dentofacial deformities, a combination of orthodontic treatment and orthognathic surgery is needed. Prediction software packages are beneficial in treatment planning and achieving improved outcomes, but before using any software, its reliability and reproducibility must be assessed. The aim of this study was to evaluate the accuracy of 2-dimensional Dolphin (version 10; Dolphin Imaging & Management Solutions, Chatsworth, Calif) and 3-dimensional Maxilim (Medicim, Sint-Niklaas, Belgium) softwares in predicting the soft-tissue profiles of patients who had Le Fort I osteotomies.

Methods: The presurgical and postsurgical cone-beam computed tomography synthesized lateral cephalograms of 13 patients were collected. Using the Dolphin and Maxilim softwares, the postsurgical profiles were predicted. The positions of the soft-tissue landmarks in profile views were compared with landmarks in the postsurgical photographs. The data were analyzed with the coefficient of reliability and paired-sample t tests.

Results: The alpha values of the interclass correlations for each landmark in the x and y planes were between 0.96 and 0.99, except for stomion superior in Maxilim (0.83). The 95% confidence interval and the absolute mean of the error showed that errors in the Dolphin software were greater than those in the Maxilim software, but the differences were not significant (P > 0.05), except for soft-tissue A-point. The greatest errors were seen in the chin region. The prediction errors of the nasolabial and mentolabial angles were greater; the prediction error in the Dolphin software was 9°, which has clinical significance.

Conclusions: The Dolphin and Maxilim softwares are both appropriate for clinical use. Their inaccuracies in the prediction of the chin region should be considered in complicated surgical planning. (Am J Orthod Dentofacial Orthop 2013;144:654-62)
in 3 planes of space. In addition, it is feasible to generate 3D facial models by color texture mapping. Three-dimensional systems are now becoming more popular because of their potential in presenting more realistic soft-tissue predictions. However, before applying prediction software in clinical practice, qualitative and quantitative validations are required; furthermore, it is important to define the magnitude, direction, and location of errors to evaluate the clinical impact of these errors. Several studies have assessed the accuracy and reliability of 2D prediction softwares, but the results were often conflicting, and most of these inaccuracies and controversies are related to soft-tissue predictions. There are limited studies evaluating Dolphin prediction accuracy; in most of them, only hard tissues were evaluated. For 3D prediction softwares, there is even less evidence regarding their validity and reliability. Maxilim software (Medicim, Sint-Niklaas, Belgium) is a 3D image-based environment for assessing the head’s anatomy and for preoperative planning for maxillofacial surgery and visualizing the expected soft-tissue outcome. In contrast to the Dolphin software, which uses linear ratios for the soft-tissue response, the Maxilim software uses computational strategies that can model the deformation behavior of the soft tissues to predict the patient’s postoperative look. Maxilim is based on a biomechanical simulation model that does not use hard-tissue or soft-tissue movement ratios to predict; instead, it mimics the elastic deformation behavior of the soft tissues. The aim of this study was assess whether a 3D orthognathic planning package could predict the changes in the soft-tissue profile more accurately than 2D programs.

MATERIAL AND METHODS

The sample consisted of 13 white adult patients over the age of 20 years (11 women, 2 men) who had orthognathic surgery in Monica Hospital in Antwerp, Belgium, in 2008 and 2009.

The inclusion criteria were the following: (1) all patients had a Le Fort I osteotomy alone or combined with a bilateral sagittal split osteotomy; (2) all had completed their growth; (3) all underwent presurgical orthodontic treatment and orthognathic surgery by the same surgeon and with a similar technique with a computer-generated surgical splint; and (4) presurgical records were taken no more than 2 weeks before surgery, and postoperative records were taken 4 months after surgery.

Preoperative and postoperative records were taken with the same CBCT machine. To minimize the effect of head position on the facial soft tissues, the following standard protocol was used during preoperative and postoperative scanning and photography procedures: the preoperative and postoperative scans were taken with the patients in natural head position while gently biting on a thin wafer in centric relation. Natural head position was obtained by asking the patient to look at his or her image in a mirror. Then the distance between the suprasternal notch and soft-tissue pogonion was measured. We called this distance the “natural head distance” and made sure that it was respected during the preoperative and postoperative scanning procedures, and the photographs were taken at the same time in this defined position.

The exclusion criteria were the following: (1) patients with congenital deformities or syndromes, such as cleft lip and palate; and (2) other adjunctive surgical procedures, such as rhinoplasty and liposuction.

There were 6 key stages for the predictions and data gathering (all performed by 1 operator [N.A.]): (1) magnification calibration, (2) landmark identification, (3) creation of predictions by Dolphin and Maxilim softwares; (4) superimposition of the 2 predictions on a postoperative profile photograph, (5) measuring the differences in soft-tissue outlines of the superimposed images, and (6) error estimation.

For the magnification calibration in the Dolphin software, 2 points 20 mm apart were marked on the lateral cephalograms using Photoshop software (CS3; Adobe, San Jose, Calif), and magnification calibration was done in the Dolphin software.

Landmarks were identified using the Dolphin software; 15 soft-tissue and 25 hard-tissue landmarks were digitized. The soft tissue landmarks were glabella, nasal dorsum, pronasale, subnasale, soft-tissue A-point, labrale superior, stomion superior, stomion inferior, soft-tissue B-point, pogonion, gnathion, menton, soft tissue nasion, labrale inferior, and throat point.

Predictions by the Dolphin software and the Maxilim software were created. Since validation of the accuracy of soft-tissue predictions was the aim of this study, the amounts of bony movement in 2 planes (anteroposterior and vertical) had to be the same as performed by the surgeon. Before creation of the predictions by the 2 softwares, the preoperative virtually osteotomized bone fragments of each patient were moved to the actual postoperative positions by the iterative closest point algorithm to achieve the best alignment and to align the surfaces of the maxilla and the mandible in 3 dimensions. The amount of bone displacement of each virtually osteotomized bony segment in the anteroposterior and vertical planes was considered to be the actual surgical movement and used by the softwares for prediction of the soft-tissue results.

For prediction by the Dolphin software, preoperative photographs were registered on CBCT synthetized lateral.
cephalometry radiographs. This process begins with the registration of subnasale of 2 images on each other, and then the magnification and orientation of the 2 images are matched manually. Using the magnified image of superimposition, it is possible to complete the process of superimposition by finite adjustment of the soft-tissue outlines of lateral cephalometry on the photographic boundaries to obtain the best fit of the images on each other.

In the Maxilim software, the 3D photographs are aligned to the CBCT skin surface automatically, and a head model is created. Although the Maxilim software can construct 3D images, just the 2D simulated profiles were used. Table I shows details of the bony movements and surgical procedures in each patient.

The 2 predictions were superimposed on the postoperative photograph. For superimposition and quantitative validation in our study, Adobe Photoshop was used, and the postoperative and the 2 predicted images were superimposed based on the best-fit method, so that the predicted and postoperative soft-tissue images were rigidly registered on each other based on unaltered surfaces. In our study, these included the surfaces above the infraorbital rim and the radix of the nose. In addition to the best fit of unaltered surfaces, well-defined facial landmarks were digitized, such as soft-tissue nasion, sella, nasion, and right-eye outer canthus. Then superimposition of the sella-nasion line and outer canthus–soft-tissue nasion line were used additionally for more accuracy (Fig 1).

The same procedure that was used for the superimposition of the preoperative photographs on the preoperative CBCT image was applied to the superimposition of the postoperative photographs on the postoperative CBCT synthetized lateral cephalometric radiographs in the Dolphin software. These images were used as the Dolphin prediction photographs in Adobe Photoshop for quantitative validation.

We measured the differences in the soft-tissue outlines of the superimposed images. Linear measurements between each predicted image and the postoperative profile photographs (Fig 2) were measured. The differences between the computer predictions and the postoperative results were calculated for each landmark across the x-axis and y-axis. A negative value indicated that the computer prediction was more posterior in the x-axis or more superior in the y-axis compared with the postsurgical result.

The nasolabial and mentolabial angles were also measured.

Error of the method was assessed by analyzing the reliability and reproducibility of the method with the double determinant method by repeated digitizations of the landmarks.

**Statistical analysis**

All statistical analyses were carried out using SPSS software (version 11; SPSS, Chicago, Ill). Descriptive statistics, paired-sample *t* tests, and interclass correlations were used.

### Table I. Detailed amounts of bony movement and surgical procedures

<table>
<thead>
<tr>
<th>Patient</th>
<th>Maxillary advancement (mm)</th>
<th>Maxillary impaction (mm)</th>
<th>Bilateral sagittal split osteotomy (mm)</th>
<th>Genioplasty (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>3.5</td>
<td>3.5</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>5</td>
<td>5</td>
<td>3.5</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>2.5</td>
<td>2.5</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>6</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>5</td>
<td>5</td>
<td>4.5</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>-</td>
<td>4.5</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>3</td>
<td>1.5</td>
<td>4.5</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>2.5</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>
RESULTS

The error range of the Dolphin software in predicting the horizontal position of the soft-tissue landmarks was between −1.41 and 1.20 mm. The greatest errors were related to pronasale, gnathion, menton, and stomion superior. For the rest of the landmarks, there was a high coefficient of reliability between the predicted and the actual postoperative photographs (Table II).

Variability in the vertical position of the predicted landmarks was between −1.85 and 1.55 mm. The maximum vertical errors were seen in subnasale, gnathion, pogonion, and menton. The reliability coefficients were above 0.97 for each predicted landmark (Table III).

The correlations of the predicted nasolabial and mentolabial angular measurement with the actual results were 0.72 and 0.81.

The range of error for the Maxilim software in the horizontal positions of the predicted landmarks was between −1.60 and 1.50 mm. The greatest errors in horizontal positions were observed in menton, gnathion, and pogonion (Table IV). In the vertical direction, the variability ranged between −4.25 and 2.42 mm. The
positions of the predicted landmarks in the Maxilim software were highly correlated to the actual results in both the horizontal and vertical dimensions. The weakest correlation was observed between the vertical positions of stomion superior, at 0.83 (Table V).

For the nasolabial and mentolabial angles, the correlations between the predicted and actual results were 0.95 and 0.92, respectively.

There were no statistical differences between the absolute prediction errors of the Dolphin and Maxilim software.
softwares and the actual results in relation to the horizontal and vertical positions of the landmarks ($P > 0.05$), except for the vertical position of soft-tissue A-point in the Maxilim simulated profile.

The 95% confidence interval (CI) showed that for the nose and lower lip the errors of the 2 prediction programs had limited clinical significance (less than 1 mm). The vertical positions of soft-tissue A-point and stomion superior had the greatest errors associated with the Maxilim prediction program. Gnathion and menton showed the greatest errors in the predictions, but there were no significant differences regarding the mean errors of each prediction program.

For the nasolabial and mentolabial angles, the differences of absolute errors in Dolphin and Maxilim were not significant ($P > 0.05$). For the 95% CI, the error in the nasolabial angle was larger with the Dolphin software (9°) relative to the Maxilim software (3°). For the mentolabial angle, the errors were 9° in Dolphin and 6.8° in Maxilim.

**DISCUSSION**

The aim of this study was to compare the postsurgical soft-tissue profiles of the patients who had Le Fort I osteotomy alone or combined with bilateral sagittal split osteotomy with the profiles predicted by the Dolphin and Maxilim softwares.

Our study showed that the postoperative positions of the soft-tissue points were highly correlated with the predicted ones in both softwares, and that there were no significant differences in errors between the predictions and the real surgical changes. The vertical and horizontal positions of the nose, upper lip, and lower lip landmarks could be produced with Dolphin and Maxilim at an accuracy of 70%; in the chin region, the accuracy was 50%. The largest systematic difference between the predicted and real changes was observed in the chin region with both softwares. These findings were consistent with the results of Loh et al., and it was assumed that this occurred because of the different amounts of mandibular autorotation between the predicted and actual results after maxillary impaction, caused by inaccuracies in locating the center of the mandibular rotational axis. The complexity of predicting in the chin region has been reported by Xia et al. because of multiple directional movements of the soft tissues in the chin region. It seems that the chin is the most difficult region to predict; however, as Magro-Filho et al. mentioned, it seems that Dolphin is more accurate in predictions of the chin area and the base of the mandible than is Dentofacial Planner. Some previous studies reported the greatest inaccuracies in the predictions of the lips, although in our study the position of the lips was predicted accurately, and the largest errors in the 95% CI in both softwares were less than 0.5 mm, with the exception of 2 points (vertical positions of labrale superior and stomion superior in Maxilim). In our subjects, the Dolphin software did not predict the changes in the shape of the lips after surgery. For example, increased show of vermilion and some other aspects such as lip lengthening and competency were not predicted and should be adjusted by the operator. In 1 patient who had an increased show of vermilion after maxillary advancement, the Dolphin prediction of the lip shape was distorted because it could not predict the changes in volume and thickness of the lips. Konstiantos et al. also found that although there were high correlations between the predicted lip landmarks and the postsurgical results, the shape of the lips was not predicted accurately. This limitation was observed not only in Class II subjects, but also in Class III subjects because the upper lip was trapped by the mandibular incisors. Differences in lip tonicity, length, posture, and mass between patients result in more inaccuracies in software that use preprogrammed fixed hard-tissue to soft-tissue ratios for prediction of the soft-tissue results. These limitations did not exist with the predictions of Maxilim, since it is based on tensor model analysis and can predict rolling the lips as free points on the teeth during simulation; it also predicts the changes in lip thickness and show of vermilion more accurately.

Although the mean errors of the vertical predictions of A-point and stomion superior in Maxilim were relatively small (−0.55 and −0.89 mm, respectively), the ranges were huge: from −4.25 to 2.42 mm, and −4.10 to 0.52 mm, respectively (Table V). This discrepancy also is expressed in the 95% CI values of these points: 2.22 mm for A-point and 2.67 mm for stomion superior; these are much larger than the values for the other points. The 95% CI was less than 0.5 mm (clinically acceptable error) for all other points in the horizontal and vertical positions. It seems that for these 2 points the range and the 95% CI reflect the discrepancy better than the mean error.

Both softwares were accurate in the prediction of nose landmarks. The systematic errors were insignificant, and inaccuracies in the 95% CI were less than 0.45 mm, with the exception of the vertical position of subnasale in Dolphin (0.64 mm), and clinically acceptable; the prediction accuracy was higher than in other softwares in similar studies.

The analysis of angular measurements showed that the amount of systematic errors and the prevalence of significant clinical errors were increased in both softwares, especially in Dolphin. It seems that this occurred...
because the errors in different landmarks of an angle were not in the same direction. If landmark prediction in 1 angle arm was overpredicted and in the other arm was underpredicted, then the sum of errors in 2 direction leads to increased inaccuracy of angle prediction. Kaipatur and Flores-Mir,25 in a systematic review of prediction computer programs stated that although differences larger than 2 mm are important, when small inaccuracies are added in each plane of space, the combined error might have clinical implications. The errors in the Dolphin predictions were 3- and 1.4-fold higher than in Maxilim for the nasolabial and mentolabial angles, respectively, and 9° for each angle in the 95% CI. De Lira Ade et al23 reported errors of 3.7° and 4.8°, respectively, in 95% CI values for the nasolabial and mentolabial angles with Dolphin predictions. In our study, these angles were predicted less accurately, but there were still acceptable correlations with the actual results, and the differences of systematic errors were insignificant between the 2 softwares. Eales et al20 used the Student t test to compare the prediction of angles by COG software (version 3.4) and found significant differences for nasolabial predictions. Using the mentioned standard error in that study, we calculated the 95% CI for nasolabial angle error; it showed that the error was 11°, comparable with the Dolphin prediction error but 4 times more than the Maxilim error. In a comparison of Dolphin with Dentofacial Planner, Magro-Filho et al22 noted that Dolphin was better for predicting the nasolabial angle.

There were some differences between our results compared with other studies that used other 2D softwares. Eales et al,20 Loh et al,10 Konstantios et al,2 and Donatsky et al26,27 have shown significant differences in lower lip prediction errors rather than the chin area. One explanation for these differences might be that different statistical analyses were used in our study in comparison with the others. Loh et al,10 Konstantios et al,2 and Donatsky et al26,27 compared the mean differences and used Wilcoxon and t tests. Jones et al24 compared the medians to overcome the problems of overestimation and underestimations compensations that might occur in mean comparisons. In our study, the differences of landmark positions between the 2 softwares and the actual positions were compared by the intraclass correlation coefficient. The t test was used for comparison of systematic errors of the 2 softwares. As Power and Hillard17 explained, significant statistical differences in the mean comparisons do not imply clinical significance because in the assessment of any software some random error is acceptable. In addition, other factors were considered in our study.

There was a wide range of disparity between studies in the timing of postsurgical assessments; some were taken immediately after surgery,10,26,27 and in 1 study during the first year after surgery.2 In our study, all postsurgical records were taken 4 months after surgery so that soft-tissue edema was minimized. An important point is that soft-tissue changes occurred even after the first year after surgery as a result of orthodontic tooth movement, relapse, or bone remodeling, so taking the records of all patients at the same time and before orthodontic tooth movement is essential.31 Other variables are impossible to control and make the surgical soft-tissue predictions difficult; eg, weight change from presurgery to postsurgery, parasthesia of the upper lip, and muscle pose can affect the results.

A number of studies that had evaluated other softwares did not consider factors such as uniformity of the samples, accuracy of the surgery, an accurate method to measure the surgical movements, or the effects of preoperative or postoperative orthodontic treatment and relapse.2,12,20 Like Jones et al,24 we mentioned that to measure the changes and assess the accuracy of surgical procedures, presurgical and postsurgical lateral cephalograms were superimposed on the sella-nasion line rather than on more variable landmarks such as overjet, overbite, or molar cusps, which were used in other studies.1,2,20 Superimpositions were also done along maxillary and mandibular bases to consider the probability of dental movements after surgery. Furthermore, in our study, CBCT synthesized lateral cephalograms were used; they had the advantage of better soft-tissue visualization and landmark identification over conventional lateral cephalograms.

Since the results showed no major systematic errors in soft-tissue predictions by these softwares, except in the chin region and as mentioned in a recent systematic review of computer program accuracy, the studied softwares gave accurate predictions with errors of less than 2 mm, except for some points in the lower third of the face.25 The most important differences between the 2 software packages were in their soft-tissue algorithms; predictions of lips and profiles; qualitative image characteristics; ability to refine images; cephalograms-photographs linking; magnification calibration; image enhancement options; and compatibility with different operating systems.16,21,28,32

Although both softwares were accurate in predictions of profiles, Maxilim has the third dimension, which makes the quantification of volumetric changes and predictions in the transverse plane possible. Maxilim’s dimensional ability is helpful, especially in complicated orthognathic surgery cases such as cleft lip and palate, and asymmetric face. However, prediction planning of the third dimention was beyond the scope of this study.
A possible limitation of this study was the few subjects (6) with Le Fort advancement surgery. In addition, the degree of anteroposterior advancement was relatively small, and future studies will be needed to validate the predictive capacity of these softwares on larger anteroposterior surgical movements.

CONCLUSIONS

This study shows that both Dolphin and Maxilim softwares are accurate in predictions of the soft-tissue profile and can be used for planning and prediction in orthognathic surgery. There are inaccuracies in predictions of the chin region that should be considered in complicated surgical planning.

ACKNOWLEDGMENTS

This paper has been entirely drawn up from a postgraduate Thesis, which was completed by Dr Niloufar Azami under the supervision of Dr Nasser Nadimi and Dr Azita Tehranchi with the cooperation of the Orthodontic Department, Dental School, Shahid Beheshti Medical Sciences University, Tehran, Iran.

REFERENCES

27. Donatsky O, Bjorn-Jorgensen J, Hermund NU, Nielsen H, Holmqvist-Larsen M, Nderer PH. Immediate postoperative outcome of orthognathic surgical planning, and prediction of positional changes.


