Accuracy of the Computer-Aided Surgical Simulation (CASS) System in the Treatment of Patients With Complex Craniomaxillofacial Deformity: A Pilot Study

James J. Xia, MD, PhD,* Jaime Gateno, DDS, MD,†
John F. Teichgraeber, MD,‡ Andrew M. Christensen, BS,§
Robert E. Lasky, PhD,‖ Jeremy J. Lemoine, BS,¶
and Michael A.K. Liebschner, PhD#

Purpose: Current surgical planning methods are usually not adequate for the treatment of patients with complex craniomaxillofacial (CMF) deformities. To this end, we have developed a 3-dimensional (3D) computer-aided surgical simulation (CASS) planning method for the treatment of patients with complex CMF deformities. The purpose of this pilot study was to evaluate the accuracy of this technique in the treatment of patients with complex CMF deformities.

Patients and Methods: Five patients with complex CMF deformities were enrolled. Surgeries were planned with the CASS planning method. Surgical plans were transferred to patients at the time of surgery via computer-generated splints. After surgery, outcome evaluation was completed by first superimposing the postoperative computed tomography (CT) model onto the planned model, and then measuring the differences between planned and actual outcomes. The criteria used to determine the accuracy of the technique were as follows: a linear difference between planned and actual outcomes of less than 2 mm, and an angular difference of less than 4°.

Results: All patients underwent surgery as planned. With the use of CASS planning, medians of the differences between planned and actual postoperative outcomes were limited to 0.9 mm and 1.7°.

Conclusion: The results of this pilot study are promising. They will be used as the basis of calculations needed to determine the sample size for a larger and more comprehensive study that will be undertaken to assess the accuracy of CASS planning methods.

The success of craniomaxillofacial (CMF) surgery depends not only on the surgical techniques used, but also on the accuracy of the surgical plan. However, significant problems have been identified in the use of currently available surgical planning methods.¹ Surgeons usually complete a physical examination and order cephalometric radiographs, plaster dental models, 3-dimensional (3D) computed tomography (CT) scans, and

*Director, Surgical Planning Laboratory, Department of Oral and Maxillofacial Surgery, The Methodist Hospital Research Institute, Houston, TX.
†Chairman, Department of Oral and Maxillofacial Surgery, The Methodist Hospital Research Institute, Houston, TX.
‡Professor, Division of Pediatric Plastic Surgery, Department of Surgery, University of Texas Houston Health Science Center, Houston, TX.
§President and Chief Technical Officer, Medical Modeling LLC, Golden, CO.
‖Professor, Departments of Pediatrics and Obstetrics and Gynecology, University of Texas Houston Health Science Center, Houston, TX.
¶Doctoral Student, Department of Bioengineering, Rice University, Houston, TX.
#Assistant Professor, Department of Bioengineering, Rice University, Houston, TX.

This work was partially supported by NIH/NIDCR grant 1 R41 DE016171-01, UT-Houston Comprehensive Research Training Program 5 T32 DE015355-01 (NIH), and UT-Houston Medical School UCRC grant M01 RR052558 (NIH).

Address correspondence and reprint requests to Dr Xia: Department of Oral and Maxillofacial Surgery, The Methodist Hospital Research Institute, 6560 Fannin Street, Suite 1228, Houston, TX 77025; e-mail: jxia@tmh.tmc.edu

© 2007 American Association of Oral and Maxillofacial Surgeons

doi:10.1016/j.joms.2006.10.005

248
sometimes CT-based 3D physical models. The cephalometric radiographs are used to perform a cephalometric analysis and to simulate the operation. The drawback of this technique is that it is 2-dimensional (2D) and cannot be used to plan surgery for patients with 3D problems.\(^2\),\(^3\) Plaster dental models are also used to simulate the operation (dental model surgery). A limitation of this technique is that it does not portray the surrounding bony anatomy, making it impossible for surgeons to visualize bony changes that may occur during surgery.\(^4\),\(^5\) Although 3D CT scans provide excellent visualization of the craniofacial skeleton, they have not been used successfully to simulate surgery because CT does not render the teeth with the accuracy necessary for surgical simulation.\(^4\),\(^6\),\(^7\) Three-dimensional CT-based physical models are generally extremely useful for surgical simulation, but unfortunately, as with 3D CT, the teeth on these models are not accurate enough for CMF surgical simulation.

To eliminate some of these problems, our Surgical Planning Laboratory has developed a 3D computer-aided surgical simulation (CASS) planning system. Although the clinical feasibility of the CASS method has been shown,\(^8\) the accuracy has not yet been quantitatively evaluated. Therefore, the purpose of this pilot study was to evaluate the accuracy of this technique in the treatment of patients with complex CMF deformities.

**Patients and Methods**

A total of 5 consecutive patients with complex CMF deformities who presented to Texas Cleft and Craniofacial Clinic between January 2004 and May 2005 were included in the study.\(^8\) Inclusion criteria were as follows: 1) Patients who were scheduled to undergo CMF surgery to treat one of the following congenital or acquired deformities: dentofacial deformity requiring double jaw surgery; craniofacial skeletal deformity, including hemifacial microsomia and craniofacial syndromes; post-traumatic craniofacial defects; craniofacial defects after tumor ablation; and deformities of the temporomandibular joint; 2) those who were scheduled to undergo CT scanning as part of their treatment; and 3) patients who agreed to participate in this study. Institutional Review Board approvals for this study were obtained through The University of Texas Houston Health Science Center and The Methodist Hospital Research Institute. Informed consent forms were signed by patients and collected by investigators.

Surgery was planned according to the authors’ CASS planning method. CT scanning was performed. The first step in the planning process was to create a composite skull model that reproduced the bony structures and dentition with a high degree of accuracy.\(^4\),\(^6\) The second step was to quantify the deformity. In particular, the anteroposterior, vertical, and transverse positions of the maxilla were measured. The third step was to simulate the entire surgical procedure in the computer. Maxillary osteotomy was usually completed first, followed by mandibular and chin surgeries. The shape and size of the bone graft, if needed, were also simulated. If simulated outcomes were not satisfactory, the surgical plan could be modified, and simulation could begin again. The final step was to create surgical splints. With the authors’ computer-aided designing/computer-aided manufacturing techniques, surgical splints were designed in the computer and were fabricated with a stereolithographic apparatus.\(^6\),\(^9\) The surgical plan was transferred to the patient at the time of surgery.

To minimize potential risks to patients, a backup surgical plan was created in accordance with standard clinical routine with the use of cephalometric analysis, prediction tracings, plaster dental model surgery, 3D CT visualization, and a stereolithographic model, if needed. Conventional acrylic surgical splints were fabricated from plaster dental models. The backup surgical plan and acrylic surgical splints that were generated by conventional planning methods were also available to the surgeon during the procedure.

The accuracy of the CASS planning method was determined by comparing planned outcomes with actual outcomes (Fig 1). This was done on the assumption that the ideal surgical planning system produces actual outcomes that are identical to predicted outcomes. Actual surgical outcomes were measured on the basis of a postoperative CT scan that was obtained within the first 6 postoperative weeks. The 6-week interval was selected to avoid bias caused by possible patient growth or orthodontic movement.
Both the composite skull model with the planned outcome and the actual postoperative CT model were imported into the computer. The outcome evaluation was completed by first superimposing the postoperative model onto the planned model, and then measuring the differences between planned and actual outcomes. The criteria used to determine the accuracy of the technique consisted of a linear difference between planned and actual outcomes of less than 2 mm,10-13 and an angular difference of less than 4°.14

To superimpose the postoperative model onto the planned model, a surface-best-fit method was used. On the planned model, we initially “hid” the bony segments that had been moved (i.e., maxilla and mandible). Only the bones that had not been moved during surgery were visualized. This was done to avoid possible operator bias during the superimposition process. In cases of maxillary surgery, only the cranial region was visualized. In cases of mandibular surgery and genioplasty, only the mandibular distal segment without teeth was visualized. This was done to avoid the bias caused by mouth opening for postoperative CT scan. We then rendered the postoperative CT model in a semitransparent mode and aligned it with the planned model, using the surface-best-fit method (Fig 2). After the postoperative model had been aligned with the planned model, all hidden bony segments of the planned model were displayed.

To evaluate whether the surgical plan was accurately reproduced in the operating room, we adopted the theory that 3 points can be used to define an object in 3D space. To do this, we first digitized certain landmarks on each planned and postoperative CT model, and then compared the differences between them (Fig 3). On the maxillary and mandibular occlusal surfaces, 3 landmarks were digitized: the midline between the 2 central incisors, and the right and the left mesiobuccal cusp tips of the first molars. During the digitization process, only 1 model (planned or postoperative) was displayed at a time. These 3 landmarks were then connected to form a triangle. To enhance visualization on the figure, the triangle is not shown. C, The planned model and the landmarks were masked as “hidden.” The postoperative model was loaded into the computer. The same 3 landmarks were digitized on the planned model. D, The 3 landmarks on the postoperative model were masked as “hidden.” The bones that had not been moved on the planned model were masked as “visible.” E, With the surface-best-fit method, the postoperative model was superimposed onto the planned model. F, Both models were masked as “hidden.” Finally, both sets of landmarks were masked as “visible.” The (x, y, z) coordinates of both sets of landmarks were recorded and paired.

![Figure 2](image-url)  
*Figure 2. Surface-best-fit superimposition method.*  

![Figure 3](image-url)  
*Figure 3. Accuracy evaluation of maxillary surgery. A, The planned computed tomography (CT) model was loaded into the computer. B, Three landmarks were digitized on the occlusal surface: the midline between the 2 central incisors, and the right and the left mesiobuccal cusp tips of the first molars. C, The planned model and the landmarks were masked as “hidden.” The postoperative model was loaded into the computer. The same 3 landmarks were digitized on the planned model. D, The 3 landmarks on the postoperative model were masked as “hidden.” The bones that had not been moved on the planned model were masked as “visible.” E, With the surface-best-fit method, the postoperative model was superimposed onto the planned model. F, Both models were masked as “hidden.” Finally, both sets of landmarks were masked as “visible.” The (x, y, z) coordinates of both sets of landmarks were recorded and paired.*

The coordinates \((x, y, z)\) of each landmark on planned and actual outcomes were recorded and paired. The same was done for the angles between the triangle edges and the horizontal plane. The difference between pairs was calculated, and the distribution of the data (difference) was screened. Scatter plot and formal tests showed that the data were not normally distributed. The authors hypothesized that this pattern was a result of the small sample size \((n = 5)\) and an outlier (patient no. 2). Therefore, the differences were presented as median and range. In addition, means and standard deviations between planned and postoperative outcomes were calculated. These will be used in the future for sample size calculation for larger and more comprehensive studies.

Finally, to ensure the accuracy of the surface-best-fit superimposition method, a test was completed with the use of patients’ planned and postoperative CT models (Fig 5). For the planned model, only the bones that had not been moved during surgery were displayed. Three clearly identified landmarks on both models were digitized. One (ie, the nasion) was located on the midline. The other 2 were located at the right and left posterior parts of the skull (ie, the lowest point of the mastoid process or the junction of frontal and temporal processes of the zygomatic bones). These landmarks were then hidden. The postoperative model was superimposed onto the planned model with use of the surface-best-fit method. After the 2 models had been aligned, the landmarks on both models were made visible. Their \((x, y, z)\) coordinates were recorded and paired, and differences were calculated. The distribution of data (the differences) was screened, and scatter plots and formal tests indicated that the data were normally distributed. Thereafter, Bland and Altman’s method for assessing agreement was used. Lack of agreement was estimated by the mean differences \((\bar{d})\) and standard deviations (SD) between planned and actual postoperative measurements. Lower and upper limits of the differences were estimated by \(d \pm 1.96SD\). The precision of the estimated mean differences of agreement was computed with the formula, \(d \pm t \sqrt{SD^2/n}\), where \(t\) is the critical value for the \(t\) distribution corresponding to the area \((2 \text{ tails at .05})\) under the curve, and \(\sqrt{SD^2/n}\), is the standard error of the mean difference.

**Results**

In this study, procedures for all 5 patients were successfully planned with the CASS planning method.
Computer-generated surgical splints were successfully used on all patients at the time of surgery. Backup surgical plans and acrylic surgical splints generated by conventional planning methods were not needed for any of the patients.

Differences between planned and actual measurements in x, y, and z directions are presented in Table 1. According to the CASS planning method, the largest linear difference between planned and actual postoperative outcomes was within 1.99 mm. In addition, the largest median of the linear difference was 0.85 mm. Finally, the means and standard deviations of the linear differences are presented in Table 1.

The differences between planned and actual orientations for maxillary, mandibular, and chin segments are presented in Table 2. With the CASS planning method, the largest angular difference between planned and actual postoperative outcomes was limited to within 3.48°. In addition, the largest median of the angular difference was 1.7°. Finally, means and standard deviations of the angular difference are presented in Table 2.

The surface-best-fit superimposition method was assessed. Mean differences between landmarks on planned and postoperative models were less than 0.12 mm, with standard deviation less than 0.19 mm (Table 3). In addition, the lowest limit of agreement between the 2 measurements was −0.37 mm, and the highest limit of agreement was 0.42 mm (Table 4). Finally, the precision of the estimated limits of the agreement was −0.24 mm for the lowest limit and 0.28 mm for the highest limit (Table 5).

Discussion

The ideal surgical plan is one that can be accurately reproduced in the operating room. The authors believe that CASS offers clinicians a better chance of achieving this goal. The results of this pilot study have shown that the accuracy of the CASS planning method (ie, within 0.9 mm and 1.7°) is promising. Previous studies have determined that a difference of less than 2 mm in linear measurement or 4° in an occlusal cant is not likely to be clinically significant.10-12,14 Thus, it can be stated that the small differences seen in our study are of no clinical relevance.

Unfortunately, the clinical accuracy of 3D surgical planning methods in the treatment of patients with complex CMF deformities has not been established. The criteria used in this study were derived from studies that evaluated the accuracy of planning methods in regular orthognathic surgery using 2D cephalograms. Donatsky et al10 reported the accuracy between planned and actual outcomes to be about 2 mm. Jacobson et al13 reported that 80% of actual
postoperative outcomes fell within 2 mm of the prediction. They documented a significant difference in planning accuracy between single-jaw and bi-maxillary procedures. In bi-maxillary orthognathic surgery, it was reported that discrepancies between linear measurements could be as great as 4.8 mm. All patients had complex craniomaxillofacial deformities and underwent bi-maxillary surgeries. Their surgeries required complex simultaneous translational and rotational movements to correct pitch, roll, and yaw of the bony segments in a 3D space. Even with these complex surgical movements, the largest error in our series was 2 mm (Table 2).

The largest discrepancy between planned and actual outcomes occurred in patient no. 2. This patient had a severe and complex asymmetric deformity that required a Le Fort I osteotomy, a sagittal split osteotomy on the right, an inverted “L” osteotomy on the left, a genioplasty with left-to-right transposition of a bony wedge, and an onlay bone graft to the inferior border of the mandible on the left. Because of the severity of his occlusal cant, mandibular osteotomy was performed first. During the surgery, generalized blood loss was difficult to control. After mandibular osteotomies, genioplasty, and onlay bone graft had been completed, we made the decision to stop the operation. Postoperative evaluation by the hematology service revealed that the patient had an undiagnosed type I von Willebrand disease. The bleeding workup required 6 weeks for completion. Therefore, the patient was not able to return to the operating room for the completion of the Le Fort I osteotomy until 2 months after the first procedure had been done. The second surgery was uneventful. The authors attribute the differences between planned and actual outcomes to postoperative relapse that occurred during this long interval, during which the mandible was not articulated against a stable base. This difference could have been smaller if the surgery had been performed in a single stage. Despite the relapse, the differences between planned and actual outcomes were limited to 2 mm and 3.5°.

Differences between planned and postoperative outcomes were measured with superimposition of planned and postoperative CTs. To accurately superimpose the postoperative model onto the planned model, the authors used the surface-best-fit method, rather than the traditional landmark-based method. This was done because of the findings of a study recently completed at our laboratory on the accuracy of various superimposition methods, in which it was concluded that the larger the number of points used for superimposition, the more accurate the superimposition becomes. The surface of the computer model contains a significant number of points, thus the surface-best-fit method should be more accurate than the landmark-based method. In addition, Bland and Altman’s method of assessing agreement was used to further confirm the accuracy of the surface-best-fit superimposition method. This test showed that the mean differences between 2 landmarks on planned and postoperative models were less than 0.1 mm, with standard deviation less than 0.2 mm (Table 3). In addition, the limits of agreement between the 2 measurements were small (Table 4). Even the largest limit was small enough that it was not clinically significant.

<table>
<thead>
<tr>
<th>Table 2. ANGULAR DIFFERENCE BETWEEN PLANNED AND ACTUAL POSTOPERATIVE ORIENTATION (PITCH, ROLL, YAW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Maxilla</td>
</tr>
<tr>
<td>Mandible</td>
</tr>
<tr>
<td>Chin</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Table 3. ACCURACY OF THE SUPERIMPOSITION (MEAN AND SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Right</td>
</tr>
<tr>
<td>Middle</td>
</tr>
<tr>
<td>Left</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Table 4. ACCURACY OF THE SUPERIMPOSITION (UPPER AND LOWER LIMITS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
</tr>
<tr>
<td>Upper</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Right</td>
</tr>
<tr>
<td>Middle</td>
</tr>
<tr>
<td>Left</td>
</tr>
</tbody>
</table>

Finally, the accuracy of the superimposition was further confirmed by precision of the estimated limits of agreement (Table 5).

In conclusion, the results of this pilot study are promising and will be used to calculate the sample size for a larger and more comprehensive study in which the accuracy of CASS planning methods will be further investigated.

Acknowledgments
The authors would like to thank Ron Kikinis, MD, Professor, Department of Radiology, and Director, Surgical Planning Laboratory, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA, for his suggestions and help with this project.
The authors also would like to thank Virginia A. Moyer, MD, MPH, Professor of Pediatrics, Center for Evidence-Based Medicine, University of Texas–Houston Medical School, Houston, TX, for her contributions in the preparation of this manuscript.

References

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Upper</td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>Right</td>
<td>−0.11</td>
<td>0.17</td>
<td>−0.16</td>
</tr>
<tr>
<td>Middle</td>
<td>−0.11</td>
<td>0.23</td>
<td>−0.24</td>
</tr>
<tr>
<td>Left</td>
<td>−0.12</td>
<td>0.27</td>
<td>−0.19</td>
</tr>
</tbody>
</table>
