Clinical and computerized assessment of mandibular asymmetry

Raymond Edler, David Wertheim* and Darrel Greenhill*
Norman Rowe Maxillofacial Unit, Queen Mary’s Hospital, Roehampton, London and
*School of Computing and Information Systems, Kingston University, Surrey, UK

SUMMARY The aim of this study was to investigate assessment of mandibular asymmetry by clinicians and to evaluate a new computerized system. Eight experienced clinicians (four maxillofacial surgeons and four orthodontists) assessed 12 standardized facial photographs of patients selected to provide a range, from normality to marked mandibular asymmetry. Photographs were taken under standardized conditions; analysis of photographs repeated after one month showed close agreement. The computerized system used four different methods for quantifying asymmetry based on right/left differences in perimeter, area, compactness, and centre of area ratios.

Good agreement (κ = 0.77) was achieved by the clinicians in assessing the likelihood of treatment need; however, when assessing an ‘acceptable’ level of asymmetry, agreement was moderate (κ = 0.46). Differences in perimeter ratios did not compare well with clinical assessment, but those for both area and compactness showed 100 per cent sensitivity and specificity to clinical assessment at ratio differences of 0.05 and 0.03 (deviation from 1), respectively. A centre of area difference ratio greater than 1 showed 75 per cent sensitivity and 85 per cent specificity to clinical assessment. These parameters could prove useful in quantifying change in asymmetries brought about by growth, treatment, or any subsequent relapse.

Introduction

It is widely recognized that what appears to be a symmetrical face in fact consists of a multitude of minor asymmetric components, the mandibular component being particularly significant. This has been demonstrated in a variety of studies of ‘normal’ that is ‘symmetric’ patients mostly based on radiographic evidence (Hewitt, 1975; Shah and Joshi, 1978; Melnik, 1992) and anthropometric evidence (Farkas and Cheung, 1981). Clearly, this evidence is important when determining the threshold beyond which ‘normal’ asymmetry becomes a clinically discernible problem for the orthodontist.

The emphasis in asymmetry assessment has been on the use of postero-anterior (PA) cephalogram analysis, involving the use of transverse linear and angular measurements (Grummons and Kappeyne van de Coppello, 1987) or triangulation (Hewitt, 1975). PA cephalogram measurements have been the basis of a variety of normal values for reference, i.e. the mid-sagittal and lateral axes of Chebib and Chamma (1981) and the use of midline to left and right differences used, in different ways, by Peck et al. (1991) and Skolnick et al. (1994). Mongini and Schmid (1987) described a computerized PA cephalogram based method using area superimpositions. However, as originally pointed out by Cook (1980), many of the landmarks used in PA cephalometry can be difficult to identify on a particular film and measurements can be prone to distortion, dependent on head position. This limitation is widely recognized (Proffit and White, 1991; A thanasiou, 1995) with recent evidence of inaccuracies in point location (Pirttiniemi et al., 1996; A thanasiou et al., 1999).
Dissatisfaction with the use of PA cephalograms has led to the use of submentovertex (SMV) radiographs (Forsberg et al., 1984; Arnold et al., 1994; Rose et al., 1994). Unfortunately, the SMV film has also not proved to be totally acceptable. Criticisms have been made about distortion and inaccuracy of point location (Peck et al., 1991). Taking the SMV film is not a particularly comfortable procedure for the patient and its measurements do not readily permit easy visualization of the patient’s actual appearance. Its use in combination with lateral and PA cephalograms was described by Grayson et al. (1983) in order to produce a three-dimensional (3D) method of evaluating the nature of a particular patient’s asymmetry, but this method does not seem to have been widely used since its introduction.

Some photographic techniques, for example a curve analysis approach, analyse asymmetry by deciding an appropriate line of symmetry and then comparing the difference in outline areas when one side is flipped contralaterally (Coghlan et al., 1987, 1993; Schmid et al., 1991). However, such techniques are dependent upon determination of an appropriate line of symmetry. This may result in the assessment depending on the precise manner in which the algorithm used applies to the image. This may make it difficult to know exactly how the computer reaches an assessment and may influence interpretation.

More recently, 3D methods have been described (Burke, 1992; McCance et al., 1997a,b). However, these require quite complex stereo-photogrammetric or scanning techniques, which are not widely available and are likely to be beyond the reach of those clinicians not working in specialized centres. Previous studies have shown that two-dimensional assessment of nasal asymmetry from photography can be of value in the evaluation of treatment for the noses of cleft lip and palate patients (Coghlan et al., 1987, 1993). Clearly, a valid and reproducible method of quantifying asymmetry could be of clinical value, as it would aid assessment in mild mandibular asymmetries, so common in orthodontic practice. This would also facilitate measurement of change in asymmetry whether by growth, treatment, or indeed, relapse after treatment.

The aims of this study were to determine the extent of agreement of mandibular asymmetry assessment between experienced clinicians from standard photographs and to compare the findings with an assessment using a new computerized method of analysis.

Methods
Twelve 35-mm transparencies, comprising standard frontal facial views of 11 patients (two males, nine females), median age 16 years (range 11–23 years), were taken as part of their clinical assessment. The patients were selected by one clinician (RE) to provide a full range from ‘normality’ to marked asymmetry. The number of images was established at 12, as this seemed both large enough to allow the clinicians to be presented with an adequate spectrum of mandibular asymmetry cases, yet small enough to be assessed in one ‘sitting’. Three of the images are shown in Figure 1a–c. For one child, one image was taken prior to treatment and one image after mandibular asymmetry correction. The images were scanned, then transferred to a computer and stored on CD-ROM. The images were then printed at a size of 11 × 16 cm using a photographic quality printer.

Photographs
The full-face photographs were taken under standardized conditions on a 35-mm slide film. The patient was seated, with head vertical, looking straight at the camera (positioned for 1:8 magnification) with the visual axis horizontal. There were two background lights (Bowens International, Clacton, Essex, UK; 250 W seconds strength) at head level, one on either side behind the patient. Two main Bowen Esprit lights (500 W seconds) with diffuse reflectors were situated either side in front of the patient, in order to avoid shadow. These were 6 feet high and aimed at 45 degrees to the patient. The patient’s chair was placed 6 feet from the background (Figure 2a,b). Spectacles were removed and the patient asked to close lightly together on the back teeth.
Clinical assessment

Four orthodontists and four maxillofacial surgeons were asked to assess the images. This number was chosen since four clinicians should provide a good balance, both allowing for an adequate range of clinical opinion within each specialty, yet permitting the presentation of the results in a reasonably practical manner (Tables 1 and 2). They were asked to categorize the images as follows:

A reasonably symmetric
B asymmetric.

If category B was chosen, the clinician then had to decide which of the following categories were appropriate:

C the individual would be unlikely to seek treatment (i.e. clinically acceptable)
D the individual would be likely to seek treatment (i.e. clinically unacceptable).

The images were presented in an album, in random order, and the same order was used for each clinician. The procedure was repeated two weeks later to allow assessment of intra-examiner agreement.

Assessment of agreement was carried out using the kappa statistic (Cohen, 1960; Fleiss, 1971; Shoukri and Pause, 1998), and interpretation made using the criteria suggested by Landis and Koch (1977) and Altman (1991).

Computerized assessment

The digitized images were analysed using a PC-based computer image analysis system (Greenhill et al., 2000). In summary, the digitized images are displayed on the computer screen, the user then marks with a mouse the inferior insertions of the ears (Otobasion inferius, Figure 3a); the computer then draws a line (A) between the ear insertions. The location of these points proved to be more straightforward than that of either the inner or outer canthus of the eyes—the natural alternative. Additionally, it had previously been identified that any errors involved in calculating asymmetry would be reduced if the baseline A was relatively long. However, when required the program can also provide an alternative baseline, based on the
inner or outer canthus of each eye, which may be more appropriate in cases of asymmetric ear position.

The user then clicks around the mandible from right to left and the computer draws line segments between the points around the mandibular outline and a line (B) at the bisection of A and perpendicular to A (Figure 3a,b). The lower portion of the face is thus effectively divided into two segments bounded by the mandibular outline. In this way the lower portion of the face can be analysed as left and right segments.

Four methods were used to assess the extent of asymmetry. First, the area of the left segment and that of the right were computed and a ratio of the two evaluated to give a dimension-independent measure. Second, the ratio of the perimeter lengths of each of the two segments was computed in a similar fashion.
The third method involved an assessment of compactness, defined as the square of the perimeter divided by the area (Gonzalez and Woods, 1993), thus providing a measure of shape. The ratio of the compactness of each side was calculated.

Finally, the centroid (centre of area) of both areas combined was calculated. This can be a useful way of determining position, for example, cell position (Hoppe et al., 1999). Its deviation from the point of bisection of line A was calculated, divided by the length of line A, and then multiplied by 100.

The combination of an area-based and shape measurement, i.e. compactness, is important, since it is possible for the areas of each segment to be similar, yet their shapes different. Area, perimeter, compactness, and deviation of the centre of area ratios were displayed with a resolution of 0.01.

Inter-observer agreement was assessed by two observers using the software to evaluate the 12 images. Agreement was assessed using the method of Bland and Altman (1986). The distribution of the difference in the measurements was assessed for normality with the Ryan–Joiner method using Minitab Release 12.1 (Minitab, Inc., USA).

Comparison of the clinical assessment and computer results was made on the basis of sensitivity and specificity (Altman, 1991), i.e. the proportion of positive/negative assessments appropriately identified by the program measurements with respect to clinical interpretation.

Assessment of variation in photographs (repeatability)

Five individuals, not involved in the main study, each had two photographs taken approximately one month apart. The slides were scanned, digitized, and then analysed by computer, as described above, to assess repeatability, so as to evaluate the variation associated with repeating the photographs.

Results

Table 1 shows the results of the clinical assessment of asymmetry on the basis of viewing the images.

Using the kappa statistic to test the agreement of the eight observers with the three categories (A, C, or D), $\kappa = 0.46$. This implies moderate agreement. The data were also analysed on the basis of whether the clinicians considered that treatment would be sought, or otherwise. For this analysis categories A and C were combined, and these results showed much better agreement with $\kappa = 0.77$, implying good agreement.
The data in the two categories for the orthodontists and the maxillofacial surgeons were also analysed separately: for the orthodontists $\kappa = 0.62$, indicating good agreement; for the maxillofacial surgeons $\kappa = 0.88$, indicating very good agreement.

Intra-examiner agreement

The assessment of the photographs was repeated approximately two weeks after the initial assessment (Table 2). When considering the three-category analysis there were differences in at least one patient assessment for seven of the observers; of these the median number of photographs with a different assessment was four. When considering the two-category analysis, there were differences in at least one patient assessment for four of the observers; in three observers there was a difference in one patient assessment; two of these were of a patient where agreement was not reached (patient 5). For one observer there were three differences in assessment, one of which was patient 5.

Computerized analysis

The results of the computerized analysis are given in Table 3. The mean of the computer assessment of the two observers is shown. Figure 4 shows the difference in area ratio from the two observers against the mean of the area ratio. The Ryan–Joiner test for normality of the difference in area ratio measurements was not consistent with a normal distribution and, hence, the range observed is indicated. The results indicate close agreement between the observers: median difference, 0.00; range, -0.01 to +0.04.

Comparison of clinical and computer assessment

When the clinical assessments were categorized as to whether treatment would be sought or not, there was one case in which four of the eight clinicians considered treatment would be appropriate (patient 5). In eight cases, there was complete agreement and in three cases seven of the eight clinicians agreed. Hence, for comparison with the computer analysis the 11 cases of good agreement were compared with diagnostic cut-offs to enable evaluation of sensitivity and specificity.

From Table 3 it can be seen that the ratio of the segment perimeters was of no obvious clear value in assessing asymmetry, in that there was very little variation and no clear association with clinical assessment.

An area ratio deviation from 1 of greater than 0.05 showed 100 per cent sensitivity and 100 per cent specificity to clinically assessed need for treatment; Figure 5a,b shows the distribution of area ratio and compactness ratio with respect to clinical assessment of treatment need. A compactness ratio deviation from 1 of greater than 0.03 showed 100 per cent sensitivity and 100 per cent specificity to clinically judged need for treatment. Additionally, an absolute centre of area difference ratio greater than 1.0 showed 75 per cent sensitivity and 85 per cent specificity to clinically judged need for treatment.

Using the above criteria, case 5 would have been judged as clinically unacceptable in area and compactness ratio, but would be clinically acceptable on the basis of the centre of area.

Assessment of variation in photographs

(repeatability)

For the five patients who had two photographs taken one month apart, the median absolute difference in area ratio was 0.01 (0.01–0.02). For length ratio, the median absolute difference
was 0.00 (0.00–0.01), for compactness ratio, the 
median absolute difference was 0.01 (0.01–0.02),
and for centre of area ratio the median absolute 
difference was 0.23 (0.09–1.04).

Discussion
The results of this study indicate that on the 
basis of whether treatment for asymmetry was 
deemed likely or not, there was a high degree 
of agreement ($\kappa = 0.77$) between the clinicians. 
However, the level of agreement was poorer
($\kappa = 0.46$) if the extra category of ‘asymmetry, 
but not likely to seek treatment’ was included. 
When assessment of photographs was repeated 
by the observers, using two categories, four of 
the eight observers had full agreement with their 
previous assessment and three differed in one 
patient assessment.

Close inter-observer agreement was observed 
in the evaluation of 12 images with the new 
computer analysis system. The system calculates 
dimension-independent measurements, such as 
area ratio, which should help to improve repeat-
ability; this is confirmed by the low variation in 
results of the photographs taken one month 
apart: the median differences of area, perimeter, 
and compactness ratios lying within the display 
resolution of 0.01. It may also be useful to identify 
reproducibility levels in repeated photographs 
taken in a more conventional setting, i.e. by 
an orthodontist. Inter-examiner identification 
errors of 0.31–4.79 mm were recorded by Major 
et al. (1994) from PA cephalogram assessments 
and a level of accuracy of 0.5 mm has been 
suggested for 3D laser scanning (Moss et al., 
1991). However, it is clearly not possible to relate 
these dimension-dependent assessments with the 
ratios presently described.

When the results of the clinical assessment 
were compared with the new computer analysis 
system, it was possible to define sensitive and 
specific diagnostic cut-off points, from which it 
appears that compactness, area, and possibly 
centre of area are potentially useful measures of 
quantifying asymmetry. However, further studies 
are required to investigate this on a larger 
cohort. Perimeter ratio was not found to be 
useful as an indicator of whether treatment 
was required; this may be due to the resolution 
of the measurement of 0.01 not being sufficient 
or it may reflect that perimeter measurement is 
more sensitive to inaccuracies in path tracing. It 
seems likely that a combination of the measure-
ments will be the most appropriate means to 
assess asymmetry. This is, in part, because there 
are different types of asymmetry. The combination 
of area ratio with compactness being a shape 

<table>
<thead>
<tr>
<th>Patient</th>
<th>Clinicians recommend treatment</th>
<th>Area ratio mean</th>
<th>Perimeter ratio mean</th>
<th>Compactness ratio mean</th>
<th>Centre of area deviation ratio mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N</td>
<td>0.99</td>
<td>1.00</td>
<td>1.005</td>
<td>0.165</td>
</tr>
<tr>
<td>2</td>
<td>N</td>
<td>0.98</td>
<td>0.995</td>
<td>1.01</td>
<td>-0.71</td>
</tr>
<tr>
<td>3</td>
<td>Y</td>
<td>0.82</td>
<td>0.96</td>
<td>1.12</td>
<td>2.515</td>
</tr>
<tr>
<td>4</td>
<td>N</td>
<td>0.96</td>
<td>0.99</td>
<td>1.02</td>
<td>0.065</td>
</tr>
<tr>
<td>5</td>
<td>Not agree</td>
<td>0.94</td>
<td>0.99</td>
<td>1.035</td>
<td>0.90</td>
</tr>
<tr>
<td>6</td>
<td>N</td>
<td>0.995</td>
<td>1.00</td>
<td>0.995</td>
<td>-0.79</td>
</tr>
<tr>
<td>7</td>
<td>N</td>
<td>0.97</td>
<td>0.99</td>
<td>1.02</td>
<td>-0.38</td>
</tr>
<tr>
<td>8</td>
<td>Y</td>
<td>0.84</td>
<td>0.96</td>
<td>1.10</td>
<td>4.01</td>
</tr>
<tr>
<td>9</td>
<td>N</td>
<td>0.985</td>
<td>0.995</td>
<td>1.01</td>
<td>1.025</td>
</tr>
<tr>
<td>10</td>
<td>N</td>
<td>1.02</td>
<td>1.005</td>
<td>0.995</td>
<td>-0.325</td>
</tr>
<tr>
<td>11</td>
<td>Y</td>
<td>1.075</td>
<td>1.015</td>
<td>0.955</td>
<td>-1.335</td>
</tr>
<tr>
<td>12</td>
<td>Y</td>
<td>1.065</td>
<td>1.01</td>
<td>0.96</td>
<td>-0.61</td>
</tr>
</tbody>
</table>

N = no (A or C); Y = yes (D).
For each parameter, the asymmetry ratio either increases or decreases from a score of 1 (representing symmetry) except 
for centre of area ratio, which increases or decreases from zero, indicating the level of asymmetry for that patient image.
measure may help to address the question of different types of asymmetry. The computerized system uses standard photographs and, hence, is non-invasive, neither requiring specialized equipment nor involving the patient in extra investigations. It is neither intended as a diagnostic aid nor for treatment planning. Additionally, it is recognized that there are other forms of facial asymmetry for which this system may not be appropriate. However, these results do suggest that the new computer analysis system could help in quantifying mandibular asymmetry and may thus be of clinical value. Further work could usefully explore the basis for selecting the most appropriate method of calculating asymmetry. This would be valuable in a number of ways:

1. in mild asymmetries, in determining the threshold level beyond which problems arise when contemplating routine non-surgical orthodontic treatment;
2. when monitoring asymmetry cases periodically, pending treatment later;
3. in auditing the quality of orthodontic/surgical correction and measuring subsequent relapse.

**Conclusions**

Good agreement was obtained by the eight clinicians in assessing the anticipated need for treatment for mandibular asymmetry. Agreement was moderate when assessing an ‘acceptable’ level of asymmetry. Good agreement was observed in the measurements when the assessments were made on repeated photographs.

In this preliminary study, compactness, area, and, to a lesser extent, centre of area ratio differences showed high sensitivity and specificity to the clinicians’ assessment of treatment need.

Thus, the computer system may be of value in assessing mandibular outlines.

**Address for correspondence**

R. Edler
Orthodontic Department
Norman Row Maxillofacial Unit
Queen Mary’s Hospital
Roehampton Lane
Roehampton
London SW15 5PN
UK

**Acknowledgements**

We are grateful to Mr Neville Cramer for statistical advice and to Mr Andy Belsham for help in digitizing the slide images. We are also grateful to Mr Neil Maffre, Photographic Department, Queen Mary’s University Hospital, Roehampton, for taking the clinical photographs.
We wish to thank the maxillofacial surgeons Malcolm Bailey, Peter Blenkinsopp, Memo Manisali, and Andrew Stewart, and the orthodontists Melissa Chen, Bob Mordecai, Fraser Mc Donald, and Claire Nightingale for undertaking the clinical assessments.

References


