Original article

**Limitations of a method used for adolescent assessment of smile aesthetics**

Sophy K. Barber*, Nadine Houghton** and R. James Spencer*

*Department of Orthodontics, Leeds Dental Institute and Pinderfields Hospital, Wakefield and **Department of Orthodontics, Leeds Dental Institute and St. Lukes’ Hospital, Bradford, UK

Correspondence to: Sophy K. Barber, Department of Orthodontics, Leeds Dental Institute and Pinderfields Hospital, Leeds LS2 9LU, UK. E-mail: sophy.barber@googlemail.com

**Summary**

**Aims:** To establish whether adolescent orthodontic patients with hypodontia have a preference between the aesthetic outcomes of two treatment strategies for lateral incisor agenesis.

**Materials/Methods:** Standardized photographs of pre-orthodontic patients with missing lateral incisors were manipulated to produce images that represented space opening and tooth replacement in the lateral incisor space and space closure with canine substitution into the lateral incisor space. Adolescent orthodontic patients with hypodontia were recruited to assess the aesthetics of the images. A control group of subjects without tooth agenesis was recruited. Each examiner undertook two tests to assess the smile aesthetics of the images: (1) rating attractiveness using visual analogue scale (VAS) and (2) choice of preference between pairs of images.

**Results:** Difficulties experienced with image manipulation and poor intra-examiner reliability of the VAS make interpretation of the results challenging. Care should be taken if findings are used to aid clinical decisions, as the validity of the main findings is questionable. Results suggest that although adolescents perceive a difference in the aesthetic result of space opening and space closure for missing lateral incisors, the impact on the smile attractiveness is not clinically significant. When forced to choose between the aesthetics of space opening or space closure, the majority of examiners chose space opening with tooth replacement. No difference was found in smile ratings or preferences between adolescents with hypodontia and those with no missing teeth.

**Clinical implications:** The methods used in this study may not be reliable for adolescent assessment of aesthetics.

**Introduction**

Hypodontia is the developmental absence of one or more primary or secondary teeth excluding the third molars (1). Reported prevalence of hypodontia in the permanent dentition varies greatly but is accepted as being approximately 4–6 per cent in Caucasians, with a 3:2 female predilection (1, 2). The maxillary lateral incisor is the second most commonly missing tooth after mandibular second pre-molar, with a prevalence of 1.55–1.78 per cent. In contrast to all other teeth, bilateral agenesis occurred more commonly than unilateral agenesis in maxillary lateral incisors (3).

Three main treatment strategies exist for the management of missing lateral incisors:

1. No orthodontic treatment with no/minimal restorative modification of adjacent teeth.
2. Open and restore lateral incisor space.
3. Orthodontic space closure with canine substitution.

Historically, opening and restoring the lateral space was considered to be the optimum treatment as it restores the dentition to fulfill Andrews’ Six Keys of Occlusion (1972). It has been suggested that space opening may produce a superior result in patients with a Class III skeletal pattern with a retrusive, concave mid-face; with no crowding and stable Class I buccal segments; or with unfavourable canine characteristics, such as large, dark-coloured teeth with high gingival margins. Additionally, if there are poor prognosis teeth or...
other missing teeth in the quadrant space opening may be preferable (4–7). The main advantages are restoration of canine guidance and optimal posterior occlusion, no destruction to tooth tissue through re-shaping and potentially superior aesthetics. These must be balanced against the cost, reliability and long-term maintenance of the restoration, and the risk of long-term loss of alveolar bone width and height (8).

More recently, significant support has been expressed for orthodontic space closure, with substitution of the canines into the lateral incisor position and canine modification where required (7, 9, 10). The substantial advantage of maintaining a natural dentition without reliance on prosthetic work maximizes biological compatibility and stability and minimizes long-term maintenance. However, a number of common problems are associated with space closure and canine substitution (11, 12). These include colour and shape of the canine, relative size discrepancy of the premolar, differing gingival heights, crown torque differences, and an unstable final dentition prone to relapse with space reopening.

This study used digital images and computer-aided manipulation to assess the aesthetic outcome of the two methods for treating developmentally absent lateral incisors. Schabel et al. (13) demonstrated that standard digital photographs are valid tools for the analysis of the post-treatment smile compared to digital video clips. Image manipulation allows virtual treatment to assess the outcome, without any detrimental effects of interventive treatment. Small facial and dental changes within a single subject can be assessed while controlling all other variables (14).

The aim of this study was to establish whether adolescent orthodontic patients with hypodontia have a preference between the aesthetic outcomes of two treatment strategies for lateral incisor agenesis. The null hypotheses were as follows:

1. Adolescent orthodontic patients will not show a difference in ratings or preference between the aesthetic outcome of space opening with prosthetic replacement and space closure with canine substitution.
2. There is no difference in smile ratings or preference between examiners with hypodontia and those with no missing teeth.

**Materials and methods**

The study was undertaken in two hospitals in Yorkshire, UK. Seventeen pre-treatment orthodontic patients aged 11–16 years with unilateral or bilateral missing lateral incisors were recruited to act as photographic subjects. Any subjects with other absent anterior teeth, poor restorations, or severe aspects of malocclusion in the anterior region were excluded. Following enrolment and consent, standardized digital colour photographs were taken of their dentitions. The photographs were then manipulated using Adobe® Photoshop® CS2 software to produce two images to represent the two treatment strategies:

- Space opening and tooth replacement in the lateral incisor space (T1).
- Space closure with canine substitution into the lateral incisor space (T2).

A total of 34 images were produced. For T1, the prosthetic teeth were generated by copying the central incisors to mimic the tooth shape but with reduced dimensions. Minimal alterations were made to the existing gingival contour. The prosthetic tooth was given a slight grey hue to mimic the common appearance of resin-bonded bridges. The aesthetics of the pontic was created to be satisfactory, rather than excellent, to reflect current standards of tooth replacement. For T2, minimal canine modification was undertaken, again to reflect current practice in the UK where tooth bleaching, gingival procedures, and veneer restorations are not widely offered. The only modifications that were made were those that would mimic the results of orthodontic means of camouflage (e.g. tooth positioning) and some enameoplasty.

Facial photographs were taken of five Caucasian female subjects aged 13–14 years to produce a composite face, into which the altered dentitions were inserted. It was decided to use a female image as non-syndromic lateral incisor agenesis has been reported as more prevalent in females than in males. The final image was considered to be of average attractiveness, symmetrical, and with frontal facial proportions closest to clinically accepted norms. These values were based on findings from an extensive literature review of components of facial attractiveness.

In the second part of the study, orthodontic patients aged 11–16 years were recruited as examiners to complete the tests. The study group consisted of 51 orthodontic patients who had developmentally missing anterior teeth (maxillary or mandibular, canines or incisors), to compare to a control group of 51 patients without missing teeth. Parental consent and examiner assent was obtained.

Two PowerPoint presentations were made to allow two tests to be undertaken to assess the aesthetic outcome of the images. The presentations were both timed and ran through on a pre-determined timing sequence from the appearance of the first image and examiners were not able to look back through the images. Standardized verbal and written instructions were given to all subjects, but participants were not given any further background information regarding the study and the images. The following tests were undertaken:

1. Rating attractiveness of each smile using a visual analogue scale (VAS): 34 test images plus 4 repeat images to test reliability.
2. Forced choice between pairs of images: 17 pairs plus 3 repeat pairs to test reliability.

When rating with the VAS, the subjects were asked to rate how attractive each image was from highly unattractive to highly attractive by placing a vertical mark on an unmarked 100 mm scale. For theipsative pairs, the examiner was asked to choose between the pair of images created from each subject (T1 and T2).

Ethical approval was obtained prior to starting the study and a small pilot study involving four subjects who were not included in the study was conducted to determine whether the experimental procedure detailed above was satisfactory.

The primary method of assessing examiner preference was the VAS. A 15% VAS difference, 15 mm on a 100 mm scale, was taken as a clinically significant difference in this study, as previously used by Parkh et al. (15) and Ioi et al. (16). For a power of 90% and a significance level of 0.05, a sample size of 17 was calculated as necessary. Therefore, 17 original images were required for manipulation. For theipsative pairs, it was agreed that, for this study, a preference of more than 20% towards either treatment would be clinically significant. For a power of 80% and a significance level of 0.05, a sample size of 51 was calculated as necessary. Therefore, 51 examiners were needed in both the study group and the control group, giving a total of 102 examiners.

Appropriate statistical analysis was undertaken using IBM SPSS Statistics version 20.0 software. The following tests were undertaken:
Visual analogue scale
• Descriptive statistics for each image—mean, standard deviation (SD), and minimum/maximum score.
• Shapiro–Wilk test to test for normality.
• Bland–Altman plot for intra-examiner agreement and for examiner agreement in the VAS measurement.
• Paired t-test to compare T1 and T2 for all examiners combined.
• Independent t-test to compare scores for T1 and T2 for those with hypodontia (study group) against those without hypodontia (control group).

Ipsative pairs
• Cronbach’s alpha to test for intra-examiner agreement.
• Z-test to establish if there is a difference in the proportion of examiners who prefer T1 or T2.

Results
Data collection was undertaken from April to May 2013, providing a total of 102 participants to act as examiners. Examiner demographics are shown in Table 1.

Examiners were recruited consecutively, with all suitable patients being invited to participate. Interestingly, the number of male and female examiners with hypodontia was equal despite the condition having a female predilection. In the control group, the majority of participants were females (3:2), which may reflect the increased demand and provision of orthodontic treatment for females.

Examination of the data revealed no differences between genders in their scoring. The majority of participants were 14–16 years old, with a slightly older demographic in the hypodontia group.

Table 1. Demographics of the examiners.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Hypodontia group</th>
<th>Control group</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>25</td>
<td>13</td>
<td>38</td>
</tr>
<tr>
<td>Females</td>
<td>26</td>
<td>38</td>
<td>64</td>
</tr>
</tbody>
</table>

Table 2. Aesthetic scores from the visual analogue scale (VAS) for the outcome of space opening (T1) and space closure (T2) for all examiners (combined hypodontia and control group).

<table>
<thead>
<tr>
<th>Image number</th>
<th>Space opening (T1)</th>
<th>Space closure (T2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean score</td>
<td>SD</td>
</tr>
<tr>
<td>1</td>
<td>46.0</td>
<td>19.8</td>
</tr>
<tr>
<td>2</td>
<td>30.2</td>
<td>16.0</td>
</tr>
<tr>
<td>6</td>
<td>38.7</td>
<td>21.6</td>
</tr>
<tr>
<td>8</td>
<td>58.2</td>
<td>22.5</td>
</tr>
<tr>
<td>11</td>
<td>42.2</td>
<td>21.8</td>
</tr>
<tr>
<td>12</td>
<td>28.2</td>
<td>18.0</td>
</tr>
<tr>
<td>15</td>
<td>54.7</td>
<td>22.3</td>
</tr>
<tr>
<td>18</td>
<td>46.5</td>
<td>25.2</td>
</tr>
<tr>
<td>21</td>
<td>58.5</td>
<td>24.0</td>
</tr>
<tr>
<td>24</td>
<td>37.8</td>
<td>20.9</td>
</tr>
<tr>
<td>25</td>
<td>38.0</td>
<td>19.3</td>
</tr>
<tr>
<td>26</td>
<td>43.7</td>
<td>19.8</td>
</tr>
<tr>
<td>27</td>
<td>48.1</td>
<td>21.0</td>
</tr>
<tr>
<td>31</td>
<td>55.0</td>
<td>22.8</td>
</tr>
<tr>
<td>32</td>
<td>43.6</td>
<td>18.4</td>
</tr>
<tr>
<td>33</td>
<td>70.5</td>
<td>20.3</td>
</tr>
<tr>
<td>Total</td>
<td>46.1</td>
<td>10.7</td>
</tr>
</tbody>
</table>

SD, standard deviation.
Space opening (T1) versus space closure (T2), \( P = 0.007 \).
Hypodontia versus control: T1, \( P = 0.559 \).
Hypodontia versus control: T2, \( P = 0.919 \).

Bold italic values indicate highest rating and bold values indicate lowest rating.
adolescent orthodontic patients preferred the aesthetic result of space opening with prosthetic tooth replacement for managing missing lateral incisors, compared to space closure with canine substitution.

Chi-square test for association failed to find a statistically significant difference in the preferences expressed by the control and study groups ($P = 0.157$), indicating that there was no difference in smile preference between examiners with hypodontia and those with no missing teeth.

**Reliability**

Reliability of the VAS measurements was tested by re-measuring five examiners’ answers, providing a total of 190 measurements. The Bland–Altman plot shows that the agreement of the measurements was excellent (Figure 3). The mean error in measurement was 0.04 mm (SD 0.488).

The intra-examiner reliability of the VAS and ipsative pairs was obtained using repeat images within the PowerPoint. The Bland–Altman plot shows poor intra-observer reliability for the repeat images, with a high variability between the scores assigned to the original and repeat images (Figure 4).

Cronbach’s alpha indicated that the reliability of the ipsative pair preferences was acceptable. Although a low value was found (0.5), this may be attributed to the fact that correlation was only being tested between two variables. In this circumstance, it has been recommended to report the inter-item correlation, which was found to be within the optimal range of 0.2–0.4 (0.334) (17).

**Discussion**

The findings of both the aesthetic scores and the preferences indicate that examiners perceived a difference between the aesthetic result of space opening with tooth replacement and space closure with canine substitution. A statistically significant preference was expressed towards space opening although this was only deemed to be clinically significant in the ipsative pairs. Contrary to our findings, Armbruster et al. (18) found that orthodontists, dentists, and lay people all rated post-treatment photographs of people with missing lateral incisors who had undergone space closure and canine substitution higher than those who had prosthetic tooth replacement. Robertsson and Mohlin (19) assessed patients’ perception of their own dental appearance following either space opening or space closure for missing lateral incisors. The results indicated that generally patients who underwent space closure were more satisfied with the aesthetic result than those who had space opening and prosthetic replacement. However, the study subjects were older and aesthetics of the prosthetic replacements were not objectively judged.

The aesthetic scores and preferences awarded by the two groups of examiners were assessed separately to establish if adolescents with missing teeth have a different perception of smile aesthetics to those without missing teeth. The VAS scores and pair preferences were not found to be statistically or clinically different between the hypodontia and control group. This suggests that those with missing teeth are no more aware or critical of missing teeth than other orthodontic patients. A study assessing the impact of hypodontia on the OHRQoL in adults found that patients became more aware of their condition with age (20). It is possible that some of the younger examiners in our study were not yet concerned by their agenesis and as a result were no more critical of others.

These results must be interpreted with caution as two key areas of concern were identified with the methodology used. First, significant
difficulty was experienced with the image creation and as a result the images used were of questionable validity. Recruitment of suitable photographic subjects was difficult, not due to unwillingness to participate, but due to a lack of suitable cases being seen in the hospital departments in the study period. Some subjects had to be excluded as they were in the mixed dentition with unerupted permanent teeth, ectopic canines, or other aspects of malocclusion that precluded image manipulation. One subject had to be excluded as racial pigmentation in the gingiva made compatibility with the composite face difficult. The recruitment period was limited, and this difficulty in recruiting photographic subjects meant some inferior initial images had to be used.

Further complications were experienced during the manipulation of the original images. Rotations, tip errors, crowding, unerupted teeth or the presence of primary teeth and discrepancy in the inter-arch relationship and occlusion made image manipulation difficult. For less severe discrepancies, the teeth were moved to an acceptable position, but this was difficult for certain malocclusions and the resulting image was less aesthetic than would have actually been produced by treatment. In the cases, where the canines were moved mesially to replace the lateral incisors the resulting maxilla was often left narrow relative to the mandible. Attempts were made to manage this by magnifying the image using a vector that recreated the correct perception of the buccal corridor, but this was limited as excessive widening of the maxillary arch to correct inter-arch relationship led to enlargement of the teeth. In some cases, it was not possible to correct the dental shade to match the face due to slight differences.

### Table 3. Results from the ipsative pairs, showing the preference of the hypodontia and control groups for space opening (T1) and space closure (T2).

<table>
<thead>
<tr>
<th>Pair</th>
<th>Image number</th>
<th>Space opening (T1)</th>
<th>Space closure (T2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total, n</td>
<td>Hypodontia, n</td>
</tr>
<tr>
<td></td>
<td></td>
<td>87</td>
<td>42</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>75</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>93</td>
<td>48</td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>81</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>92</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>27</td>
<td>68</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>1</td>
<td>81</td>
</tr>
<tr>
<td>7</td>
<td>21</td>
<td>85</td>
<td>38</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>89</td>
<td>41</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>83</td>
<td>44</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>96</td>
<td>49</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>97</td>
<td>49</td>
</tr>
<tr>
<td>12</td>
<td>31</td>
<td>99</td>
<td>50</td>
</tr>
<tr>
<td>13</td>
<td>31</td>
<td>95</td>
<td>46</td>
</tr>
<tr>
<td>14</td>
<td>18</td>
<td>47</td>
<td>24</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>90</td>
<td>48</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>90</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1362 (78.5%)</td>
<td>685 (39.5%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>685 (39.5%)</td>
<td>677 (39%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>182 (10.5%)</td>
<td>190 (11%)</td>
</tr>
</tbody>
</table>

Preference between T1 and T2: \( P < 0.001 \).
Preference between those with hypodontia and controls: \( P = 0.157 \).
Values in bold italic indicates preferred image.

Figure 3. Bland–Altman plot for agreement of visual analogue scale (VAS) measurements.

Figure 4. Bland–Altman plot showing poor intra-examiner agreement in the visual analogue scale (VAS) scores.
in exposure, which restricted the amount of post-exposure modification that could be undertaken. The complications with image manipulation can be seen in the low scoring images (Figure 2), where there are problems with colour, tooth position, and inter-arch relationship.

The second concern was with the reliability and validity of the VAS. The VAS is widely reported in dental research, and it has been shown to be simple, easy to use in a variety of settings, and reproducible (15, 16, 21–30). However, the Bland–Altman plots that were used to assess intra-examiner agreement indicated poor reliability between the test–retest images. The repeat images were at the end of the tests and possible explanations for the variation in scores on the repeat images could be examiner fatigue or a change in the examiner scoring after seeing a number of images and dentitions. This should have been avoided by altering the order of the images, including the repeat images, during testing. The VAS has been used extensively for assessing perception of aesthetics (15, 16, 21–30), but in these studies the examiners were adults. Only one study that assessed aesthetics using a VAS involving adolescents aged 13–17 years could be found (31). The study reported the validity and reliability of results to be acceptable, although the exact reliability score was not reported.

The high frequency of maximum and minimum scores from the extreme ends of the scale suggests that the examiners may not have fully grasped the concept of the scale. Examiners were told that 0 represented ‘highly unattractive’ and should be only used for images that they felt were considerably worse than average, whereas 100 was ‘highly attractive’ and should be used for images that were felt were much nicer than an average smile. While the most and least attractive images may have fulfilled this description, it is possible that the examiners did not actually feel this strongly about all images that were given an extreme score. The wide range in scores for each image may have occurred because no calibration image was used prior to the testing, so each examiner had to set their own baseline score against which to measure images. A means for reducing this ‘self-calibration’ would have been to use a control image against which the examiners could score, as used in other studies (32, 33). This method was not used as the aim was to obtain an absolute, rather than relative judgement of attractiveness, but in hindsight the use of a reference image may have reduced the wide SD and improved reliability.

The second means of assessing aesthetic preference was through ipsative pairs, a type of forced-choice testing requiring respondents to rate images, during testing. The VAS has been used extensively for assessing perception of aesthetics (15, 16, 21–30), but in these studies the examiners were adults. Only one study that assessed aesthetics using a VAS involving adolescents aged 13–17 years could be found (31). The study reported the validity and reliability of results to be acceptable, although the exact reliability score was not reported.

The high frequency of maximum and minimum scores from the extreme ends of the scale suggests that the examiners may not have fully grasped the concept of the scale. Examiners were told that 0 represented ‘highly unattractive’ and should be only used for images that they felt were considerably worse than average, whereas 100 was ‘highly attractive’ and should be used for images that were felt were much nicer than an average smile. While the most and least attractive images may have fulfilled this description, it is possible that the examiners did not actually feel this strongly about all images that were given an extreme score. The wide range in scores for each image may have occurred because no calibration image was used prior to the testing, so each examiner had to set their own baseline score against which to measure images. A means for reducing this ‘self-calibration’ would have been to use a control image against which the examiners could score, as used in other studies (32, 33). This method was not used as the aim was to obtain an absolute, rather than relative judgement of attractiveness, but in hindsight the use of a reference image may have reduced the wide SD and improved reliability.

The second means of assessing aesthetic preference was through ipsative pairs, a type of forced-choice testing requiring respondents to make a choice between two or more options from the same scale with no neutral response allowed. Research has demonstrated increased validity in forced-choice testing due to a reduction in central tendency bias and social desirability responding (34, 35). Examiners reported that the ipsative pairs testing was easy to understand and acceptable, and the comparison of from repeat images this study show intra-examiner reliability to be acceptable. Unfortunately, as it is unclear if the VAS was a valid tool against which to compare, it is difficult to draw any conclusions or make recommendations about ipsative pairs as a method of testing.

Conclusions
1. Adolescent orthodontic patients expressed a preference in the aesthetic result of space opening with prosthetic tooth replacement for managing missing lateral incisors, compared to space closure with canine substitution. This was both statistically and clinically significant.
2. There was no difference in smile ratings or preferences between adolescents with hypodontia and those with no missing teeth.
3. Other aspects of malocclusion may limit image manipulation of pre-orthodontic dentitions. Difficulties in image manipulation may have compromised the validity of the findings of the study.
4. The VAS may not be valid or reliable for measuring perception of aesthetics in adolescents.

Acknowledgements
Many thanks to Colin O’Sullivan for his help with image creation. The research was undertaken as part of an MSc with sponsorship from the University of Leeds.

References


