Cephalometric evaluation of the predictability of bimaxillary surgical-orthodontic treatment outcomes in long face pattern patients: A retrospective study

Carla Maria Melleiro Gimenez¹, Francisco Antonio Bertoz², Marisa Aparecida Cabrini Gabrielli², Oswaldo Magro Filho³, Idelmo Garcia³, Valfrido Antonio Pereira Filho³

Objective: The aim of this study was to compare by means of McNamara as well as Legan and Burstone’s cephalometric analyses, both manual and digitized (by Dentofacial Planner Plus and Dolphin Image software) prediction tracings to post-surgical results. Methods: Pre and post-surgical teleradiographs (6 months) of 25 long face patients subjected to combined orthognathic surgery were selected. Manual and computerized prediction tracings of each patient were performed and cephalometrically compared to post-surgical outcomes. This protocol was repeated in order to evaluate the method error and statistical evaluation was conducted by means of analysis of variance and Tukey’s test. Results: A higher frequency of cephalometric variables, which were not statistically different from the actual post-surgical results for the manual method, was observed. It was followed by DFPlus and Dolphin software; in which similar cephalometric values for most variables were observed. Conclusion: It was concluded that the manual method seemed more reliable, although the predictability of the evaluated methods (computerized and manual) proved to be reasonably satisfactory and similar.

Keywords: Corrective orthodontics. Predictive value of tests. Oral surgery. Face.

Objetivos: a proposta desse trabalho foi comparar, por meio das análises cefalométricas de McNamara, Legan e Burstone, os traçados de previsão manuais e os digitalizados pelos programas Dentofacial Planner Plus e Dolphin Imaging, com os resultados pós-cirúrgicos. Métodos: foram selecionadas as telerradiografias pré- e pós-cirúrgicas (seis meses) de 25 pacientes face longa submetidos a cirurgia ortognática combinada. Foram realizados os traçados de previsão manual e computadorizados de cada paciente, comparando-os, cefalométricamente, com os resultados pós-cirúrgicos. Esse protocolo foi repetido para avaliação do erro do método e realizou-se a avaliação estatística por meio da análise de variância e sobre teste de Tukey. Resultados: observou-se uma maior frequência de variáveis cefalométricas que não diferiram estaticisticamente do resultado pós-cirúrgico real para o método manual, seguido dos programas DFPLUs e Dolphin; observando-se valores cefalométricos similares para a maioria das variáveis. Conclusão: concluiu-se que o método manual pareceu mais fidedigno, embora a previsibilidade dos métodos avaliados (computadorizados e manual) tenha se mostrado razoavelmente satisfatória e similar.

INTRODUCTION

Accurate diagnosis of both structures involved in malocclusions and the severity of occlusal conditions, facial and functional, leads to the decision of surgical-orthodontic treatment. It is extremely important to carry out an individualized plan in order to obtain successful and consistent results. Initially, preparatory orthodontic procedures are performed prior to orthognathic surgery, aiming to position the teeth in their bone bases. At the latest stage of ideal arch placement, molding is performed in order to articulate the upper and lower models in Class I relationship. Should this relationship be suitable in sagittal, transverse and vertical dimensions, the patient is conducted to the oral and maxillofacial surgeon who will perform the surgical planning in all its particularities.1

The prediction tracings show the inclination of the incisors and anticipate all necessary surgical movements, providing visualization of potential results from the tangent to the soft tissue as well as from the tangent to the skeletal tissue. Based on these data, model surgery is performed in semi-adjustable articulator, in which the information concerning the prediction tracing is transferred. This phase will accurately determine both the magnitude and the direction of surgical movements performed to obtain proper occlusion. Afterwards, surgery is performed, followed by orthodontic finishing, removal of orthodontic appliance, placing of retention and post-retention monitoring phases.

This paper focuses on the phase of prediction tracings which is important for carrying out proper surgical planning as well as for guiding the patient and establishing communication with him. Reliability of the proposed result is a constant concern. These tracings create a situation in which it is possible to describe in detail all surgical alterations, leading to an optimized conduct of the case. 13 Conventionally, these tracings are manually carried out, however, there are computer software that perform the prediction of results based on the digitization of cephalometric teleradiographs landmarks, for instance: Dentofacial Planner, OPAL, Quick Ceph Image, COGsoft, TIOPS, Dolphin.1,6 These software are able to simulate the effect of incisor decompensation and the resultant movements of bony bases, translating them into illustrations and providing a silhouette of post-surgical skeletal and soft tissue profiles. However, it is worth noting that the prediction of these profile changes is difficult due to the variability of soft tissue behavior and differences in their translation accompanying skeletal changes promoted by orthognathic surgery.17

The present study is set within this context, with the purpose of comparing, through cephalometry, the accuracy of manual prediction tracings as well as those performed by both Dentofacial Planner Plus and Dolphin Image software, in relation to post-surgical results of long face patients subjected to bimaxillary orthognathic surgery.

MATERIAL AND METHODS

This study was approved by UNESP/Araçatuba Institutional Review Board, and analyzed pre and post-surgical lateral teleradiographs (6 months) of a sample comprised of 25 adult, long face patients, Angle’s Class II, who were subjected to combined orthognathic surgery. These teleradiographs were obtained from the Center for Research and Treatment of Buccofacial Deformities (CEDEFACE – Araraquara/SP) and from the Oral and Maxillofacial Surgery and Traumatology course, given at UNESP College of Dentistry — Araraquara.

The inclusion criteria were:

1) Leucoderm, Brazilian, dentate and adult patients of both sexes.
2) Hyperdivergent patients.
3) Class II dental relationship, showing no open bite.
4) Bimaxillary surgical-orthodontic treatment performed in a minimum period of 6 months before the research, as this period assures sufficient regression of edema caused by surgery.
5) Radiographic documentation of initial periods, immediate preoperative phase, and well-performed post-surgical phase.
6) Absence of pathologies, fissures, facial anomalies or asymmetries.
7) Conventional Edgewise orthodontic technique was used for orthodontic preparation.
8) Absence of any other cosmetic and reconstructive surgeries performed on the patient’s face during or after surgical-orthodontic treatment.

The sample comprised 22 (88 %) females and 3 (12 %) males with a mean age of 32.24 years (17 to 45 years).

Each teleradiography was traced three times (alternately and at weekly intervals in order to avoid memorization of traces), at an environment with controlled
lighting (dark room). Seventy one landmarks were marked as they are necessary for the digitalization process performed with the Dentofacial Planner Plus software. The last trace of each patient, regarding pre and post-surgical teleradiographs (6 months), was used as a guide for an organized and sequential scan of these cephalometric landmarks. A scan of each trace was repeated twice in order to evaluate the method reproducibility (Intraexaminer Method Error).

Dentofacial ShowCase 2.0 for Microsoft Windows 95 and for Microsoft Windows NT 4.0 was used to scan the preoperative tracing, the prediction tracing obtained from this software as well as manual prediction tracing and post-surgical tracing. Dolphin Imaging 10.5 software (in the case of prediction tracings performed by this program) was used for the cephalometric evaluation, in which a single examiner was calibrated for tracing and digitizing the cephalograms.

Prediction tracings were built manually as well as with the use of Dentofacial Planner Plus (DFPlus) and Dolphin Imaging 10.5 (Dolphin) software. This process was based on data obtained from the surgery that was performed (clinical records in file folders of patients). The following sequence of tracings was obtained for each patient: manual prediction, DFPlus prediction, Dolphin prediction and post-surgical tracing (actual result). This sequence of tracings was subjected to McNamara’s as well as Legan and Burstone’s analyses for cephalometric evaluation.

The results (linear and angular measures obtained from the cephalometric analyses) were tabulated into Excel. A cephalometric analysis on the results obtained from manual and computerized prediction tracings as well as post-surgical tracings was carried out in order to check prediction error. This comparison was developed in three steps:

1) Evaluation carried out with Student’s t test (paired) to determine whether or not there was prediction error statistically different from zero (for each cephalometric analysis measure cited for all prediction methods). The prediction error was given by subtracting post-surgical actual result from the value of cephalometric prediction.

2) Analysis of variance for comparison between post-surgical tracings and manual and computerized prediction tracings.

3) Tukey’s test.

RESULTS

Method error

The method error analysis was indicated due to the importance of carrying out critical evaluation to verify the possibility of reproducibility as well as the effectiveness of the methodology used. It should be noted that the data was read twice and that there was mutual agreement between readings, which proves the procedures to be reliable.

Table 1 - Means (M) and standard deviation (SD) of McNamara’s cephalometric post-surgical measures and prediction errors means (M.E.) with their standard deviation (SD), according to the prediction methodology used.

<table>
<thead>
<tr>
<th>Cephalometric measurement</th>
<th>Post-surgical</th>
<th>DFPlus</th>
<th>Dolphin</th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-Nperp</td>
<td>6.5</td>
<td>4.8</td>
<td>-6.6</td>
<td>3.9*</td>
</tr>
<tr>
<td>Co-Gn</td>
<td>124.3</td>
<td>6.5</td>
<td>-23</td>
<td>4.4**</td>
</tr>
<tr>
<td>CO-A</td>
<td>96.1</td>
<td>6.7</td>
<td>-74</td>
<td>4.5*</td>
</tr>
<tr>
<td>Dif. Mx-Md</td>
<td>28.3</td>
<td>4.2</td>
<td>5.2</td>
<td>51.1*</td>
</tr>
<tr>
<td>Al/FH</td>
<td>77.7</td>
<td>5.5</td>
<td>21</td>
<td>3.9*</td>
</tr>
<tr>
<td>Pg-Nperp</td>
<td>0.1</td>
<td>6.3</td>
<td>-4.0</td>
<td>4.8**</td>
</tr>
<tr>
<td>1-A perp</td>
<td>2.5</td>
<td>3.3</td>
<td>2.8</td>
<td>3.4*</td>
</tr>
<tr>
<td>1-A-Pg</td>
<td>1.8</td>
<td>3.1</td>
<td>3.5</td>
<td>2.0*</td>
</tr>
<tr>
<td>FMA</td>
<td>30.6</td>
<td>5.0</td>
<td>-1.8</td>
<td>3.6*</td>
</tr>
<tr>
<td>Facial Axis</td>
<td>87.3</td>
<td>4.1</td>
<td>-1.1</td>
<td>3.6*</td>
</tr>
<tr>
<td>Nasolabial Ang.</td>
<td>100.7</td>
<td>12.2</td>
<td>0.7</td>
<td>13.3*</td>
</tr>
</tbody>
</table>

Means with the same letters in a row are not significantly different for Tukey’s test at 5%.
* Means significantly different from zero for Student’s t test at 5%.
** Bold: Measures of which prediction error was statistically equal to zero (actual result similar to the prediction).
Underline: Measures with lower prediction error among the three evaluated methods.
Cephalometric evaluation of the predictability of bimaxillary surgical-orthodontic treatment outcomes in long face pattern patients: A retrospective study

when using the manual method (5 cephalometric variables), followed by the DFPlus computerized method (2 cephalometric variables) and Dolphin (only one cephalometric variable).

Analyses of variance were used to compare the three methods concerning the prediction error. When such analyses indicated significant difference between methods, Tukey’s test was used for multiple comparisons of means. Despite presenting statistically significant differences, the following measures were close to the actual result: 1-APg (manual method), Pg-Nperp (DFPlus), CO-A (manual method and Dolphin), CO-Gen (DFPlus), A-Nperp (Dolphin). In this case, the three methods (Manual method, DFPlus and Dolphin) presented the same frequency (2 cephalometric variables for each method).

Results also demonstrated that some cephalometric measures showed very close values between two methods: A-Nperp (manual method and DFPlus), CO-A (manual method and Dolphin), CO-Gen (DFPlus), A-Nperp (Dolphin). In this case, the three methods (Manual method, DFPlus and Dolphin) presented the same frequency (2 cephalometric variables for each method).

The results displayed in Table 2 refer to the cephalometric measures established by Legan and Burstone.9 From a total of 37 cephalometric means evaluated by Student’s t test, 14 were not significantly different

**DISCUSSION**

Each measure regarding the selected analyses was systematically evaluated from a cephalometric point of view9,12 in order to relate the prediction error: the difference between the actual post-surgical measure and the prediction measure of each method (DentoFacial Planner Plus, Dolphin Imaging, manual prediction tracing). If the difference was zero, it would mean that the prediction would have been identical to the actual post-surgical results, indicating an excellent degree of accuracy concerning predictability. Therefore, the closer the cephalometric measures are to zero, the more accurate the method of predictability. Moreover, positive or negative prediction error indicate that the predicted value is, respectively, higher or lower than the actual value.

In accordance with Student’s *t* test, the results shown in Table 1 (McNamara Jr Cephalometric Analysis,12) demonstrate that only 8 out of 33 evaluated means were not significantly different from zero; i.e., only eight presented a prediction result that did not differ statistically from the actual final result. These measures were: Maxillomandibular difference (manual method), AIFH (manual method), 1-Aperp (manual method), FMA (manual method and Dolphin), Facial axis (DFPlus), nasolabial angle (manual method and DFPlus). Therefore, the predictions which did not differ statistically from the actual post-surgical result were more frequent

<table>
<thead>
<tr>
<th>Cephalometric measurement</th>
<th>Post-surgical</th>
<th>DFPlus</th>
<th>Dolphin</th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sn-G Vert</td>
<td>10.7</td>
<td>4.9</td>
<td>-4.0</td>
<td>2.6**</td>
</tr>
<tr>
<td>Pg-G Vert</td>
<td>4.0</td>
<td>7.1</td>
<td>-4.2</td>
<td>5.4**</td>
</tr>
<tr>
<td>Middle third</td>
<td>677</td>
<td>56</td>
<td>0.9</td>
<td>3.8**</td>
</tr>
<tr>
<td>Lower third</td>
<td>74.4</td>
<td>60</td>
<td>-11</td>
<td>2.9**</td>
</tr>
<tr>
<td>Upper Lip Protrusion</td>
<td>5.2</td>
<td>2.3</td>
<td>-11</td>
<td>1.8**</td>
</tr>
<tr>
<td>Lower Lip Protrusion</td>
<td>3.2</td>
<td>3.4</td>
<td>-0.3</td>
<td>3.5**</td>
</tr>
<tr>
<td>Upper Incisor Exposure</td>
<td>3.5</td>
<td>3.7</td>
<td>-0.6</td>
<td>2.2**</td>
</tr>
<tr>
<td>Interlabial Space</td>
<td>3.6</td>
<td>3.7</td>
<td>-2.7</td>
<td>2.7**</td>
</tr>
<tr>
<td>Facial Convexity</td>
<td>154</td>
<td>5.3</td>
<td>-2.7</td>
<td>4.5**</td>
</tr>
<tr>
<td>Cervico-mental angle</td>
<td>75.4</td>
<td>9.6</td>
<td>-2.0</td>
<td>10.2**</td>
</tr>
<tr>
<td>Nasolabial angle</td>
<td>100.7</td>
<td>12.2</td>
<td>0.7</td>
<td>13.3**</td>
</tr>
<tr>
<td>Mentolabial sulcus</td>
<td>5.4</td>
<td>14</td>
<td>-1.6</td>
<td>1.9**</td>
</tr>
<tr>
<td>Medium / inf.%</td>
<td>91.5</td>
<td>9.4</td>
<td>2.8</td>
<td>6.7**</td>
</tr>
</tbody>
</table>

Means with the same letters in a row are not significantly different for Tukey’s test at 5%.

* Means significantly different from zero for Student’s *t* test at 5%.

» Bold: Measures of which prediction error was statistically equal to zero (actual result similar to the prediction)

» Underline: Measures with lower prediction error among the three evaluated methods.

NOTE: The averages of the middle and lower facial thirds are not provided by Dolphin Imaging 10.5 software.

### Table 2 - Means (M) and standard deviation (SD) of Legan Burstone’s cephalometric post-surgical measures, linear (mm) and angular (degrees), and prediction error means (E.M.) with their standard deviation, according to the prediction methodology used.
from zero, i.e., not statistically different from the actual post-surgical result, which indicates a good level of prediction. These cephalometric references are: middle third of the face (DFPlus and manual method), lower third of the face (manusal method), upper lip protrusion (manual method), lower lip protrusion (DFPlus and manual method), upper incisor exposure (Dolphin and DFPlus), interlabial space (Dolphin), facial convexity (Dolphin), cervico-mental angle (DFPlus), nasolabial angle (DFPlus and manual method) proportion of the facial thirds - Medium/Inf.% (manual method). The manual method had the highest number of cephalometric variables (6) with no statistically significant difference concerning the actual post-surgical result; followed by DFPlus (5) and Dolphin (3).

Except for the last two cephalometric measures, the means which significantly differ from zero, regarding a single measure, present equal signs: positive or negative, thus, demonstrating the same behavior towards cephalometric prediction. Analyses of variance were used to compare the three methods concerning the prediction error. Such analyses were complemented by Tukey’s test.

Despite presenting statistically significant differences, the following measures were close to the actual result: Sn-G Vert (DFPlus), Pg-G Vert (DFPlus), mentolabial sulcus (manual method). In this regard, the prediction of DFPlus software (2 cephalometric variables) was more often closer to the actual result.

Results also demonstrate that some cephalometric measures showed very close values between two methods: Sn-G Vert (manual method and DFPlus), upper incisor exposure (DFPlus and Dolphin), facial convexity (DFPlus and Dolphin), cervico-mental angle (manual method and Dolphin), nasolabial angle (DFPlus and manual method), Medium / inf.% (DFPlus and manual method). Therefore, manual and DFPlus methods proved to have a higher frequency of agreement, followed by DFPlus and Dolphin. The study of Power et al.16 corroborates the findings of the present study. These authors compared the accuracy of prediction using Dolphin Imaging Software (Version 8.0) and the traditional manual technique; in comparison with actual post-surgical results. Manual tracings proved to be more predictable. The comparison of actual results to the predictions of the software demonstrated clinically significant differences for all measures.

Similarly, in the study carried out by Chunmaneecho and Friedman,2 prediction proves to be higher in the manual method. The authors concluded that conventional prediction tracings were significantly closer to post-surgical results than the pre-programmed proportions ($p < 0.05$).

Smith, Thomas and Proffit18 evaluated five software (Dentofacial Planner Plus, Dolphin Imaging, Orthoplan, Quick Ceph Image and Vistadent) and their differences in the ability to simulate results in orthognathic surgery. Dentofacial Planner Plus software was considered the best simulator. The results showed that the differences in the ability to simulate results depend on several factors, such as: software performance, easy to use, cost, compatibility, image quality and practical application of available resources. In the present study, the frequency of cephalometric variables that were closer to post-surgical results proved to be higher when using DFPlus software in comparison to Dolphing Imaging software, corroborating the aforementioned authors.

Furthermore, Schultes et al.17 also claim that the prediction of DFPlus software is appropriate. The authors found that the software was frequently in accordance with the real situation regarding the nasal and labial areas, while the highest margins of errors were seen in the submental region. In general, predictability was greater than 80%, which ensures accurate planning.

Csazar, Bruker-Csaszar and Niederllmann5 also evaluated the accuracy of prediction of DFPlus software and concluded that this proves to be satisfactory, although the profile of the labial region presents difficulties of predictability, which indicates the need for further development of this software.

In the study carried out by Gosset et al.7 which compared the traditional prediction tracings (manual method) and the Dolphin Imaging software tracings with actual post-surgical results, it was shown that seven out of the sixteen measures showed statistically significant differences for the conventional method, while nine measures were statistically significant different for Dolphin Imaging software. Based on these data, the authors concluded that both methods seem to demonstrate reasonable predictability, thus, being similarly accurate. This statement corroborates the findings of this study, since the evaluated methods also demonstrated reasonable predictability, the differences between them were slight and the degree of accuracy observed was similar.
Claiming that it is necessary to have common sense when using any system regarding prediction of results, Lu et al.\textsuperscript{10} emphasized that although computerized imaging systems are valuable for establishing communication with patients as well as giving explanations to them, further efforts are needed to improve their accuracy, including considerations related to the stress of soft tissue and quality of muscle tissue. Therefore one must consider that the prediction obtained by imaging systems must be carefully interpreted.

Facing the possibility of high individual variability, Cousley et al\textsuperscript{4} compared post-surgical results with prediction tracings and noted that the predictions of some cephalometric variables were reasonably accurate in terms of mean values. However, there was high individual variation for most measures, with the presence of systematic error. In particular, there was a tendency toward orthognathic surgery “overprediction”, inducing backward mandibular rotation.

Pektas et al\textsuperscript{15} evaluated the accuracy of predictability of tegumental tissue response resulting from surgical-orthodontic treatment and observed that in the sagittal plane, the tip of the nose was the most accurate area, while the upper lip area presented the highest level of difference. In the vertical plane, the subnasal area was the most accurate, while the lower lip area was the least precise. The authors suggested that the predictions in sagittal plane were superior to those in the vertical plane.

Several authors suggest that variability factors related to the soft tissue (such as thickness, soft tissue tonicity, shape, functional application, and free functional space) be carefully considered when interpreting the predictability within the context of surgical-orthodontic treatment.\textsuperscript{5,6,14}

Statistical analysis of the method error proved that the methodology used for marking cephalometric landmarks as well as angular, linear and proportion measures, was reliable and reproducible, which confirms the considerations of Buschang et al\textsuperscript{2}, Martins\textsuperscript{11} and Trajano et al.\textsuperscript{19}

**CONCLUSION**

The experimental conditions of this study show that:
- The manual method proved to be closer to the cephalometric variables evaluated in relation to the actual post-surgical results, followed by DFPlus and Dolphin.

**REFERENCES**