The term evidence-based medicine (EBM), first cited in the literature in 1992, has since become of paramount importance in the health care industry. A number of medical journals (e.g., Core Evidence, Evidence-Based Healthcare, International Journal of Evidence-Based Healthcare) focus solely on the subject of EBM. One of the prerequisites for practicing EBM is an objective system for outcomes measurement, but the field of aesthetic surgery lacks such an algorithm. This is understandable because patient satisfaction is the ultimate measurement of outcome and is composed of innumerable subjective factors, but an objective system for measuring surgical outcomes would support research in the field of aesthetic surgery. Specifically, variations in techniques and other surgical parameters could be rigorously compared and more evidence-based recommendations could be made. As stated by Chung et al, “Plastic surgery should lead in the evidence movement or others will be more than happy to lead us.” The advent of an objective outcome measurement system, particularly in facial aesthetic surgery, seems like a remote (if not impossible) prospect. In our opinion, the first step is defining what makes a human face attractive.

The field of attractiveness psychology has made significant advances in the past 15 years, particularly regarding the concept of composite (or averaged) faces. The original studies by Langlois et al in 1990 and 1994 and Rhodes et al in 1996 found that composite faces are generally more attractive than the individual faces from which they are created. Elaborating on this concept, further studies found that composites are more attractive when they are composed of faces that are more attractive versus those that are composed of all faces.

Safa E. Sharabi, MD; Daniel A. Hatef, MD; and Larry H. Hollier Jr., MD

Abstract

Background: In order for aesthetic surgery to participate in the movement toward evidence-based medicine, a system that quantifies aesthetic outcomes must be developed based on reliable and reproducible data. Previous studies have investigated the attractiveness of averaged (or composite) faces. Data has shown that these faces are deemed more attractive than their constituent faces.

Objectives: The authors seek to elucidate the mechanism behind the phenomenon in which more attractive faces included in a composite increase the overall attractiveness rating.

Methods: Ten composites, five male and five female, were generated with resources provided by the www.faceresearch.org research group. The five composites contained the top two, three, and five most attractive faces, a composite of all faces, and the five least attractive faces. An online survey was conducted in which participants were asked to rank each of the five faces in order of attractiveness.

Results: Analysis of variance (ANOVA) on ranks was performed on the data provided by the 245 participants. Attractiveness increased from the top two to the top three to the top five composite faces, and then decreased from the top five to all faces to the bottom five composite faces. This trend was present and statistically significant (with a P value <.05) across all genders, ages, and races.

Conclusions: This study provides statistically significant evidence that averaging more attractive faces, specifically the top 10% of a population, renders a more attractive composite face. This optimal composite face could potentially be a standard to which aesthetic surgical outcomes could be objectively compared.

Keywords

facial attractiveness, aesthetic surgery, facial aesthetic surgery, facial composites

Accepted for publication July 6, 2009.
In a study by Perrett et al in 1994, two facial composites were generated with digital imaging software. The first composite was of the entire study population. The second composite was of the most attractive 50% of the study population. The most attractive 50% of the study population rendered a more attractive composite face. Interestingly, using digital image subtraction, the authors determined the differences between these two composite faces and then used digital imaging software to exaggerate these differences. The product of this exaggeration was more attractive than either of the two composite faces.

The study by Perrett et al is the basis for the current consensus, which is that average faces are attractive, but perhaps not optimally so. There are some unusual features that deviate from the norm and add to the overall attractiveness of the face. Averaging more attractive faces pools together more of these unusual features, rendering the composite face more attractive. If this is, in fact, true, an optimal facial composite must exist for any given population that combines the benefits of averaging all faces and the benefits of averaging only attractive faces. This composite face would have the property that both adding and subtracting faces from the composite would decrease the overall attractiveness. The current consensus, based on the study by Perrett et al, does not specify the level at which adding more faces adds to or detracts from the overall attractiveness.

To provide objective data to test this hypothesis and to locate the point at which optimal attractiveness exists, we conducted a blinded study in which subjects were asked to rank different composite faces, each composed of a different number of faces, in order of attractiveness. The optimal composite face in both male and female facial composites was the composite of the top five (approximately the top 10%) most attractive faces in each population. Composites of more and fewer faces were statistically significantly less attractive.

**METHODS**

Photographs of 43 women and 48 men (age range, 20-30) were identified using the www.faceresearch.org resource (Aberdeen School of Psychology, University of Aberdeen, Scotland, UK). Subjects were of varying race. A pilot study was performed that identified the five most attractive and five least attractive faces for both men and women. Using these data, five composites were generated for each sex. They consisted of the following combinations of faces: the top two most attractive, top three most attractive, top five most attractive, the average of all faces, and the five least attractive (see Figures 1 and 2).

An online, anonymous survey was administered that prompted the participants to rank the composites on a scale of one (most attractive) to five (least attractive). The study was blinded in that the participants were not aware of the composite nature of the individual photographs. The photographs were arranged horizontally in a random order. The participants were prompted to rate the five male faces in order of attractiveness, followed by the five female faces. The participants were then prompted for their gender, age range, ethnicity, and status as a medical student, resident, fellow, faculty, or other participant. All of the demographic questions were optional.

In total, 510 subjects were invited to participate in the online survey. The medical student body, plastic surgery residents, fellows, faculty, and other staff members at Baylor College of Medicine were invited via email to participate in the survey.

**RESULTS**

Of the 510 invitees, 245 (48%) completed the survey. Of these, 126 (51.4%) were men and 119 (48.6%) were women. Medical students comprised 83.5% of the total; 4.7% were plastic surgery residents, fellows, or faculty; 1.3% were other residents, fellows, or faculty; and 10.6% did not fit into any of these categories. The age distribution was as follows: 0.8% were younger than 21 years old, 90.3% were 21 to 30 years old, 5.5% were 31 to 40 years old, 0.4% were 41 to 50 years old, 2.1% were 51 to 60 years old, and 0.8% were older than 60. Caucasians comprised 61.9% of the participants; 3.4% were African American, 20.3% were Asian, 0.8% were Native American, 3.4% did not wish to disclose their race, and 10.2% identified with none of the above groups.

The mean rank values and 95% confidence intervals (CI) were calculated for each face based on the responses from all participants. The mean rank value, which is a number between one and five, can be viewed as a measure of attractiveness, with one being the most attractive and five being the least attractive. For the male faces, the top two composite’s mean rank was 3.225 (95% CI: 3.070-3.380), the top three composite’s mean rank was 3.029 (95% CI: 2.914-3.144), the top five composite’s mean rank was 1.525 (95% CI: 1.426-1.624), the “all faces” composite’s mean rank was 2.634 (95% CI: 2.475-2.793), and the bottom five composite’s mean rank was 4.558 (95% CI: 4.445-4.671). For the female faces, the top two composite’s mean rank was 3.424 (95% CI: 3.265-3.583), the top three composite’s mean rank was 2.440 (95% CI: 2.308-2.572), the top five composite’s mean rank was 1.863 (95% CI: 1.735-1.991), the “all faces” composite’s mean rank was 2.826 (95% CI: 2.657-2.995), and the bottom five composite’s mean rank was 4.422 (95% CI: 4.302-4.542). Viewed as a trend, in both male and female faces and in all of the subgroups, the composites increased in attractiveness from the top two to the top three to the top five composites and then decreased in attractiveness from the top five to all faces to the bottom five composites (see Figures 3 and 4).

Kruskal-Wallis one-way analysis of variance (ANOVA) on ranks was performed with the Dunn method and revealed that the mean ranks (attractiveness) for all of the male facial composites were unique with a P value <.05, except for the top two and top three composites, which were not significantly different. The same analysis of the female facial composites revealed that the mean ranks (attractiveness) of all composites were statistically unique, with a P value <.05 for all pairwise comparisons. In other words, the differences in attractiveness between each facial composite in male and female faces were all statistically significant, except for the top two and top three composites in men.
A Student’s $t$ test comparing male and female survey participants revealed no significant differences between the male and female responses for any given facial composite (see Figures 5 and 6). A Student’s $t$ test comparing the responses of participants between 21 and 30 years old and those older than 30 revealed that there were no statistically significant differences between the two groups’ choices, except for the top three composite in men (see Figures 7 and 8). One-way ANOVA on ranks was performed to analyze the differences in the survey results for participants of different races. Caucasians, African Americans, and Asians had no statistically significant differences in their choices for any facial composites (see Figures 9 and 10).

**DISCUSSION**

The Phi Mask, developed by Dr. Stephen Marquardt, is a mathematical model of the prototypical face. It has been suggested that this model could be the basis for an

---

objective system for measuring outcomes in facial aesthetic surgery. Bashour demonstrated that when the mask was imposed onto a face, deviation from the dimensions of the mask did correlate negatively with attractiveness. However, this correlation was not very strong and only accounted for, at most, 25% of variation of facial attractiveness. The authors stated that one of the reasons for the weak correlation was the fact that the Phi Mask relies on the “attractiveness is averageness” hypothesis.

The results of this study demarcate a statistically significant trend that is consistent among both male and female faces and deviates from the “attractiveness is averageness” hypothesis. That is, the optimally attractive composite is composed of the top five attractive faces, or

approximately the top 10% of the given population. Adding faces to and subtracting faces from this optimal composite both decreased the attractiveness rating. This is counterintuitive because even subtracting unattractive faces (e.g., from the top five to the top three composite) decreased the attractiveness.

The most unattractive composites in men and women, however, were the composites of the five least attractive faces, or the bottom 10%. This displays a mirror opposite response to the effect of averaging the top 10% of the population. These results support the theory that, although average is attractive, optimal attractiveness is the average of more attractive (in this study, the top 10%) faces.

The concept of a universal standard of attractiveness across race, gender, and age of the perceiver is controversial. Previous studies, however, have maintained this concept. In this study, controlling for race, gender, and age of the study participants did not reveal any differences in the trends of attractiveness of the composites. This study, therefore, supports the concept that a universal standard for attractiveness does in fact exist across races, genders, and age groups. Further studies would be helpful to determine the trends in attractiveness in faces of different age groups. This would be especially helpful to the aesthetic surgeon because...
attractiveness in older individuals is more likely to be related to soft tissue arrangement, which is currently the most feasible target of the facial aesthetic surgeon.

The argument can be made that the goal of aesthetic surgery is to make a patient more satisfied with his or her appearance, not to make the patient look more like a certain prototype. Therefore, it is important to distinguish that the results of this study support the development of a universal standard for outcomes measurement and not a tool on which operative planning for a specific patient should be based. For example, when performing a brow lift, the specific parameters of the operation should be based on the characteristics...
unique to the patient, their goals, the surgeon, and the situation. However, when comparing brow lift techniques on a large scale, the results of this study can be used as a standard upon which to base outcomes, thereby introducing an evidence-based method to an otherwise subjective field.

The evidence generated in this study is an objective analysis of subjective opinions. Of course, what determines whether a face is attractive to a particular person is the result of innumerable and intangible factors. The subjective opinion of an individual in no way represents substantial data upon which to base clinical decisions. However, a highly statistically significant trend among persons of different ages, ethnicities, and genders approaches, if not embodies, evidence-based material in the field of aesthetic surgery.

Previous studies that have utilized composites generated by digital imaging software have referred to the “smoothing effect” associated with averaging images. For example, it is unlikely that two faces will have blemishes in the same location. Therefore, when the faces are averaged, the blemishes will fade, yielding a more smooth appearance. In 1994, Langlois et al performed a study that controlled for this variable and determined that the smoothing effect is not responsible for the attractiveness of averaged faces. However, this variable was not controlled in our study.

Another potential bias of computer-generated composite faces is the “familiarity effect,” purporting that averaged faces are more likely to seem familiar to the viewer. Also, familiar faces are more likely to be attractive to the viewer. Previous studies have postulated that the attractiveness of averaged faces may be due to their familiarity. Although this variable was not controlled in this study, our results do not show a consistent increase in attractiveness with an increase in the number of faces in the composite. This phenomenon cannot be the sole cause of our results.

There are certain biases inherent in this study. First, a selection bias is present because 89.4% of the subjects who completed the study are in some way associated with the medical field. Second, 61.9% of the population surveyed was Caucasian. In order to determine the absolute effects of race on perception of attractiveness, equal numbers of subjects from each race would need to be surveyed. Similarly, 99.1% of the subjects surveyed were between the ages of 21 and 30. For the effects of age on perception of attractiveness to be studied rigorously, equal numbers of subjects of each age range would need to be surveyed.

CONCLUSIONS

The results of this study support the concept that there is a particular relationship between the number of faces included in a facial composite and its attractiveness. Composite faces are more attractive than their component parts up to a certain point (in this study, up until the top 10% of the study population was included), after which the attractiveness decreases with additional faces and reaches its nadir when the five (10%) least attractive faces are averaged. This information, although rudimentary in comparison to the ultimate goal of developing an objective system for measuring outcomes in facial aesthetic surgery, provides an initial effort in determining what an attractive face entails. Combining the results of this study with the concept of the Phi Mask, we postulate that generating a model of facial dimensions that is a representation of the top 10% of the population, rather than the average of the population, is the first step toward an objective system for measuring outcomes in facial aesthetic surgery.

Disclosures

The author(s) declared no conflicts of interests with respect to the authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

REFERENCES

7. Alley TR, Cunningham MR. Average faces are attractive, but very attractive faces are not average. Psychol Sci 1991;2:123-125.
13. Marquardt SR. Marquardt beauty analysis. Available at: www.beautyanalysis.com