University of Groningen

Beauty Is in the Eye of the Beholder

van Veen, Martinus M; Broekstra, Dieuwke C; Mureau, Marc A M; Werker, Paul M N; Dijkstra, Pieter U

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Link to publication in University of Groningen/UMCG research database

Beauty Is in the Eye of the Beholder: Esthetic Outcome Assessment in Smile Reanimation Surgery in Patients With Facial Palsy

Martinus M. van Veen, MD,* Dieuwke C. Broekstra, MSc, PhD,* Marc A.M. Mureau, MD, PhD,†
Paul M.N. Werker, MD, PhD,* and Pieter U. Dijkstra, PT, PhD‡§

Abstract: Layperson assessments are becoming increasingly important in the evaluation of surgical procedures of the face, including smile reanimation. In this study, the authors set out to answer 3 questions: (1) are esthetic scores more dependent on the assessor or the person that is being assessed, (2) how does smile reanimation change esthetic scores, (3) do sex and age of the patient and assessor explain some of the esthetic outcomes?

Thirty-five assessors scored pre and postoperative photographs of 21 facial palsy patients undergoing smile reanimation. Linear mixed-effect models were used to investigate the effects of assessor and patient factors on esthetic outcome assessments, to examine changes after smile reanimation, and to determine whether sex and age explained part of the esthetic outcomes.

Fifty-eight percent of variation in the esthetic scores can be explained by some assessors being more positive in their esthetic scoring compared to other assessors. Twenty-nine percent was attributed to patient baseline esthetic scores. Overall esthetic scores improved after smile reanimation. Sex and age of the patient and assessor could not explain variation in the esthetic scores.

Esthetic appearance highly depends on "who is looking." These findings are important for preoperative counseling, and for those treating and educating patients with facial palsy.

Key Words: cosmetic, esthetic, facial palsy, facial reanimation, layperson

(*J Craniofac Surg* 2021;32: 159–163)

Layperson assessments are becoming increasingly important in the evaluation of cosmetic and reconstructive surgical procedures of the face, including smile reanimation surgery in patients with facial palsy.[1,2] Layperson assessments offer the ability to analyze effects of these procedures, amongst others on attractiveness,[3–5] disfigurement,[6,7] and trustworthiness[7] from a more societal perspective. The major benefit of using laypersons for these assessments is that the researcher is able to include the perspective of many different assessors at the same time, representing the general population.

Assessment of attractiveness and disfigurement after cosmetic and reconstructive surgical procedures is inherently subjective and dependent on the assessor and patient. Some assessors will overall be more positive than others in their assessment and some patients are generally more attractive than others.[4,8] Additionally some assessors may rate some patients higher while other assessors rate them lower. How much of the variation in these assessments can be attributed to the assessor, the patient, and their interactions – or an intervention – is unknown. Knowledge about the (relative) influence of assessor and patient factors on esthetic outcome scores is needed to place treatment effects into perspective, and to be able to adequately counsel patients seeking treatment for facial palsy.

In the present study, we performed an analysis of assessments of the esthetic outcomes of smile reanimation surgery in longstanding facial palsy. We set out to answer 3 questions:

(1) are esthetic scores more dependent on the assessor or the person that is being assessed,

(2) how does smile reanimation surgery change esthetic scores, and

(3) do sex and age of the patient and assessor explain some of the esthetic outcomes?

METHODS

This study was a repeated assessment of photographs of patients treated in 2 tertiary university hospitals in the Netherlands, collected for a previously published study.[9] The institutional review boards of both centers did not deem formal approval necessary before the start of this study since the patients included received standard care (Erasmus MC 2016-699/ UMCG 2016.383).

Assessors and Patients

Thirty-five adult laypersons acted as assessors. They were visitors of the outpatient plastic surgery clinic of the University Medical Center Groningen – consulting a physician for a health condition unrelated to their esthetic appearance (eg, thumb base osteoarthritis) – and family or friends accompanying them. Among these 35 assessors were 5 persons consulting a physician for facial palsy. They only participated in this study as "assessors" and were not included as "patient." All assessors were naïve to the purpose of the study. After completion of the assessments they were informed

about the aim of the study. Laypersons were chosen as the assessors in our study since they represent the view of the common man, and we hypothesize that the majority of the social anxiety that comes with facial palsy results from these encounters and encounters with friends and family or doctors or other health professionals.

Adult facial palsy patients that had been treated with either a free functioning gracilis muscle transplantation (gracilis FFMT) neurotized by the masseteric nerve (Erasmus MC Rotterdam) or a modified temporalis muscle transposition (University Medical Center Groningen) were identified and approached for visiting our institution to provide photographs used in this study.[9] If a patient wanted to participate but was not able to visit our institution, most often due to a large distance from our institution, the patient was visited at home by one of our investigators, ensuring adequate quality of the photographs. Unwillingness to participate at all and the lack of a set of preoperative photographs were exclusion criteria.

For clarification: the "assessors" in this study are the laypersons who viewed the photographs and scored them, the "patients" are the persons with facial palsy whose photographs were used. There was no overlap between "assessors" and "patients."

Procedure

A PowerPoint slide show of 21 preoperative and 21 postoperative slides of the facial palsy patients was presented to the assessors. They were asked to rate the esthetic appearance of the face. Each slide showed 3 photographs of the face simultaneously: one in repose, one with a closed mouth smile, one with a maximum smile. A combined rating of the 3 photographs on each slide was asked on a 100 millimeter visual analogue scale (VAS), with zero standing for "absolutely not beautiful" and 100 standing for "absolutely very beautiful." The term "beautiful" (Dutch: *mooi*) was chosen opposed to "attractive" in an attempt to minimize possible bias due to sexual orientation. All 35 assessors rated all slides, in a random order, resulting in a total of 1470 observations. Sex, age, type of smile reanimation surgery and the follow-up time since smile reanimation surgery was collected from patients. Additionally, sex and age of the assessors was recorded.

Statistical Analysis

Descriptive statistics were calculated in IBM Statistical Package for the Social Sciences (SPSS) version 23 (IBM, NY) and given as frequencies and percentages in case of nominal data, and medians and interquartile ranges in case of continuous data due to non-normality. Normality was checked by evaluating histograms. Data-handling, visualization and analyses were done in R (packages lme4,[10] sjstats,[11] lmerTest,[12] haven,[13] dplyr,[14] and ggplot2[15]), software version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) to analyze the effect of smile reanimation on esthetic outcome, and the influence of assessor and patient characteristics. An in-depth description of our statistical analyses can be found in the supplementary material (Supplementary material 1, http://links.lww.com/SCS/B873).

For answering the first research question (Are esthetic scores more dependent on the assessor or the person that is being assessed?) several linear mixed-effect models with different structures regarding the assessors and patients (and combinations) were compared. The best model was chosen based on model fit (Akaike Information Criterion, AIC, and Bayesian Information Criterion, BIC). Intra-class correlation coefficients were then calculated, indicating the amount of variation in the outcome that can be attributed to the assessor or the patient.

To establish whether or not gracilis muscle transplantations or modified temporalis muscle transpositions resulted in improved postoperative esthetic results (research question 2), the previously

chosen model was extended with a variable for the type of surgery (gracilis FFMT versus temporalis transposition). The effect was corrected for "patient age," "patient sex" and "follow-up duration" since the time since reanimation varied for all patients and was hypothesized to be off influence on the outcome.

Lastly, to determine whether patient age, patient sex, assessor age, and assessor sex – and interactions between these patient and assessor characteristics – influenced the assessment scores (research question 3), the model was extended with variables for patient and assessor age, and patient and assessor sex, and all possible interactions. Again, model fit criteria (AIC and BIC) were used to determine if the addition of a variable was a significant improvement.

RESULTS

Patients were described in detail elsewhere.[9] Assessors were more often male compared to patients. Assessors and patients were of similar age (Supplementary Digital Content, Table 1, http://links.lww.com/SCS/B872).

A large variation in the esthetic scores was observed at both the assessor and the patient level. Some assessors demonstrated much more homogeneous assessments (ie, they gave a similar esthetic score to all patients) (Fig. 1) than other assessors (Fig. 2). Interestingly, some assessors rated strong improvement of a patient while other assessors rated strong deterioration of that same patient (Figs. 3 and 4). Six unconditional mean models were established and compared for the best random effects structure (Supplementary Digital Content, Table 2, http://links.lww.com/SCS/B872). A model including random intercepts for assessor and patient, and a random slope for patient had best model fit (model D). We also fitted models including a random slope for assessor and 2 random slopes for patient and assessor. However, these models were singular or did not converge, both indicating that the model specification was too complex for the data. In model D, 58% of the variation in the esthetic scores could be explained by clustering at the assessor level, 29% by clustering at the patient level, and an additional 2% by the random slope for patients. This means that a large majority of variation in esthetic scores (58%) could be attributed to some assessors generally being more positive in their esthetic scoring of patients compared to other assessors. Some patients having a higher baseline esthetic score accounted for 29% of the variation, and only 2% of the variation was explained by allowing patients to vary in their in- or decrease in esthetic scores.

FIGURE 1. Line graphs showing all observations of a single assessor. Each line depicts an individual patient (21 lines in each graph) connecting the preoperative score to the postoperative score. Shown are the observations of one assessor with rather homogeneous and consistent observations.
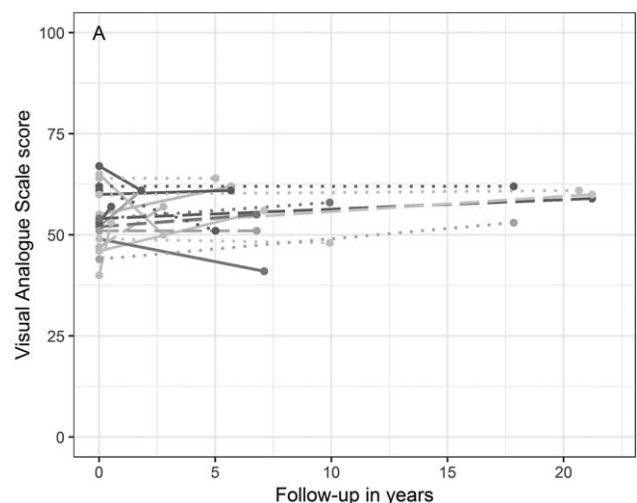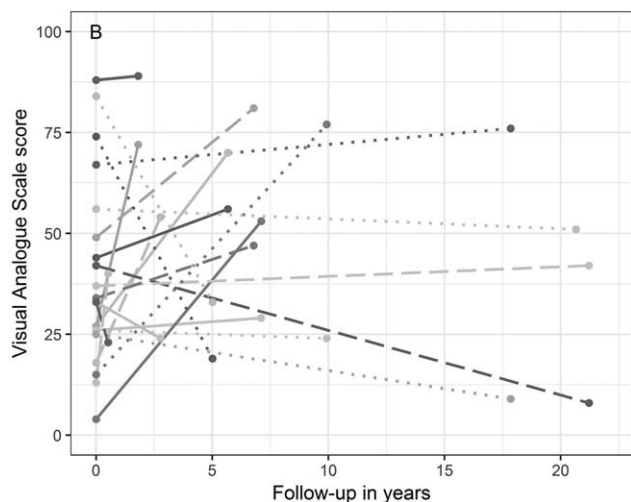
FIGURE 2. Line graphs showing all observations of a single assessor. Each line depicts an individual patient (21 lines in each graph) connecting the preoperative score to the postoperative score. Shown are the observations of one assessor with more variety in the esthetic assessments of the patients.

The type of smile reanimation surgery was not significantly associated with esthetic scores. A significant positive effect of the length of follow-up on the esthetic scores was found (Supplementary Digital Content, Table 3, http://links.lww.com/SCS/B872). Esthetic scores increased on average 1.9 points per year after reanimation.

Finally, the effects of age and sex of the assessor and patient on the esthetic scores were examined. Since "surgery type" was not significantly associated to the esthetic outcome it was not incorporated in the models; follow-up duration was included since it was significantly associated to the outcome. We observed that none of the added patient and assessor factors, nor any interactions between these factors, improved model fit (Supplementary Digital Content, Table 4, http://links.lww.com/SCS/B872), meaning that age and sex of the patient or assessor did not affect the esthetic scores. Neither were interactions between age and sex of the patient and assessor
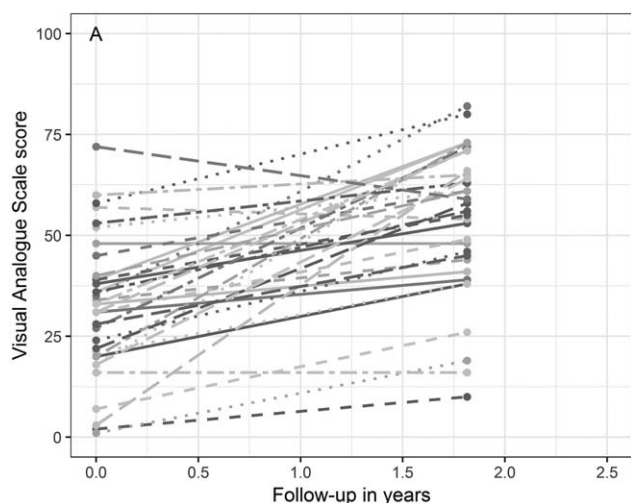

FIGURE 3. Line graphs showing all observations of a single patient. Each line depicts an individual assessor (35 lines in each graph) connecting the preoperative score to the postoperative score. Much variation in the assessments can be seen within a single patient: some assessors show strong improvement of the esthetic appearance, while other assessors show strong deterioration for the same patient. Shown is example patient A.

FIGURE 4. Line graphs showing all observations of a single patient. Each line depicts an individual assessor (35 lines in each graph) connecting the preoperative score to the postoperative score. Much variation in the assessments can be seen within a single patient: some assessors show strong improvement of the esthetic appearance, while other assessors show strong deterioration for the same patient. Shown is example patient B.
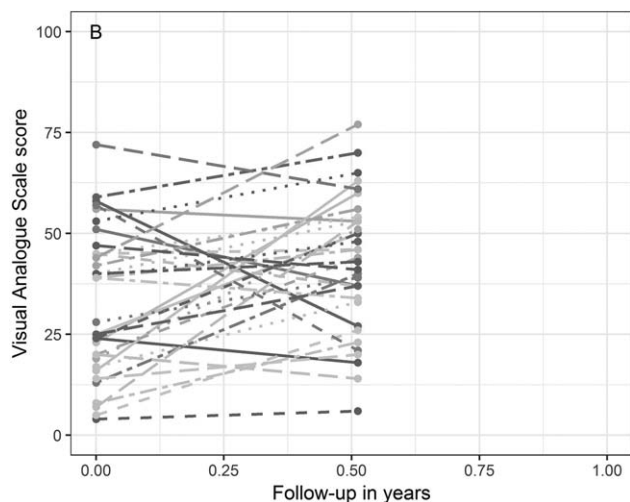
present, meaning that men and women did not rate women and men differently, and that younger assessors did not rate older patients differently than younger patient and vice versa. A model examining a difference in assessments between assessors with and without facial palsy did not converge, indicating that this model specification was too complex for the data. On visual inspection of the line graphs of each patient, the assessments of assessors with facial palsy were not systematically different from assessments of assessors without facial palsy (Supplementary Digital Content, 2, http://links.lww.com/SCS/B874).

DISCUSSION

This study aimed to answer 3 questions:

(1) what is the effect of assessors and patients on esthetic outcome assessments,
(2) how does smile reanimation surgery change esthetic scores, and
(3) do sex and age of the patient and assessor explain some of the esthetic outcomes?

We found that the data from our study were highly clustered. In the optimal unconditional mean model, we found that 58% of the variation in esthetic scores of the patient could be explained by including a random intercept for assessor. This means that 58% of the differences in esthetic scores could be attributed to the fact that some assessors in general rated esthetic appearance of the patients higher compared to other assessors. Clinically this means that patients should be aware of the fact that how they are seen by others highly depends on who is looking and depends much less on themselves. Although there obviously are baseline beauty characteristics – also present in the explained variance of 29% by the patient intercept – the clustering at assessor level was twice as high as at patient level. This finding puts a limit on what the effect of a facial palsy surgeon can be on the patient's esthetic appearance and perception by others. From a research standpoint, the high level of clustering on assessor level means that results from studies involving only one or few assessors cannot directly be generalized to the population at large. Additionally, the high level of data clustering requires advanced statistical methods to analyze this data.

When looking at the effect of smile reanimation surgery, we observed an increase in esthetic appearance after smile reanimation surgery. Others have reported an increase in attractiveness after smile reanimation surgery.[4] We also observed that the esthetic outcome was better with increasing duration of follow-up. However, we only had 2 observation moments. Therefore, trajectories of changes in esthetic appearance in smile reanimation surgery other than linear could not be investigated. Potentially a patient's control over his or her smile continues to increase with time, resulting in better esthetic scores after a longer follow-up time.[4,16] We also examined whether the type of smile reanimation surgery would make a difference in the esthetic outcome. We could not find a statistically significant difference between both types of surgery, meaning that the assessors did not rate one of both procedures as resulting in a better esthetic appearance postoperatively. We believe this indicates that the donor site morbidity of temporalis transposition is not more severe than the swelling created by gracilis FFMT, although no other comparative studies evaluating esthetic outcomes of types of smile reanimation surgery have been published that could support these findings. On the other hand, our small sample size could have limited the power of finding a difference between both procedures.

As a third step we analyzed whether part of the clustering at the assessor and patient level could be attributed to differences between men and women or was related to the assessor's or patient's age. We examined all possible combinations of sex and age of the assessor and patient, and the interactions between those, and could not find an improvement of the model. The high level of clustering may be related to other factors that we did not measure. Some recent research suggests that patient and assessor ethnicity may play a role,[17] but all patients and virtually all assessors in the current study were Caucasian (reflecting the demographic composition of the region[18]). Hence, we believe ethnicity has not played a major role in the current study. Potentially, psychological factors may also partly explain clustering at assessor level. Women's own attractiveness and dissatisfaction with their appearance were found to independently predict how they judge other women.[19] Self-perceived attractiveness or body image of the assessor could have explained part of the clustering at the assessor level. Additionally, it is likely that character traits such as perfectionism could explain some assessors being more critical in their assessments than others. Lastly, factors such as mood or happiness at the time of assessing may have been of influence. Hypothetically, assessors in a good mood were more positive in their assessments.

We were unable to examine the fit of unconditional models including a random slope for assessor and 2 random slopes for patient and assessor. For the complexity of the multilevel structure of our data, a larger dataset would have been ideal. The relatively small data set also limited us from examining a difference in assessments from assessors with facial palsy compared to those without. If patients with facial palsy are more or less critical in their assessment of other patients with facial palsy compared to assessors without facial palsy could not be examined, although inspection of the line graphs did not show large or obvious differences.

Another limitation is the separation of both reanimation procedures over the 2 participating centers. This means that the effect of the operation can partly be explained by center effects. However, the peri and postoperative procedures of both centers were very similar, therefore we believe the influence of center effects was limited. For future studies comparing reanimation procedures, it would be better to include both procedures at the same center performed by the same surgeon and a random allocation to the procedures. On the other hand, the procedure in our study resembles clinical practice more closely: each surgeon will perform his or hers preferred surgical procedure.

Our assessors were a mixed group of visitors to the outpatient plastic surgery clinic of the University Medical Center Groningen: both patients with a health condition unrelated to their esthetical appearance and family or friends accompanying them. Although we believe this resulted in a group of assessors fairly representative of the general population – approximately half of the assessors were female and median (interquartile ranges) age was 59.0 (42.0; 68.0) years – selection bias from sampling in a hospital cannot be excluded. Furthermore, we chose to use laypersons as assessors, since we wanted to study the view of the "common man" as explained in the methods. Including doctors for example could change the results of this study since they might be either more critical, because they are trained in looking at facial palsy or facial appearance, or less critical, because they are familiar with the disorder and the extent of severity it can have. Our results cannot be extrapolated to other types of assessors.

Postoperative photographs of 4 patients were taken at their home by one of the investigators, since they were not willing to visit our institution because of a larger distance from the institution. Although photographs were taken with great care, in the same standardized way used in our institution, by the investigator trained by the medical photographer, the photographs of these patients undoubtedly introduced additional variation.

Although the current study has some limitations, we believe the findings presented contain important information for clinicians working in the field of facial plastic surgery. The observation that 58% of the variance in esthetic appearance can be explained by who is looking, opposed to the patient themselves, is novel and important for preoperative counseling and treating and educating those with a facial deformity.

CONCLUSION

The perception of esthetic outcomes varies greatly depending on the assessor: that is, 58% of the variation in esthetic scores can be attributed to the assessor versus only 29% to the patient. Therefore, beauty truly is in the eye of the beholder. Esthetic outcome studies based on one or few assessors are not necessarily generalizable to the general population of patients or to the population of assessors. What factors explain this assessor clustering remains unknown and should be topic for future studies.

REFERENCES

1. Ishii LE. Moving toward objective measurement of facial deformities: exploring a third domain of social perception. *JAMA Facial Plast Surg* 2015;17:189–190
2. Dusseldorp JR, van Veen MM, Mohan S, et al. Outcome tracking in facial palsy. *Otolaryngol Clin North Am* 2018;51:1033–1050
3. Ishii L, Godoy A, Encarnacion CO, et al. Not just another face in the crowd: society's perceptions of facial paralysis. *Laryngoscope* 2012;122:533–538
4. Dey JK, Ishii M, Boahene KDO, et al. Changing perception: facial reanimation surgery improves attractiveness and decreases negative facial perception. *Laryngoscope* 2014;124:84–90
5. Dey JK, Ishii LE, Nellis JC, et al. Comparing patient, casual observer, and expert perception of permanent unilateral facial paralysis. *JAMA Facial Plast Surg* 2017;19:476–483
6. Lyford-Pike S, Helwig NE, Sohre NE, et al. Predicting perceived disfigurement from facial function in patients with unilateral paralysis. *Plast Reconstr Surg* 2018;142:722–728
7. Li MK, Niles N, Gore S, et al. Social perception of morbidity in facial nerve paralysis. *Head Neck* 2016;38:1158–1163
8. Devcic Z, Karimi K, Popenko N, et al. A web-based method for rating facial attractiveness. *Laryngoscope* 2010;120:902–906
9. van Veen MM, Dijkstra PU, le Coultre S, et al. Gracilis transplantation and temporalis transposition in longstanding facial palsy in adults:

patient-reported and aesthetic outcomes. *Journal of Cranio-Maxillo-Facial Surgery* 2018;46:2144–2149

10. Bates D, Ma?chler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. *J Statist Softw* 2015;67:1–48

11. Ludecke D. Sjstats: Statistical functions for regression models (version 0.17.3). https://CRAN.R-project.org/package=sjstats. Updated 2019. Accessed July 15, 2019.

12. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. *J Statist Softw* 2017;82:1–26

13. Wickham H, Miller E. Haven: Import and export 'SPSS', 'stata' and 'SAS' files. R package version 2.0.0. https://CRAN.R-project.org/package=haven. Updated 2018. Accessed July 15, 2019.

14. Wickham H, Francois R, Henry L, Muller K. Dplyr: A grammar of data manipulation. R package version 0.8.0.1. https://CRAN.R-project.org/package=dplyr. Updated 2019. Accessed July 15, 2019.

15. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag; 2016

16. Terzis JK, Olivares FS. Long-term outcomes of free-muscle transfer for smile restoration in adults. *Plast Reconstr Surg* 2009;123:877–888

17. Darrach H, Ishii LE, Liao D, et al. Assessment of the influence of "other-race effect" on visual attention and perception of attractiveness before and after rhinoplasty. *JAMA Facial Plast Surg* 2019;21:96–102

18. Central Bureau of Statistics. Bevolking naar migratieachtergrond. Updated 2016. https://www.cbs.nl/nl-nl/achtergrond/2016/47/bevolking-naar-migratieachtergrondAccessed August 16, 2019.

19. Watkins CD. Own attractiveness and dissatisfaction with physical appearance independently predict the salience of facial cues to size when women judge other women's attractiveness. *Perception* 2017;46:1321–1328