Quantitative Analysis of Aesthetic Results: Introducing a New Paradigm

Al Aly, MD, FACS; Andre Tolazzi, MD; Shehab Soliman, MD; and Albert Cram, MD

When perusing a plastic surgery journal or attending a plastic surgery meeting, it is evident that the results shown in any given aesthetic presentation are considered by some to be excellent, whereas others deem the same results to be average or less than optimal. This disparity occurs when the interpretation of posttreatment results is based solely on subjective opinion. Certainly, the task of quantifying the results of aesthetic surgery (rather than justsubjectively assessing their quality) is immense, but it is essential for aesthetic surgery to follow the trend toward evidenced-based medicine (EBM) that is becoming ingrained in the fabric of the medical profession as a whole. In fact, the quantification of aesthetic surgery results has more far-reaching ramifications than simply determining objective measures by which results can be judged. Objectively assessing the results of our cosmetic surgeries has the potential to change the way surgery is performed.

As we all learn more about the philosophies behind EBM (eg, in the Editorial by Dr. Felmont Eaves and Dr. Andrea Pusic in this month’s issue, on page 117), it is helpful to also find support among colleagues who have begun implementing it in their own practices. To that end, we would like to share with you the ways in which adding quantitative outcomes assessment, which is the cornerstone of EBM, has changed some of our own clinical approaches. Rigorous research has been conducted and published on how to quantify (instead of merely qualify) patient satisfaction outcomes. However, as Millard taught us, patient satisfaction or dissatisfaction with surgical results should never dissuade us from critically evaluating the results themselves objectively. Thus, it is necessary for us, as plastic surgeons, to adopt a two-pronged approach to the critical evaluation of our surgical results: we must understand our patients’ satisfaction/dissatisfaction with those results and conduct objective evaluations of them. Our ultimate goal in this editorial is to introduce a practical framework for incorporating quantitative analysis measurements into the clinical practice of aesthetic surgery.

To begin, it is important that we recognize that what is considered “aesthetic” is based, at least partially, on previous life experiences. A classic example of this from popular media is the nevus on supermodel Cindy Crawford’s left cheek. It is an abnormality, but because a previous cultural beauty icon, Marilyn Monroe, had a similar nevus, Crawford’s nevus is considered attractive. There is no method by which we can quantitate this aspect of aesthetics because it varies tremendously between individuals and can sometimes even lead to certain individuals finding considerably unattractive appearances pleasing. Thus, our discussion about quantifiable results will be limited to the “non-environmentally influenced” aspects of aesthetics.

THE QUANTIFICATION PROCESS

There are a few essential steps involved in objectively determining the quality of surgical results. To assess the results, we must be able to:

1. measure, calibrate, and average specific parameters or data points (usually through photographs);
2. define an “ideal”;
3. measure and compare the same parameters preoperatively and postoperatively; and
4. compare the postoperative results to the previously-established “ideal.”

Dr. Aly is Professor and Director of Aesthetics at the University of California, Irvine. Dr. Tolazzi is a Voluntary Plastic Surgeon and a postgraduate surgical student at the Hospital de Clinicas of Federal University of Parana (UPPR), Parana, Brazil. Dr. Soliman is Assistant Professor of General and Plastic Surgery at the Kasr El Aini Hospitals, Cairo University, Egypt. Dr. Cram is Professor Emeritus of Plastic Surgery at the University of Iowa, Coralville, Iowa.

Corresponding Author:
Dr. Al Aly, 200 S Manchester Ave, #650, Orange, CA 92868 USA.
E-mail: mdplastic@aol.com
Measure, Calibrate, and Average

Quantitative measurements require accuracy and consistency. In the arena of aesthetic surgery, we currently judge our results through two-dimensional (2D) photographs. As three-dimensional and video photography technology continues to develop, these options will likely become the standard, and the same algorithm we outline here would still apply. However, our current discussion will focus on the existing standard of 2D photography, which can provide distance and surface area measurements.

Again, as a prerequisite to the process of quantifying results, one must be able to standardize photography with regard to position, camera settings, and focal distances. The photographs must also be calibrated—meaning that if the distance between two anatomic reference points is $x$, it must remain $x$ in any other subsequent photograph of the same patient. We utilize the Canfield Imaging Systems software (Fairfield, New Jersey) to accomplish these goals, but a variety of other software is available. Our specialty is in need of commercial software that will allow investigators—and eventually the entire plastic surgery community—to quickly and efficiently average the appearance of many individuals into composite images. The concept of averaging is important in establishing an “ideal.”

Determining an “Ideal”

In order to determine the quality of a postsurgical result, it is necessary to have a “yardstick” against which the results can be measured. In his landmark book, Principilization of Plastic Surgery, Millard\(^5\) referred to this “yardstick”—the ideal—as the “beautiful normal.” This does not necessarily imply extreme levels of beauty but more simply describes “that which is considered normal/attractive.” The process by which the ideal is determined involves “averaging” different characteristics of a population. For example, plastic surgeons know that the appropriate nasolabial angle in a Caucasian woman of normal height should be between 100 and 110 degrees. This number was determined by averaging the nasolabial angles of individuals in a large sample population. Langlois, a researcher in the field of beauty, explains that humans are “cognitive averagers,” or “koinophiles.”\(^6\,7\) (In Greek, koinos means “usual” and philos means “love.”) In essence, we determine attractiveness based on a mental average of characteristics in the population.\(^6\,8\,9\) Cognitive averaging specifically refers to the ongoing, evolving nature of the process, evidenced by the fact that opinions of what is attractive change over time.\(^10\) There is evidence that this tendency is hardwired in humans and is based on evolutionary advantage.\(^11\,12\,13\,14\)

The concept of “averageness” is not new. In the late 1870s, Francis Galton\(^15\) devised an experiment wherein he attempted to identify archetypal faces for certain subsets of the population—specifically, the “criminal archetype.” When he averaged the faces of a group of “criminals,” he was surprised to find that the resultant composite face was similar to composites from other “noncriminal” groups. As a matter of fact, he found that the averaged faces of all groups, including the “criminal” group, were better looking than the individual faces of which the composites were comprised. Johnston, a more recent beauty researcher in this century, also averaged two faces to create a composite.\(^5\,16\,17\) The two original faces, as well as the composite, were then rated by blinded observers for attractiveness. Johnston found that the composite face was rated higher than the originals. He then added another face and repeated the process continually until he had reached a composite of 16 faces. In every case, the composite photograph was rated higher than the pictures utilized to create the composite. In a similar study published in March 2010 in Aesthetic Surgery Journal, Sharabi et al\(^9\) found “statistically significant evidence that averaging more attractive faces, specifically the top 10% of a population, renders a more attractive composite face.”

Historically, the process of averaging faces and creating composites has been difficult and time-consuming. It involves a significant amount of tedious work, which few investigators have undertaken. For example, as described previously, the average/ideal Caucasian female nose was determined after over a century of research in which a large number of investigators (and patients) were involved. However, today’s computer technology has given us the ability to average and quantify the anatomy of the face and body in a timely fashion and with relative ease. Although software that would allow us to render composites is not currently commercially available, it is only a matter of time and demand—which I hope we, as plastic surgeons, will generate.

Comparing Preoperative and Postoperative Photographs

The critical step in evaluating the effects of a surgical procedure is to quantify the difference between the preoperative and postoperative states. It is true that this requires a significant amount of work, but again, future advances in software technology will undoubtedly streamline these processes. Many distances, angles, surface area measurements, and anatomic relationships can be quantified in any region. Surgeons themselves must select which measurements are relevant to each operation. Establishing these parameters, recording the measurements, and comparing those measurements to the same areas postoperatively are the mechanisms by which we can assess the extent of change in a quantitative (rather than qualitative) manner. For example, instead of noting that the nasolabial angle was increased in a postoperative rhinoplasty, quantitative measurements allow us to say, “The change was from $x$ degrees preoperatively to $y$ degrees postoperatively.” This type of information is very powerful not only on an individual patient basis but also as a long-term evaluation tool. If collated from a multitude of rhinoplasty patients, these measurements could determine the effectiveness of a technique in manipulating the particular parameter being studied. For example, if the surgeon...
increased the nasolabial angle by resecting the caudal septum in one group of patients, whereas in another group of patients a cephalic strip resection was utilized, he or she could quantitate the difference between the two techniques and apply this information when selecting which technique is best suited to the needs of a particular patient.

Comparing Surgical Results to the “Ideal”

Once we have decided on the ideal anatomy, we can judge the surgical result not only against the patient’s individual preoperative anatomy but also against the ideal. This would seem to be fairly simple, but a few issues complicate this process. First, the ideal is generally not exact; rather, it is a range within which the result could fall. For example, the previously-mentioned ideal for the nasolabial angle in a Caucasian woman includes a range of acceptable angles, rather than an exact measurement. Second, in most anatomic areas of the body, the surgical result will approximate the ideal in some aspects but will fall short in others. For example, in the case of a postoperative rhinoplasty result, it is possible that the dorsal height and nasolabial angle would be within the normal ranges of the ideal, whereas other aspects, such as tip definition or middle vault width, could fall outside the acceptable range. The third and most significant obstacle to the entire quantification process is the problem of choosing the appropriate ideal. We can facetiously conceptualize this as the “Best in Show” dilemma. This term comes from dog shows, where in the initial stages of the competition, dogs are judged against each other to determine the “best” dog of each breed. In the “Best in Show” segment of the competition, the winning dogs of each breed are judged against each other. How can one judge decide which is the best dog when each dog is so very different? This dilemma is very much like the one we face as plastic surgeons in deciding which result is better since we compare patients from widely variable population groups (Caucasians, Asians, women, men, tall, short, etc). To truly and accurately assess the quality of surgical results, we must determine an ideal for each population and compare individual surgical results only to the ideal for the population category into which the patient falls. This may seem like a daunting task, and historically, similar studies have taken decades to accomplish. However, armed with current and forthcoming software technology, we can photograph patients from different populations that we regularly see and eventually assimilate and render averaged ideals for our patients. For example, if a plastic surgeon treats a large Asian population, patients who are deemed to have aesthetically-pleasing noses can be recruited and photographed. “Averaged composites” can then act as ideals for Asian patients undergoing rhinoplasty, and individual results can be compared accordingly.

An alternative method of delineating multiple ideals or averaged composites is for our plastic surgery societies to collate this information through a project in which the entire plastic surgery community can participate. The ideal composites and measurements could then be distributed to all members of the society who want to utilize them. Whichever method is utilized, it is essential that any ideals/composites undergo continuous updating since cognitive averaging is an ongoing process that is adaptive to time and fashion.
To help crystallize how the process of quantification works and how it can affect our practices, we would like to share a clinical example from our practice. We evaluated a previously-published brachioplasty technique\textsuperscript{18,19} for (1) the degree of arm reduction achieved and (2) the comparison of our results with an ideal.

We began with 20 patients who underwent brachioplasty. They were photographed in a consistent manner, both preoperatively and postoperatively. We then recruited 10 patients who presented to our office seeking breast augmentation but who were also deemed to have “normal/attractive” arms. We photographed their arms in the same manner as the patients who were undergoing brachioplasty. The breast augmentation group comprised our “ideal” population. At the time of the study, we did not have the technology to generate a true composite of the 10 ideal patients, but we measured distance and surface area on their 2D photographs and averaged these values to establish an ideal range of measurement.

Figure 1A demonstrates the distances and surface area measurements we recorded with the Canfield Imaging System, which provided a great deal of accuracy. Figure 1B shows the same measurements on the postoperative photograph. Both sets of photographs were calibrated to ensure consistency and accuracy from picture to picture. In comparing preoperative and postoperative photographs, we found that most of the results had moved closer toward ideal numbers. For example, in Figure 1, the patient experienced a 27% reduction of the surface area (purple area), a 33% reduction of the greatest width of the upper arms, and a 5-cm increase in the distance from the midline to the greatest width (white arrows). However, when we measured the width-to-length ratio of each arm, we found that the postoperative ratio was smaller than the “ideal” group’s ratio, meaning that the postoperative arms were narrower than the ideal (Figure 2).

In this study, we clearly utilized quantitative data to better understand our results, which led to a change in our surgical technique. We obviously made our patients’ arms smaller, but we were able to quantify the reduction, rather than subjectively state that the arms were simply smaller. We also changed some of the relationships of our patients’ arm anatomy (ie, the location of the greatest width relative to the entire upper arm length), which we had simply never noticed prior to embarking on the process of quantification. Delineating an “ideal” to which we could compare our postoperative results, although not in the exact form of the composites suggested in this article, allowed us to determine which of the postoperative changes had brought our patients closer to the ideal and which results we could improve. As a result of this information, we have since altered our technique to leave more tissue behind, in order to align our results even more closely with the ideal. Thus, in our practice, quantification of the aesthetic results has (1) led to a better understanding of anatomy, (2) allowed objective assessment of surgical changes, and (3) inspired changes in surgical technique to better approximate ideal anatomy.

Figure 2. In our previously-published patient series, quantitative data analysis showed that the postoperative width-to-length ratio of each arm was smaller than the “ideal” group’s ratio, meaning that the postoperative arms were narrower than the ideal.
CONCLUSIONS

In order to truly incorporate EBM into the fabric of aesthetic surgery, it is essential that we begin to quantitatively evaluate our surgical results. Armed with the ability to measure and quantitate physical parameters on photographs, we can compare each patient’s preoperative and postoperative states and subsequently determine whether those changes have moved the patient closer toward the “ideal.” It is our hope that, with this editorial, we have demonstrated to the reader how this traditionally difficult task might become a part of clinical practice with the aid of modern photography and software technology and inspired readers to embrace, participate in, and refine the process of quantitative outcomes evaluation.

Disclosures

The author declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Funding

The author received no financial support for the research, authorship, and publication of this article.

REFERENCES