Int. J. Oral Maxillofac. Surg. 2019; 48: 77–83
https://doi.org/10.1016/j.ijom.2018.07.010, available online at https://www.sciencedirect.com

International Journal of Oral & Maxillofacial Surgery

Research Paper
Orthognathic Surgery

Applying artificial intelligence to assess the impact of orthognathic treatment on facial attractiveness and estimated age

R. Patcas[1], D. A. J. Bernini[1], A. Volokitin[2], E. Agustsson[2], R. Rothe[2,3], R. Timofte[2,3]

[1]Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Plattenstrasse 11, 8032 Zurich, Switzerland; [2]Computer Vision Laboratory, D-ITET, ETH Zurich, Sternwartstrasse 7, 8092 Zurich, Switzerland; [3]Merantix GmbH, Skalitzer Str. 104, 10997 Berlin, Germany

R. Patcas, D. A. J. Bernini, A. Volokitin, E. Agustsson, R. Rothe, R. Timofte: Applying artificial intelligence to assess the impact of orthognathic treatment on facial attractiveness and estimated age. Int. J. Oral Maxillofac. Surg. 2019; 48: 77–83.

Abstract. This observational study aimed to use artificial intelligence to describe the impact of orthognathic treatment on facial attractiveness and age appearance. Pre- and post-treatment photographs ($n = 2164$) of 146 consecutive orthognathic patients were collected for this longitudinal retrospective single-centre study. Every image was annotated with patient-related data (age; sex; malocclusion; performed surgery). For every image, facial attractiveness (score: 0–100) and apparent age were established with dedicated convolutional neural networks trained on >0.5 million images for age estimation and with >17 million ratings for attractiveness. Results for pre- and post-treatment photographs were averaged for every patient separately, and apparent age compared to real age (appearance). Changes in appearance and facial attractiveness were statistically examined. Analyses were performed on the entire sample and subgroups (sex; malocclusion; performed surgery). According to the algorithms, most patients' appearance improved with treatment (66.4%), resulting in younger appearance of nearly 1 year [mean change: −0.93 years (95% confidence interval (CI): −1.50; −0.36); $p = 0.002$), especially after profile-altering surgery. Orthognathic treatment had similarly a beneficial effect on attractiveness in 74.7% [mean difference: 1.22 (95% CI: 0.81; 1.63); $p < 0.001$], especially after lower jaw surgery. This investigation illustrates that artificial intelligence might be considered to score facial attractiveness and apparent age in orthognathic patients.

Key words: orthognathic surgery; malocclusion; convolutional neural networks; attractiveness; facial appearance.

Accepted for publication 19 July 2018
Available online 4 August 2018

Orthognathic therapy focuses on the treatment of dentofacial discrepancies which are beyond the scope of conventional orthodontic treatment, particularly severe Class II and Class III, anterior open bite or facial asymmetry[1]. The objective of orthognathic treatment is to correct the functional and aesthetic impairments of dentofacial deformities through combined orthodontic and surgical efforts. While patients may present to an orthognathic clinic for a number of reasons, improvement of facial appearance clearly constitutes a prime concern[2–5].

Much research has been devoted to accurately evaluate the aesthetic outcome of orthognathic treatment. Improvements in facial appearance have been studied based on self-reported scores of patients[6–8] or orthodontists and maxillofacial surgeons appraising the changes in soft tissue proportions and facial aesthetics[9]. However, all historical approaches fall short to adequately address the assessment of *social* attractiveness[10]. Professional appraisal of attractiveness, as performed by orthodontists, surgeons or general dentists, relies on taught rules of beauty dictated by 'ideal' facial features, golden ratios and other established proportions[11–13]. These classic rules of facial harmony will, however, not reflect the attractiveness as perceived by peers[14–16]. Conversely, the assessment of attractiveness performed by a limited number of lay people remains equally inconclusive. The subjectivity involved is too important to be ignored[17,18] and may corrupt the evaluation altogether[10,19].

In recent years, the field of artificial intelligence (AI) has demonstrated some impressive advances, empowering computers not only to perform rudimentary cognitive functions such as optical facial recognition, but also to exceed in simulating much more complex cognitive tasks, including the *analysis and interpretation* of a recognized face. Hence, AI seems a promising tool to overcome the above-mentioned concerns related to the evaluation of facial attractiveness. Though it is admittedly very difficult to quantify beauty, AI enables from a single face image an assessment of attractiveness by characterizing the attractiveness of particular facial traits, and their combinations. Moreover, the appraisal of these traits may be utilized to calculate an apparent age. Employing an algorithm based on a convolutional neural network trained on big data that mirrors relevant opinion may indeed prove helpful in objectively and reproducibly interpreting facial appearance.

It seems that AI has never been applied to assess clinical results in dentistry in general, and facial changes in orthognathic patients in particular. The objectives of this present study were (1) to assess the effect of orthognathic therapy on facial attractiveness and apparent age, by applying a dedicated algorithm, validated on a large dataset from a dating site, on pre-and post-treatment facial photographs, and (2) to review the observable changes according to gender, underlying malocclusion and chosen osteotomy. Due to the exploratory nature of this investigation, no specific hypotheses were formulated.

Material and Methods

Material

This was a retrospective longitudinal cohort study of the most recent 150 consecutive patients who completed orthognathic therapy at the local university. All records were collected from the archives of the Clinic of Orthodontics and Paediatric Dentistry. Patients with craniofacial syndromes, cleft lip and palate, reported previous maxillofacial surgery or facial traumata were excluded. The collected records included facial photographs and variables related to patient (sex; age; type of malocclusion) or treatment (type of surgery; duration of treatment). Overall, four cases had to be excluded due to partially missing data of the patient-related variables.

Pre-treatment malocclusions were assessed on cephalograms and categorized as (1) skeletal Class II, (2) skeletal Class III, (3) anterior open bite and (4) asymmetry (cases could be assigned to one or more categories). The types of surgery were grouped in the following categories: (1) Le Fort I osteotomy of the maxilla (upper jaw surgery), (2) sagittal split ramus osteotomy of the mandible (lower jaw surgery), (3) chin osteotomy and (4) other osteotomies (cases could be assigned to one or more categories).

All photographs were taken with a single-lens reflex camera, a dedicated flash reflector and against a monochrome background. The standardized images were taken both initially and after completion of treatment and consisted of several viewing angles (frontal, profile, 45° oblique) and different characteristics (resting posture, smile, habitual occlusion). Apart from altering contrast or brightness, no digital image enhancement was performed. Of the 2628 possible combinations, 2164 annotated facial images were retrieved, digitally archived, processed in JPEG format at a resolution of 600 dpi, and used for the statistical analysis. Missing images were disregarded for analyses.

Method

Apparent age and facial attractiveness were determined by applying a computational algorithm comprising a face detector[20] and convolutional neural networks (CNNs) for the prediction of apparent age[21] and facial attractiveness[22].

Prior to CNN prediction, faces were detected in all images, roughly aligned, and each image was cropped to ensure that all faces were equal in size, orientation and position. All face images were brought to a size of 256×256 pixels with a centred face and a 40% background margin and used as input for the CNN models. The algorithms did not explicitly use facial landmarks. The CNN models employed VGG-16 architecture[23] and were pre-trained on >0.5 million facial images with age labels acquired from the Internet Movie Database and Wikipedia (IMDb-Wiki, age range: 0–100 years)[21].

For apparent age prediction, the CNN model was fine-tuned on the APPA-REAL[24] face images with apparent age labels (age range: 0–95 years).

For attractiveness prediction, the CNN model was fine-tuned using a dataset from a dating site containing >13,000 face images with more than 17 million ratings for attractiveness[22]. Since images from the employed dataset were taken in conditions dissimilar to medical assessment, the pre-trained attractiveness prediction network was further adjusted, using the Chicago Face Dataset (CFD)[25]. The latter contains 597 photographic images all taken in identical lighting conditions against a monochrome background, and each labelled for facial attractiveness. For this task, CFD was partitioned into 469 train and 128 validation images, binning the attractiveness scores into four bin-classes and fine-tuning the CNN model for classification. The expected attractiveness score was subsequently computed using attractiveness = $\sum_{i=0}^{3} i \times$ *probability of class i*.

To facilitate the interpretation, facial attractiveness was scaled from 0 to 100 (0: extremely unattractive; 100: extremely attractive).

As shown in recent investigations[22,24,26], trained CNN models are not only sensitive to facial features, but also to background. In order to suppress background distractors, the margin of the image (i.e. a 30-pixel-wide rim) was set to black for both apparent age and facial attractiveness prediction.

Apparent age and attractiveness scores were established for every single facial photograph ($n = 2164$), and were subsequently averaged for every patient before treatment and after treatment, respectively.

Statistical analysis

Data were analysed in SPSS software [IBM SPSS Statistics for Windows, version 23 (IBM Corp., Armonk, NY, USA]. A Kolmogorov–Smirnov test was applied on all variables to investigate normal distribution, and depending on the test's outcome parametric or non-parametric descriptive analysis and statistical testing were operated. Real age, computed apparent age, appearance and computed facial attractiveness were descriptively reviewed. Appearance was defined as difference between computed apparent age and real age, and the patient's appearance was considered as looking older when apparent age was greater than real age. Changes in appearance before and after treatment were calculated and the impact of treatment on appearance was analysed with a paired Student's *t*-test.

Scores of facial attractiveness before and after treatment were compared to each other, and the difference in attractiveness was assessed with a Wilcoxon signed-rank test.

These statistical analyses were performed for the entire sample and independently repeated for certain subgroups according to gender (males; females), type of underlying malocclusion (Class II, Class III, anterior open bite, facial asymmetry), or type of treatment (upper jaw surgery, lower jaw surgery, combined upper and lower jaw surgery with or without chin osteotomy).

To evaluate possible associations between changes in attractiveness and age at start of treatment or attractiveness at start of treatment, respectively, scatter plots were created using data of the entire sample, and two separate models for linear regression analysed. For all analyses, differences were deemed significant at $p < 0.05$.

Results

The pre-treatment malocclusions of the 146 orthognathic patients (females: $n = 77$; 52.7%) were categorized as Class II ($n = 62$; 42.5%), Class III ($n = 68$; 46.6%), anterior open bite ($n = 57$; 39.0%) and asymmetry ($n = 42$; 28.8%). In most cases, upper jaw surgery ($n = 124$; 84.9%) and lower jaw surgery ($n = 122$; 83.6%) was performed. Chin osteotomy was carried out in nearly half of the patients ($n = 71$; 48.6%), and other different surgical approaches were chosen in around 10% of the cases ($n = 15$; 10.3%). Thus, most cases were treated with multiple osteotomies [upper and lower jaw surgery with chin osteotomy ($n = 53$; 36.3%) or without (n=48; 32.9%)]. The minority of the cases were treated exclusively in the upper jaw ($n = 15$; 10.3%) or solely in the lower jaw ($n = 21$; 14.4%).

The descriptive values of real age and computed apparent age are listed in Table 1. According to the algorithm used, orthognathic patients appeared in general older than their actual age. Most patients' appearance improved with treatment (66.4%). Table 2 describes the changes in appearance. On average, orthognathic patients appeared 1.75 years older than their actual age before treatment, but only 0.82 years older after treatment. This statistically significant improvement (i.e. reduction of the difference between apparent age and real age) was, however, not equally reproducible in all subgroups. In particular, females benefited more than males, as did Class II and Class III patients. Combined surgical treatment (lower and upper jaw surgery) seemed to achieve the most beneficial effect with regard to appearance.

Treatment also had a positive impact on attractiveness (Fig. 1, Table 3). Altogether, 74.7% of the patients were more attractive after treatment. Improvement in attractiveness was observed in both sexes and across all malocclusions, and the combined therapy was most likely to produce the best effect in terms of attractiveness. Descriptive results, however, reveal small effect sizes and may relativize the clinical relevance.

Two possible associations were investigated with linear regression analysis. The first model in Fig. 2A indicates that changes in attractiveness were unlikely connected to the age at start of treatment [regression coefficient: −0.01 (95% CI: −0.06; 0.03); p = 0.57]. The second model in Fig. 2B discloses a relationship between the improvement in attractiveness with the *attractiveness* at start of treatment [regression coefficient: −0.10 (95% CI: −0.17; −0.02); p = 0.02]. Hence, greater improvement in attractiveness could be expected in patients with lower attractiveness score at start of treatment.

Discussion

As outlined in the introduction, improvement of facial attractiveness is an essential aspect when appraising the outcome of orthognathic therapy. However, facial attractiveness is elusive[13] and its scoring

Table 1. Real (chronological) age and apparent (computed) age, at start and end of treatment. Descriptive values (median, first and third quartiles) are given for entire sample and specific subgroups.

		At start of treatment		At end of treatment	
	n (percentage of total)	Real age (years) Median (Q1; Q3)	Apparent age (years) Median (Q1; Q3)	Real age (years) Median (Q1; Q3)	Apparent age (years) Median (Q1; Q3)
All	146 (100.0%)	19.17 (17.08; 27.35)	24.02 (19.35; 28.49)	22.34 (20.21; 30.21)	26.00 (22.05; 31.28)
Males	69 (47.3%)	19.04 (17.67; 26.96)	23.90 (19.96; 28.63)	22.29 (20.54; 29.84)	26.85 (22.98; 31.86)
Females	77 (52.7%)	19.63 (16.25; 28.50)	24.14 (18.90; 27.88)	22.46 (19.88; 30.59)	25.31 (21.53; 30.91)
Subgroups according to malocclusion					
Class II	62 (42.5%)	19.17 (16.69; 26.92)	23.81 (19.00; 28.07)	22.25 (20.38; 29.23)	25.78 (21.75; 30.62)
Class III	68 (46.6%)	18.88 (17.42; 27.63)	23.68 (20.00; 28.40)	21.80 (20.21; 30.44)	24.90 (21.87; 30.90)
Anterior open bite	57 (39.0%)	18.79 (17.04; 24.38)	22.02 (18.35; 26.44)	21.88 (20.13; 26.88)	25.36 (21.20; 29.39)
Asymmetry	42 (28.8%)	19.04 (17.27; 27.81)	24.46 (19.85; 28.10)	22.17 (20.27; 30.59)	26.30 (23.12; 31.20)
Subgroups according to treatment					
Upper jaw	15 (10.3%)	18.13 (16.13; 25.46)	19.94 (17.16; 26.15)	21.63 (20.04; 28.79)	21.54 (19.27; 28.78)
Lower jaw	21 (14.4%)	18.54 (16.75; 29.13)	24.41 (19.75; 29.73)	21.46 (19.84; 31.71)	26.66 (23.81; 31.39)
Combined	48 (32.8%)	21.00 (18.02; 27.10)	24.75 (21.40; 28.63)	23.75 (20.90; 30.59)	26.70 (23.52; 31.79)
Combined with chin	53 (36.3%)	18.38 (16.50; 27.50)	23.72 (9.15; 28.12)	21.63 (19.75; 29.59)	24.85 (21.90; 31.09)

Table 2. Appearance, based on calculated differences (mean and standard deviation) between real age and apparent age, at start and end of treatment.

	n	Appearance at start of treatment Difference between real age and apparent age (y)	Appearance at end of treatment Difference between real age and apparent age (y)	Impact of therapy on appearance	
		Mean (±1 SD)	Mean (±1 SD)	Paired *t*-test *P*	(years) Mean (95% CI)
All	146	1.75 (±3.68)	0.82 (±4.37)	0.002*	−0.93 (−1.50; −0.36)
Males	69	2.13 (±2.87)	1.48 (±3.59)	0.060	−0.64 (−1.30; 0.03)
Females	77	1.42 (±4.28)	0.22 (±4.91)	0.011*	−1.20 (−2.11; −0.28)
Subgroups according to malocclusion					
Class II	62	1.82 (±3.93)	0.39 (±5.14)	0.005*	−1.43 (−2.41; −0.45)
Class III	68	2.07 (±3.29)	1.09 (±3.67)	0.010*	−0.98 (−1.71; −0.24)
Anterior open bite	57	1.97 (±3.35)	1.37 (±3.56)	0.123	−0.60 (−1.37; 0.17)
Asymmetry	42	2.19 (±3.57)	1.54 (±3.86)	0.153	−0.65 (−1.56; 0.25)
Subgroups according to treatment					
Upper jaw	15	0.45 (±3.01)	−0.19 (±3.88)	0.492	−0.64 (−2.59; 1.30)
Lower jaw	21	2.03 (±3.86)	1.54 (±4.65)	0.464	−0.48 (−1.82; 0.86)
Combined	48	1.82 (±3.60)	0.80 (±4.53)	0.011*	−1.02 (−1.80; −0.25)
Combined with chin	53	1.99 (±3.92)	0.98 (±4.23)	0.077	−1.02 (−2.14; 0.11)

Positive values signify that the apparent age is greater than the real age (''older appearance''). The changes in appearance are tested with Student's paired *t*-tests for the entire sample and specific subgroups, and given as mean difference with corresponding 95% confidence interval (CI). A negative mean value relates to a reduction of an older appearance.

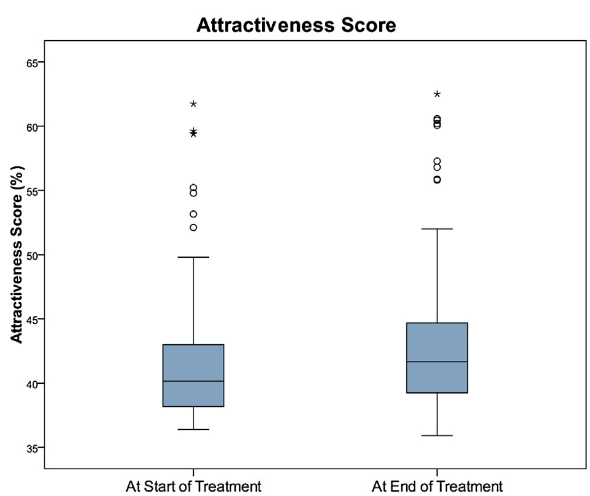

Fig. 1. Box and Whisker plot for attractiveness score across entire sample ($n = 146$). Box depicts the interquartile range (IQR), with the median indicated by the internal line. Whiskers enclose the lowest to highest values excluding the outliers (i.e. values within 1.5 × IQR of the lower and upper quartiles). Stars mark extreme outliers (i.e. value more than 3 × IQR beyond the upper quartile); circles mark mild outliers.

inherently problematic; as both lay people's and professional's assessments suffer from serious limitations in mirroring social attractiveness. The purpose of this present research was to overcome the subjectivity common in all traditional rating protocols by means of AI, using an algorithm trained on large data and fine-tuned for medical assessment. With a model that has been validated on 17 million ratings retrieved from a dating site, the approach can be considered a robust technique to reflect *social* attractiveness. Additionally, this investigation is presumably the very first attempt ever made in dentistry to introduce AI to identify not only changes in attractiveness, but also in apparent age, a task in which the human reference is evidently outperformed[21]. But perhaps most importantly, the introduced method would be a welcome additional clinical tool. In contrast to panel-based scores, which are unavailable for individual planning, the use of AI would allow the surgeons to predict the outcome of surgical procedures on the appearance of patients. As such, AI would allow eliminating the inherent subjectivity in the planning and possibly obtaining more favourable aesthetic outcomes.

The sample, consisting of >2100 images of 146 represented patients, is clearly adequate for the analysis performed. Both the types of malocclusions and the surgery performed were sufficiently represented and evenly distributed across both sexes. Mean age, both at start and at the end of the therapy, as well as the duration of 3.1 years for treatment, can all be considered representative for a typical cohort of orthognathic patients.

The analysis demonstrated that orthognathic treatment significantly improves facial attractiveness, both in males (mean difference: 1.22; $p < 0.001$) and females (mean difference: 1.33; $p < 0.001$). The evaluation of the subgroups revealed a differential effect relative to the treated malocclusion; the highest score for attractiveness improvement was documented for corrected asymmetries, followed by Class II and Class III patients. This observation is in agreement with previous studies attesting a beneficial aesthetic outcome for corrected asymmetries[27], Class II and Class III patients[28]. Moreover, the correction of facial asymmetry indisputably constitutes the change most tangible for patients (in contrast to any other profile-related dysgnathia). Thus, the fact that the correction of asymmetry produced the highest increase in attractiveness score underlines the appropriateness and usefulness of the AI-based scoring.

The assessment of different surgical approaches showed equally that not all

Table 3. Computed attractiveness at start and end of treatment.

	n	Attractiveness at start of treatment (%) Median (Q1; Q3)	Attractiveness at end of treatment (%) Median (Q1; Q3)	Comparison between median attractiveness at start and at end of treatment Wilcoxon signed-rank test	Difference in attractiveness (%) Mean (95% CI)
All	146	40.16 (38.17; 43.03)	41.68 (39.23; 44.69)	<0.001*	1.22 (0.81; 1.63)
Males	69	40.35 (38.10; 44.24)	42.52 (39.60; 44.70)	<0.001*	1.33 (0.70; 1.95)
Females	77	39.94 (38.44; 42.31)	41.03 (39.15; 44.62)	<0.001*	1.13 (0.28; 1.68)
Subgroups according to malocclusion					
Class II	62	40.03 (38.14; 42.59)	41.60 (39.18; 44.14)	<0.001*	1.12 (0.59; 1.66)
Class III	68	39.67 (38.17; 44.06)	41.31 (39.23; 44.67)	0.001*	0.98 (0.35; 1.61)
Anterior open bite	57	39.42 (38.06; 41.83)	41.63 (38.66; 44.81)	<0.001*	1.23 (0.63; 1.84)
Asymmetry	42	40.80 (38.18; 44.03)	42.26 (39.49; 44.69)	<0.001*	1.49 (0.73; 2.25)
Subgroups according to treatment					
Upper jaw	15	38.16 (37.36; 41.96)	39.69 (37.10; 41.63)	0.173	0.74 (−0.69; 2.17)
Lower jaw	21	41.37 (37.95; 42.59)	42.15 (40.15; 44.60)	0.006*	1.53 (0.55; 2.52)
Combined	48	40.29 (38.89; 44.06)	42.77 (39.73; 44.73)	<0.001*	1.33 (0.67; 2.00)
Combined with chin	53	40.17 (38.10; 42.75)	41.58 (39.11; 44.83)	0.001*	1.07 (0.29; 1.86)

Descriptive values (median, first and third quartiles) are given for entire sample and specific subgroups. The difference in attractiveness (between at start and at end of treatment) is tested by Wilcoxon signed-rank test for the entire sample and specific subgroups. The resulting difference in attractiveness is shown (mean difference and 95% confidence interval (CI)).

* Statistically highly significant at $p < 0.01$.

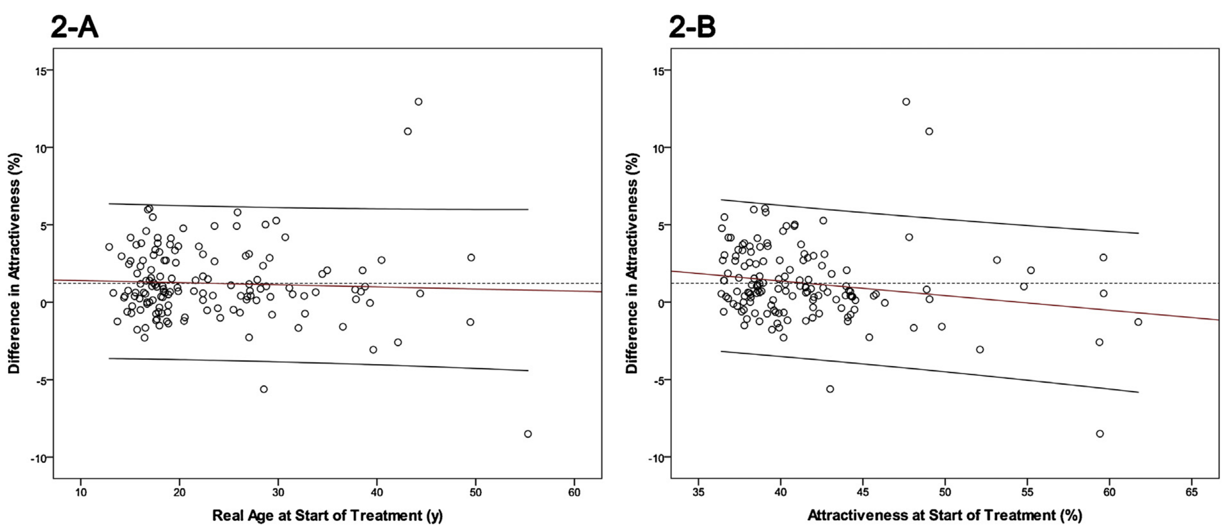

Fig. 2. Scatter diagram with changes in attractiveness caused by the treatment (y-axis) plotted against real (chronological) age at start of treatment (A) and attractiveness at start of treatment (B), respectively. The linear regression (red line) is drawn together with its 95% confidence interval (black lines). The black dotted line represents the mean improvement of attractiveness (y = 1.22). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

osteotomies produce identical improvements in facial aesthetics. It is evident, and clearly clinically relevant, that the various osteotomies in the lower jaw produced the most significant beneficial facial changes (mean differences: 1.07–1.53; $p < 0.006$), yet single jaw surgery in the maxilla failed to affect facial attractiveness significantly (mean difference: 0.74; p = 0.173).

Although pre- and post-treatment results differed significantly, the estimated effect size (given as mean difference) remained, perhaps somewhat surprisingly, rather small. The following considerations might be submitted. First, ageing was disregarded. While it is an established fact that facial attractiveness diminishes with age[29], the outcome of orthognathic treatment was compared to pre-treatment attractiveness. Thus untreated, the patients would have suffered a steady decrease simply owing to ageing, and this effect is insufficiently accounted for when comparing to pre-treatment score. Second, it is a noted phenomenon seen in numerous investigations that the use of a discrete scale to score intra-individual changes of facial aesthetics in orthognathic patients ordinarily results in small quantitative increments[9,16,27].

With regard to age appearance, people with severe malocclusions looked older than their real age, an observation more accentuated in males. Orthognathic therapy was able to reduce, but not to bridge, the gap between apparent and real age. The treatment had a significant impact, especially on age appearance of female patients, resulting in a younger facial appearance after orthognathic therapy (age appearance: −1.2 years; p = 0.011). Interestingly, patients with malocclusions affecting their profile (Class II and Class III) appeared significantly younger after treatment, in contrast to patients with malocclusions affecting chiefly their frontal appearance (anterior open bite and asymmetries), for whom the juvenescent effect remained insignificant. In terms of treatment modality, the greatest impact on age appearance was achieved through combined surgery in both jaws. The general finding that orthognathic treatment may alter age appearance is anything but elementary. While one of the primary intentions in plastic surgery is indisputably to change age appearance, orthognathic surgery is mainly focused to achieve a balanced facial harmony. The observation that mandibular sagittal split ramus osteotomy may influence age appearance is therefore novel and seminal.

Regression analyses revealed that changes in attractiveness were not linked to age at start, but to attractiveness at start. Hence, patients with impaired initial attractiveness are especially subject to significant improvement, an observation independently corroborated by an earlier study[12].

Applying AI to assess clinical results is indeed a promising venture, as it potentially enables to overcome several major pertinent drawbacks. At the same time, the introduction to medicine of a new and unique algorithm must be evaluated with great caution.

First, while clinicians may benefit greatly from an AI-based assessment with regard to treatment planning, it will never replace the patient's own perceptions and expectations, which remain of primordial importance. The clinician's responsibility is to inform the patient in an honest and realistic way about the aesthetic outcome, in order not to elicit illusory expectation. To achieve this task, AI seems to be indeed a helpful tool.

Second, the algorithm was used in connection with social media[22], but never against the backdrop of medical interventions. Patients undergoing orthognathic treatment might identify certain features as important and worth correcting, while some of these features could be underrepresented in the model. This obviously is not a methodological shortcoming, but rather a general observation that dissimilarities between the subjective patient's view and the computed score could exist. Or, to put it more simply: having mastered a system to objectively assess treatment outcome does not necessarily mean that orthognathic patients themselves will think accordingly.

Lastly, the appropriateness of using data retrieved from a dating platform to qualify attractiveness should be discussed. Attractiveness is generally defined as the quality to cause interest and desire in the observer. As such, subjectivity and cultural influences are indeed an inherent part of the definition. Attractiveness can surely be scored by panel (of laypeople, artists, surgeons, Caucasians, etc.), but every panel will remain a representation of its observers. Albeit precise, measurements based on AI are just a quantifiable representation of a particular opinion. But what do the AI-based results exactly represent? This is indeed a thought-provoking question, and probably best answered with coining the results as 'social attractiveness', i.e. the quality to cause interest and desire in our present globalized society. Based on millions of ratings retrieved from a globalized dating site, validated and fine-tuned on medical images, the proposed AI-score is unquestionably a fitting tool to mirror social opinion on treated patients. And perhaps this is what should be considered most important for patients. Treatment outcome should not be measured by specific panels (or historic and cultural definitions of attractiveness), but by how society views the aesthetic results. As such, scores of a dating site are probably most appropriate to train the algorithm to discern facial qualities that cause interest and desire in the observer.

This study is the first of its kind to use AI to analyse the impact of orthognathic treatment on facial attractiveness and age appearance. Outperforming past approaches, this investigation offers a new key which permits to score facial attractiveness and apparent age objectively and reproducibly. The results are to be considered clinically relevant, as they highlight the need to differentiate between the malocclusions with which the patients present. Clinicians should be aware that lower jaw surgery has a far greater impact on attractiveness than osteotomies in the maxilla, and that age appearance can especially be altered beneficially in patients with profile-related malocclusions. Finally, orthognathic treatment in patients with low initial attractiveness is more likely to produce greater aesthetic outcome than in other patients with similar malocclusions.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests

The authors have no conflicts of interest to disclose.

Ethical approval

Ethical guidelines were strictly followed, and irreversible anonymization was performed in accordance with the Declaration of Helsinki and in compliance with State and Federal Law. Ethical approval was given by governmental ethics committee (reference no.: BASEC 2016-00990).

Patient consent

Written informed consent for secondary use of the patients' data including their facial photographs was obtained from all patients (and their legal guardians) prior to treatment.

References

1. Cunningham SJ, Johal A. Orthognathic correction of dento-facial discrepancies. *Br Dent J* 2015;**218**:167–75.
2. Oland J, Jensen J, Elklit A, Melsen B. Motives for surgical-orthodontic treatment and effect of treatment on psychosocial well-being and satisfaction: a prospective study of 118 patients. *J Oral Maxillofac Surg* 2011;**69**:104–13.
3. Patcas R, Cunningham SJ, Shute J, Lloyd T, Obwegeser JA, Arjomand L, Sharma S. Motivation for orthognathic treatment and anticipated satisfaction levels-a two-centre cross-national audit. *J Craniomaxillofac Surg* 2017;**45**:1004–9.
4. Nurminen L, Pietila T, Vinkka-Puhakka H. Motivation for and satisfaction with orthodontic-surgical treatment: a retrospective study of 28 patients. *Eur J Orthod* 1999;**21**:79–87.
5. Ryan FS, Barnard M, Cunningham SJ. Impact of dentofacial deformity and motivation for treatment: a qualitative study. *Am J Orthod Dentofac Orthop* 2012;**141**:734–42.
6. Hunt OT, Johnston CD, Hepper PG, Burden DJ. The psychosocial impact of orthognathic surgery: a systematic review. *Am J Orthod Dentofac Orthop* 2001;**120**:490–7.
7. Schwitzer JA, Albino FP, Mathis RK, Scott AM, Gamble L, Baker SB. Assessing patient-reported outcomes following orthognathic surgery and osseous genioplasty. *J Craniofac Surg* 2015;**26**:2293–8.
8. Islam S, Aleem F, Ormiston IW. Subjective assessment of facial aesthetics after maxillofacial orthognathic surgery for obstructive sleep apnoea. *Br J Oral Maxillofac Surg* 2015;**53**:235–8.
9. Storms AS, Vansant L, Shaheen E, Coucke W, de Llano-Perula MC, Jacobs R, Politis C, Willems G. Three-dimensional aesthetic as-

sessment of class II patients before and after orthognathic surgery and its association with quantitative surgical changes. *Int J Oral Maxillofac Surg* 2017;**46**:1664–71.
10. Jesani A, DiBiase AT, Cobourne MT, Newton T. Perceived changes by peer group of social impact associated with combined orthodontic-surgical correction of class III malocclusion. *J Dent* 2014;**42**:1135–42.
11. Edler R, Agarwal P, Wertheim D, Greenhill D. The use of anthropometric proportion indices in the measurement of facial attractiveness. *Eur J Orthod* 2006;**28**:274–81.
12. Proffit WR, Phillips C, Douvartzidis N. A comparison of outcomes of orthodontic and surgical-orthodontic treatment of Class II malocclusion in adults. *Am J Orthod Dentofac Orthop* 1992;**101**:556–65.
13. Harrar H, Myers S, Ghanem AM. Art or Science? An evidence-based approach to human facial beauty a quantitative analysis towards an informed clinical aesthetic practice. *Aesthetic Plast Surg* 2018;**42**:137–46.
14. Knight H, Keith O. Ranking facial attractiveness. *Eur J Orthod* 2005;**27**:340–8.
15. Kiekens RM, Kuijpers-Jagtman AM, van't Hof MA, van't Hof BE, Straatman H, Maltha JC. Facial esthetics in adolescents and its relationship to ideal ratios and angles. *Am J Orthod Dentofac Orthop* 2008;**133**. 188 e1–188 e8.
16. Chung EH, Borzabadi-Farahani A, Yen SLK. Clinicians and laypeople assessment of facial attractiveness in patients with cleft lip and palate treated with LeFort I surgery or late maxillary protraction. *Int J Pediatr Otorhinolaryngol* 2013;**77**:1446–50.
17. Bell R, Kiyak HA, Joondeph DR, McNeill RW, Wallen TR. Perceptions of facial profile and their influence on the decision to undergo orthognathic surgery. *Am J Orthod* 1985;**88**:323–32.
18. Yin L, Jiang M, Chen W, Smales RJ, Wang Q, Tang L. Differences in facial profile and dental esthetic perceptions between young adults and orthodontists. *Am J Orthod Dentofac Orthop* 2014;**145**:750–6.
19. Maple JR, Vig KW, Beck FM, Larsen PE, Shanker S. A comparison of providers' and consumers' perceptions of facial-profile attractiveness. *Am J Orthod Dentofac Orthop* 2005;**128**:690–6.
20. Mathias M, Benenson R, Pedersoli M, Van Gool L. Face detection without bells and whistles. *Paper presented at: European Conference on Computer Vision*. 2014:720–35.
21. Rothe R, Timofte R, Van Gool L. Dex Deep expectation of apparent age from a single image. *Paper presented at: Proceedings of the IEEE International Conference on Computer Vision Workshops* 2015:10–5.
22. Rothe R, Timofte R, Van Gool L. Some like it hot-visual guidance for preference prediction. *Paper presented at: Proceedings CVPR* 2016;**2016**:1–9.
23. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 2014.
24. Agustsson E, Timofte R, Escalera S, Baro X, Guyon I, Rothe R. Apparent and real age estimation in still images with deep residual regressors on APPA-REAL database. *Paper presented at: Automatic Face & Gesture Recognition (FG 2017)12th IEEE International Conference 2017 IEEE* 2018:87–94.
25. Ma DS, Correll J, Wittenbrink B. The Chicago face database: a free stimulus set of faces and norming data. *Behav Res Methods* 2015;**47**:1122–35.
26. Rothe R, Timofte R, Van Gool L. Deep expectation of real and apparent age from a single image without facial landmarks. *Int J Comput Vision* 2016;**126**:144–57.
27. Blockhaus M, Kochel J, Hartmann J, Stellzig-Eisenhauer A, Meyer-Marcotty P. Three-dimensional investigation of facial surface asymmetries in skeletal malocclusion patients before and after orthodontic treatment combined with orthognathic surgery. *J Orofac Orthop* 2014;**75**:85–95.
28. Sinko K, Jagsch R, Benes B, Millesi G, Fischmeister F, Ewers R. Facial aesthetics and the assignment of personality traits before and after orthognathic surgery. *Int J Oral Maxillofac Surg* 2012;**41**:469–76.
29. Kathleen MK, Irene T. Perceived age and perceived physical attractiveness. *Percept Mot Skills* 1982;**54**:1251–8.

Address:
Raphael Patcas
Clinic of Orthodontics and Pediatric Dentistry
Center of Dental Medicine
University of Zurich
Plattenstrasse 11
8032 Zurich
Switzerland
Tel.: +41 (0)44 634 32 89
E-mail: raphael.patcas@zzm.uzh.ch