Appreciation of cleft lip and palate treatment outcome by professionals and laypeople

Despina A. Papamanou*, Nikolaos Gkantidis**, Nikolaos Topouzelis*** and Panagiotis Christou*

*Department of Orthodontics, Dental School, University of Athens, Greece, **Department of Orthodontics and Dentofacial Orthopedics, Dental School, University of Bern, Switzerland and ***Department of Orthodontics, Dental School, Aristotle University of Thessaloniki, Greece

Correspondence to: Despina A. Papamanou, Department of Orthodontics, Dental School, University of Athens, Thivon 2, Goudi, 11527, Greece. E-mail: dpapamanou@dent.uoa.gr

SUMMARY The aim of the study was to analyse the aesthetic evaluation of head photographs of treated individuals with clefts by laypeople and professionals and to investigate how certain cephalometric variables could be related to their rating. A set of five standardized head photos (frontal, both laterals, three-quarter right and left) of 12 Caucasian patients with treated unilateral cleft lip and palate were presented to 12 adult laypeople, 12 orthodontists, and 12 maxillofacial surgeons. For each set of photos the judges had to answer four questions on a visual analogue scale (VAS). The answers were analysed for intra- and inter-panel level of agreement and correlations of assessments with certain cephalometric parameters were determined. There was a high level of agreement for all assessments of each panel of raters. However, laypeople were less satisfied with lip and nose aesthetics compared to professionals. The three groups were similarly satisfied with the aesthetics of the jaws and the face. The anterior position of the maxilla (SNA) influenced positively professionals’ ratings of facial aesthetics. Orthodontists were negatively influenced when the vertical dimension of the face or the distance of the lower lip to E-plane were relatively increased. The latter was the only cephalometric parameter correlated with lower aesthetic scores obtained from laypeople. Professionals report greater satisfaction from the treatment outcome and evaluate cleft consequences with less severity than laypeople. According to cephalometric findings, the relative positions of the lips seem to dominate facial aesthetics’ appreciation by laypeople, while specialists appear to focus on different features of the face.

Introduction

One of the principle aims of cleft lip and palate treatment is to alleviate the undesirable sequelae and improve the facial appearance of individuals with cleft. Cleft treatment and rehabilitation procedures can be quite extensive, while previous research demonstrated that under the correct circumstances the functional rehabilitation can be successful (Friede and Katsaros, 1998). However, even after the end of treatment, patients still seem to have concerns about the appearance of cleft-related features (Marcusson et al., 2002; Sinko et al., 2005). Their appearance can be disturbed due to scars formed after surgical interventions (Ritter et al., 2002) and this should not be neglected. Facial appearance is clearly important to an individual’s social well being (Rhodes, 2006; Anderson et al., 2008) especially in adolescents, where dentofacial aesthetics are important for their self-concept and self-esteem (Shaw, 1981; Thomas et al., 1997; Turner et al., 1998).

Patients with scarring or variable degree of facial deformity suffer from a variety of psychosocial problems as they can be subjected to various forms of social discrimination (Shaw, 1981; Rumsey et al., 1986; Hunt et al., 2006; Noor and Musa, 2007). Self-assessment of the facial appearance by an individual with cleft may be biased as previous experiences related to the problem may affect judgments (Pitak-Arnop et al., 2010), although this is not always the case (Broder et al., 1994). Furthermore, assessments by the peers may be biased (Shaw, 1981) as members of the same family can be influenced either because of personal affection to the patient or because they are used to his/her appearance (Broder et al., 1992). However, the patient’s social interactions depend on the assessments of members of the non-immediate environment.

Although extensive previous research has addressed cleft individuals’ self-perception (Broder et al., 1994; Hunt et al., 2006; Noor and Musa 2007) and/or assessments by professionals (Sinko et al., 2005; Russell et al., 2009) or comparisons of cleft to non-cleft individuals (Oosterkamp et al., 2007; Russell et al., 2009), less attention is given on the evaluation of the aesthetic appearance of the face after treatment by laypeople and on attempts to correlate findings to objective parameters. The only study that compared the
aesthetic assessment of individuals with clefts between laypeople and professionals and attempted to correlate findings to objective evaluations was the one by Meyer-Marcotty and Stellzig-Eisenhauer (2009) by using 3D measurements of facial asymmetry. However, this is not a method that can be applied in everyday practice and it cannot provide clear discrimination between different features of the facial complex.

The aim of the present study was to analyse the aesthetic evaluation of head photographs of treated cleft individuals by laypeople and professionals and to investigate potential impact of certain cephalometric variables on this evaluation.

**Materials and methods**

The material consisted of photographs and cephalometric radiographs of 12 Caucasian patients (9 male, 3 female) with unilateral cleft lip and palate (right: 3, left: 9) who had completed all the stages of their cleft treatment within the last 10 years (mean age: 22.1, range: 17.5–27.4). Patients with syndromes and other congenital anomalies or psychological disorders were excluded from the study. The surgical treatment of the patients followed various protocols, but all patients received orthodontic treatment at the Postgraduate Orthodontic Clinic of the University of Athens, Greece. Consecutively treated patients, who were appropriate for inclusion in the study, along with their parents were thoroughly informed for the study. The sample consisted of the first 12 pairs of patients–parents who all accepted to participate and sign an informed consent statement. All study records were taken within a 2 week time interval. The study protocol was approved by the Ethical Committee of the Dental School of the University of Athens (Protocol No. 135/26.01.2010).

Five head photos from each patient (frontal face, right/left lateral face, three-fourth right/left face) were taken by one investigator (DAP), in one visit, under standardized conditions, identical light conditions, and with the same photographic set-up. All cephalometric radiographs were taken in the same cephalometric set-up and depicted a reference ruler on the cephalostat for correction of the magnification factor.

The sets of photos were presented in a standardized manner (Figure 1) to three panels of raters: 12 adult laypeople, 12 qualified orthodontists, and 12 qualified maxillofacial surgeons. All raters were not related to the patients and were not implicated whatsoever in any stage of their treatment. Specialists had more than 10 years of professional experience after acquisition of the specialty title. Laypeople and orthodontists were matched for gender, whereas males were predominant in surgeons’ panel. Age of all raters was matched within a 5 year divergence between corresponding raters. In addition, laypeople were randomly selected from a population that lived in the same region with the professionals and presented a wide range of socio-economical status.

The raters were asked to fill in a questionnaire consisted of four questions for each set of photos (Figure 2) and their answers were registered on a 100 mm Visual Analogue Scale (VAS) (Aitken, 1969) (Figure 3). The distances between the start of the scale and the markings of each rater were measured (mm) by one investigator (DAP) with an
electronic digital pointed jaw calliper (Jainmed Inc., Seoul, Korea) to transform ratings to continuous metric variables.

The cephalometric radiographs of the same patients were digitized by one researcher (DAP) using Viewbox software (version 3.1.1.13, dHAL Software, Kifissia, Greece) and the values of five angular (SNA, SNB, ANB, GoGn–SN, nasolabial angle) and two linear (upper lip to E-plane, lower lip to E-plane) measurements, representing skeletal and soft-tissue relationships, were assessed (Figure 4).

Statistical analysis
Cronbach’s alpha (α) (Schmitt, 1996) was calculated to assess intra-class level of agreement for each panel, regarding all the questions together and each question separately, based on individual scores of the 12 observers of each group. Cronbach’s alpha (α) above 0.8 indicates high reliability.

The inter-panel agreement was determined with Spearman correlation coefficient based on the median VAS score of the 12 observers of each group for each patient. It was calculated for each parameter as it was scored by each group of observers.

Non-parametric Friedman test was used to evaluate differences between the ratings of the three panels (Friedman, 1940). The aesthetic assessment score of each subject was the median of the VAS scores calculated for each group of raters. These variables were used for comparing the rating of each parameter by the different panels. In cases where Friedman test identified significant differences, pairwise comparisons were performed by the non-parametric Wilcoxon signed-rank test (Wilcoxon, 1945). Spearman correlation coefficient was calculated to evaluate correlations between each of the measured cephalometric variables and the median VAS scores given by the raters to patients. Correlation coefficient above 0.8 defines high correlation, while moderate correlation is defined between 0.6 and 0.8. The level of significance (α) is set at \( P \leq 0.05 \).

Error of the method
Two weeks after the initial measurement, 30 VAS scores were remeasured by the same researcher (DAP). Paired
Table 1 Internal consistency of the three panels measured by Cronbach’s α coefficient (α > 0.8 = high consistency, 0.8 > α > 0.6 = average consistency, α < 0.6 = low consistency). “All” refers to all types of questions tested together.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Surg</th>
<th>Lays</th>
<th>Ortho</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.906</td>
<td>0.870</td>
<td>0.902</td>
</tr>
<tr>
<td>Nose</td>
<td>0.885</td>
<td>0.843</td>
<td>0.908</td>
</tr>
<tr>
<td>Upper lip</td>
<td>0.901</td>
<td>0.889</td>
<td>0.942</td>
</tr>
<tr>
<td>Jaws</td>
<td>0.960</td>
<td>0.878</td>
<td>0.914</td>
</tr>
<tr>
<td>Face</td>
<td>0.926</td>
<td>0.871</td>
<td>0.896</td>
</tr>
</tbody>
</table>

Table 2 Inter-panel agreement among the appreciation of laypeople, surgeons and orthodontists as determined by Spearman’s correlation (*P < 0.05).

<table>
<thead>
<tr>
<th>Parameter assessed</th>
<th>Laypeople versus orthodontists</th>
<th>Laypeople versus surgeons</th>
<th>Orthodontists versus surgeons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rho</td>
<td>P</td>
<td>rho</td>
</tr>
<tr>
<td>Nose</td>
<td>0.76</td>
<td>0.00*</td>
<td>0.93</td>
</tr>
<tr>
<td>Upper lip</td>
<td>0.80</td>
<td>0.00*</td>
<td>0.89</td>
</tr>
<tr>
<td>Jaws</td>
<td>0.82</td>
<td>0.00*</td>
<td>0.85</td>
</tr>
<tr>
<td>Face</td>
<td>0.84</td>
<td>0.00*</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Table 3 Descriptive statistics and pairwise comparisons between groups’ assessments by Wilcoxon signed-rank test. Pairwise comparisons were not performed for Jaws and Face parameters since significant differences were not evident by Friedman test (n.s. = not significant, *P < 0.05).

<table>
<thead>
<tr>
<th>Parameter assessed</th>
<th>Median VAS Lays</th>
<th>P value</th>
<th>Median VAS Lays</th>
<th>P value</th>
<th>Median VAS Ortho</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nose</td>
<td>35.9</td>
<td>0.04*</td>
<td>35.9</td>
<td>0.00*</td>
<td>53.4</td>
<td>0.08, ns</td>
</tr>
<tr>
<td>Upper lip</td>
<td>39.4</td>
<td>0.00*</td>
<td>39.4</td>
<td>0.00*</td>
<td>50.6</td>
<td>0.10, ns</td>
</tr>
<tr>
<td>Jaws</td>
<td>64.8</td>
<td>—</td>
<td>64.8</td>
<td>69.6</td>
<td>61.9</td>
<td>69.6</td>
</tr>
<tr>
<td>Face</td>
<td>60.3</td>
<td>59.7</td>
<td>60.3</td>
<td>67.0</td>
<td>59.7</td>
<td>67.0</td>
</tr>
</tbody>
</table>

ns, not significant.
nose. In all cases, the evaluation of the upper lip was negatively correlated with the distance of the upper and the lower lip from E-plane. SNA was significantly correlated with the assessments of the jaws only for orthodontists, while GoGn-SN and lower lip to E-plane measurements were negatively correlated to the satisfaction of all panels. With regards to the face, the anterior position of the maxilla (SNA) positively influenced the ratings by orthodontists and surgeons. Orthodontists were also negatively influenced by high values of GoGn-SN and lower lip to E-plane measurements. The increasing distance of the lips from E-plane was the only factor that negatively affected the assessment of the face by laypeople in a significant manner.

Discussion

The aim of this study was to evaluate whether there are differences in the assessment of facial attractiveness of treated cleft individuals by raters of variable background and to investigate how certain cephalometric parameters might be related to their assessments. Cleft patients undergo a quite copious treatment with successive surgeries that may result in extensive scarring and disfigurement. Although several studies evaluated subjective assessments of treatment results (Strauss et al., 1981; Sinko et al., 2005; Oosterkamp et al., 2007; Meyer-Marcotty and Stellzig-Eisenhauer, 2009), no previous study examined potential differences between trained professionals and laypeople on the aesthetic outcome of these procedures and their correlation with objective cephalometric measurements. Meyer-Marcotty and Stellzig-Eisenhauer (2009), in a similar set-up, evaluated dentofacial aesthetic perception of patients, laypeople, and professionals using a 9-point scale and qualitative parameters and they correlated these assessments with the degree of 3D facial asymmetry. They concluded that asymmetry, especially of the midface, exerts a negative influence on how facial appearance is self-perceived or perceived by others, which explains the primary desire or need for nose correction. The self-perception of patients affected by cleft lip and palate was not correlated with objective variables or with the way they were perceived by others.

In our study, the panels of raters were of two different types as we could distinguish laypeople and professionals. Laypeople had no previous experience related to clefts. This kind of laypeople forms the major part of the social environment of individuals with clefts (Rumsey et al., 1986). Orthodontists were selected to represent professional judgment since soft and hard tissue relationships are everyday routine in orthodontic practice for evaluating facial aesthetics. Assessments by surgeons are equally important as they are more experienced in evaluating scarred soft tissues. However, both surgeons and orthodontists are expected to have a type of professional bias since they are familiar with the difficulties of treating such cases and this may influence them to be more tolerant to undesirable aesthetic results.

Males and females were matched concerning the panels of orthodontists and laypeople but not the panel of surgeons. However, previous research (De Smit and Dermaut, 1984) did not identify any significant difference between male and female participants and thus the possibility of gender bias is minimal.

Photographic records were collected under standardized conditions. The only potential variation of the images could be attributed to different stages of biological development of the patients or to the presence of facial hair (two cases). We decided not to exclude these cases from the consecutive selection since this could create selection bias. These people function and socialize properly with their present appearance and therefore, they should be judged so.

Specialists are expected to be more adept and familiar with the evaluation of facial and profile photos compared to unexercised laypeople. To compensate for this, in the present study, both profiles with a three-quarter view of the face were included. This compilation is closer to 3D approach providing the best impression of a person’s facial appearance (Van der Linden and Boersma, 1987) and it simulates everyday social contacts since one usually looks at another person at various angles.

Photos may have certain limitations (Asher-McDade et al., 1991), since they represent a 2D representation with no further information about form and function. For this reason, video recordings have been used in order to evaluate function and aesthetics, but the reported intra- and inter-rater reliability was generally poor (Morrant and Shaw, 1996; Ritter et al., 2002). When using full-face pictures, the general facial attractiveness may influence the evaluations of the raters (Tobiasen et al., 1991; Asher-McDade et al., 1991). Therefore, a valid proposal could be to mask patients’ characteristics or to crop the photos to illustrate the affected area. However, performing the evaluation on modified photos is generally not a skill familiar to the raters (especially to laypeople), and thus, it may insert bias to the results (Lo et al., 2002).

Furthermore, the set-up of the present study is one-shot evaluation of photographs without any previous training of the raters. It is practically the closest representation of an everyday social interaction (Peerlings et al., 1995). For this type of assessment, a panel size of four seems to be adequate (Peerlings et al., 1995), while in our study, the panel size was increased to enlarge reliability of judgments. On the other hand, the sample size of 12 patients may be considered relatively small, but the reduced size aimed not to discourage or fatigue raters by presenting too many sets of photos for evaluation (Ritter et al., 2002; McLaughlin et al., 2009).

The answers of the raters were registered on VAS. Its principle disadvantage is that we cannot directly connect numbers to certain categories or ordered scales because
usually data are not normally distributed in all cases (Svensson, 2000; Knutsson et al., 2010). On the other hand, in categorical rating scales, the selection of the number of points in the scaling may influence the reliability of the results (Grant et al., 1999). Apparently, VAS is more objective (Grossman et al., 1992), repeatable, reliable, and sensitive (Grant et al., 1999) than categorical rating scales with or without verbal descriptors (Williamson and Hoggart, 2005).

Since each patient was rated three times, between-group differences were evaluated in a paired manner between corresponding ratings of the same subject and not in a group-pooled way. This option was selected to add power to the analysis counteracting the effect of relatively small sample size. All panels seem to evaluate cleft faces in a similar way. In all cases, they agree on order but not on magnitude and there is high consistency among raters. Homogeneity within each panel was shown to be good (Table 1) and along with inter-panel correlations (Table 2) reveal a high degree of consensus in the evaluation of facial aesthetics.

The results show a high degree of agreement between the specialists’ panels in spite of their divergent educational background, which corresponds to previously published research (Meyer-Marcotty and Stellzig-Eisenhauer, 2009; Fabré et al., 2010). It should be noted that both specialist categories presented higher scores than laypeople (Table 3), which is an indicator of greater satisfaction from treatment outcome. Similar findings have been reported by previous research (Sinko et al., 2005). Meyer-Marcotty and Stellzig-Eisenhauer (2009) concluded in the same result by qualitative methods. This could be attributed to the fact that specialists are more familiar with the aesthetic consequences of the cleft and the difficulties of treating them, and thus evaluate clefts with less severity than laypeople. This is enhanced by the fact that in our study, differences were mostly evident in nose and lip aesthetics and less evident in jaw or face aesthetics. However, there are studies which report that professionals rated individuals with clefts more critical than laypeople (Ritter et al., 2002).

While nose and lip assessments revealed a certain degree of dissatisfaction of laypeople from treatment outcome (Table 3), nose assessments were not significantly correlated to any of the cephalometric parameters evaluated (Table 4). This may be attributed to the 2D nature of the cephalometric radiograph. According to previous research, asymmetry in the dimension of the frontal level, which is not illustrated in a cephalometric radiograph, is the main reason for the disturbed nasal aesthetics in individuals with clefts (Sinko et al., 2005; Meyer-Marcotty and Stellzig-Eisenhauer, 2009).

The aesthetic assessment of the upper lip was negatively correlated to the distances of the upper and lower lip from E-plane for all panels in a moderate to high degree (Table 4). The antero-posterior position of the lower lip exerted also a

---

**Table 4**

<table>
<thead>
<tr>
<th>Parameter assessed</th>
<th>Laypeople</th>
<th>Orthodontists</th>
<th>Surgeons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper lip E-plane</td>
<td>$\rho$</td>
<td>$\rho$</td>
<td>$\rho$</td>
</tr>
<tr>
<td>$P$</td>
<td>$P$</td>
<td>$P$</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SNA</td>
<td>0.22</td>
<td>0.59</td>
<td>0.37</td>
</tr>
<tr>
<td>NBI</td>
<td>0.08</td>
<td>0.79</td>
<td>0.33</td>
</tr>
<tr>
<td>ANB</td>
<td>-0.23</td>
<td>0.35</td>
<td>0.12</td>
</tr>
<tr>
<td>GADSN</td>
<td>0.08</td>
<td>0.76</td>
<td>0.34</td>
</tr>
<tr>
<td>NAS</td>
<td>0.06</td>
<td>0.81</td>
<td>0.31</td>
</tr>
<tr>
<td>Upper lip to E-plane</td>
<td>$\rho$</td>
<td>$\rho$</td>
<td>$\rho$</td>
</tr>
<tr>
<td>$P$</td>
<td>$P$</td>
<td>$P$</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$\rho$</td>
<td>$\rho$</td>
<td></td>
</tr>
<tr>
<td>$P$</td>
<td>$P$</td>
<td>$P$</td>
<td>---</td>
</tr>
</tbody>
</table>

---

Notes: *Statistically significant values ($P < 0.05$)
significant influence in the judgments of the jaws and the face in almost all cases. Furthermore, the upper and the lower lip distances from E-plane were the only parameters that were significantly correlated (negatively) with the assessment of the face by laypeople. Thus, dissatisfaction becomes greater while the distance of the lips from E-plane increases. It should be noted that this increase does not imply a significant lip protrusion compared to normal standards but relative lip protrusion. For example, in the specific study sample, the values for lower lip were close to normal standards, but the upper lip was in most cases significantly retruded (N Gkantidis, DA Papamanou, P Christou, N Topouzelis, unpublished data). Consequently, the relative lip protrusion in regards to E-plane may indicate a significantly retruded nose, and thus, their relative antero-posterior positions may comprise the feature that is judged more negative. These results underscore the importance of proper antero-posterior position of the lips in the face, but also relative to the nose, in influencing the appearance of cleft individuals.

On the other hand, a more anterior position of the maxilla, as determined by SNA, influenced positively the assessments of the jaws only for orthodontists and the aesthetics of the face for both specialists. These results may indicate that the position of the upper jaw is crucial for specialists, while laypeople in front of a unilateral cleft face tend to focus directly on the problem and its sequelae and are not much affected in their judgment by other parameters. In a similar manner, an increased vertical height was associated with poor jaw aesthetics according to all panels, while this was evident for the facial appearance only according to orthodontists.

Conclusions

Professionals report greater satisfaction with the aesthetic outcome of cleft lip and palate treatment and evaluate the cleft aesthetics more favourably than laypeople in this study.

From a cephalometric viewpoint, the position of the lips is correlated with the aesthetic perception in the eyes of laypeople, while specialists were focused on different cephalometric parameters, such as the antero-posterior position of the maxilla or the vertical dimension. This can account as a result of specialists’ education and experience which may lead to an evaluation of facial aesthetics from a different perspective compared to laypeople.

However, a specialist should consider the opinion of the patient as well as that of his/her social environment when planning treatment and not only what the specialist considers important according to his/her background.

Further research is needed to objectively detect the primary aesthetic needs of cleft individuals, which, along with the subjective needs defined by the patient, should determine the aim of the planned treatment interventions.

References


Friede H, Katsaros C 1999 Current knowledge in cleft lip and palate treatment from an orthodontist’s point of view. Journal of Orofacial Orthopedics 59: 313–330

Friedman M 1940 A comparison of alternative tests of significance for the problem of m rankings. The Annals of Mathematical Statistics 11: 86–92


Noor S N, Musa S 2007 Assessment of patients’ level of satisfaction with cleft treatment using the Cleft Evaluation Profile. The Cleft Palate-Craniofacial Journal 44: 292–303


rhinoplasty with versus without columellar grafting. Journal of Cranio-Maxillofacial Surgery Sep 8: [Epub ahead of print]
Svensson E 2000 Concordance between ratings using different scales for the same variable. Statistics in Medicine 19: 3483–3496