OPEN

Levels of Evidence in Cosmetic Surgery: Analysis and Recommendations Using a New CLEAR Classification

Eric Swanson, MD

Background: The Level of Evidence rating was introduced in 2011 to grade the quality of publications. This system evaluates study design but does not assess several other quality indicators. This study introduces a new "Cosmetic Level of Evidence And Recommendation" (CLEAR) classification that includes additional methodological criteria and compares this new classification with the existing system.

Methods: All rated publications in the Cosmetic Section of *Plastic and Reconstructive Surgery*, July 2011 through June 2013, were evaluated. The published Level of Evidence rating (1–5) and criteria relevant to study design and methodology for each study were tabulated. A new CLEAR rating was assigned to each article, including a recommendation grade (A–D). The published Level of Evidence rating (1–5) was compared with the recommendation grade determined using the CLEAR classification.

Results: Among the 87 cosmetic articles, 48 studies (55%) were designated as level 4. Three articles were assigned a level 1, but they contained deficiencies sufficient to undermine the conclusions. The correlation between the published Level of Evidence classification (1–5) and CLEAR Grade (A–D) was weak ($\rho = 0.11$, not significant). Only 41 studies (48%) evaluated consecutive patients or consecutive patients meeting inclusion criteria.

Conclusions: The CLEAR classification considers methodological factors in evaluating study reliability. A prospective study among consecutive patients meeting eligibility criteria, with a reported inclusion rate, the use of contemporaneous controls when indicated, and consideration of confounders is a realistic goal. Such measures are likely to improve study quality. *(Plast Reconstr Surg Glob Open 2013;1:e66; doi: 10.1097/GOX.0000000000000001; Published online 1 November 2013.)*

The Level of Evidence pyramid was introduced to *Plastic and Reconstructive Surgery* in July 2011 as a means of grading articles and encouraging a higher level of evidence in plastic surgery publications.[1] On the occasion of its second anniversary, it is timely to assess its validity.

The lack of science in plastic surgery is well recognized.[1–5] Efforts to incorporate evidence-based medicine[6,7] in plastic surgery are justified. Both the Level of Evidence[8] and Grade[2] concepts originated in a seminal Canadian Task Force Report published

From the Swanson Center, Leawood, Kans.

Received for publication July 15, 2013; accepted September 6, 2013.

DOI: 10.1097/GOX.0000000000000001

Disclosure: *The author has no financial interest to declare in relation to the content of this article. The Article Processing Charge was paid for by the author.*

Supplemental digital content is available for this article. Clickable URL citations appear in the text.

in 1979.[9] Evidence-based medicine challenges traditional clinical practice based on unsystematic clinical observations, basic principles, common sense, experience, and expert opinions.[7,10–12] Ironically, the Level of Evidence classification[8] itself is a product of experience and expert opinion. Evidence-based medicine is not intended to be static but rather a dynamic, lifelong process[12,13] that recognizes the need to evolve.[10] There is no grandfather clause that shields it from scientific scrutiny.[14] When analyzed, medical practice guidelines often fall short in meeting methodological standards.[14] About half the guidelines are outdated in 6 years.[15] This study endeavors to use the components of evidence-based medicine,[4,12] including "tracking down the best evidence" and "critically appraising that evidence," to investigate evidence-based medicine. Such a study has not been reported in the plastic surgery literature.

METHODS

A 2-year period of publications in *Plastic and Reconstructive Surgery*, July 2011 through June 2013, was retrospectively evaluated. All articles with a Level of Evidence rating published in the Cosmetic Section were included. Each article was designated a quality rating by the author using a new Cosmetic Level of Evidence And Recommendation (CLEAR) scale (Table 1). This classification modifies the traditional Level of Evidence ranking[1] and grade of recommendation (Table 2).[2–5] Table 3 and Figure 1 compare the classifications. Table 4 provides the study design and methodology characteristics for the first 10 articles. To conserve article space, the complete data for all 87 articles are given in **Supplemental Table** (**Supplemental Digital Content 1**, which shows the publications in cosmetic surgery, 2011–2013, and quality of evidence criteria, ***http://links.lww.com/PRSGO/A12***). Table 5 summarizes the findings. Correlation between the assigned Level of Evidence and CLEAR Grade was tested using a Spearman rank correlation coefficient.

RESULTS

Forty-eight studies (55%) were designated as level 4 by *Plastic and Reconstructive Surgery* using its Level of Evidence rating. Three articles were assigned a level 1. Forty-one articles (48%) evaluated consecutive patients or consecutive patients subject to inclusion criteria. Thirty-five studies (40%) consisted of chart reviews and a recording of complication and reoperation rates. Twenty-five studies (29%) reported physical measurements on patients or images. An equal number of studies (29%) featured subjective evaluations of the result by the investigators. Patient-derived data were collected in 18 studies (21%). The correlation between the published Level of Evidence classification (1–5) and CLEAR Grade (A–D) was weak ($\rho = 0.11$, not significant).

DISCUSSION

Levels of Evidence Hierarchy

A level 1 study is often considered the "gold standard" of evidence.[10,11,16,17] A grade A recommendation is usually assigned to such studies.[4,13] A level 5 study, on the other hand, constitutes expert opinion that is often open to question. A level 2 study is a prospective comparison of treatment cohorts, a level 3 study is a retrospective case-control study, and a level 4 study is a case series.[4] Level of Evidence categories are qualitative and nonlinear. The numbers of studies designated to each group are not normally distributed (Fig. 1).[18–21] Consequently, the numerical scores cannot be compared using common statistical techniques that assume normality.[18–20]

Grade (A–D) Recommendation

The present grade classification used by *Plastic and Reconstructive Surgery*[4] provides recommendations based on current knowledge irrespective of the study. A deficient study could receive an "A" grade if existing high-level studies support its con-

Table 1. Cosmetic Level of Evidence and Recommendation

Level	Description	Recommendation
1	Randomized trial with a power analysis supporting sample sizes	A
2	Prospective study, high inclusion rate (≥ 80%), and description of eligibility criteria Objective measuring device (ie, not surgeon's opinion) or patient-derived outcome data Power analysis if treatment effect is compared No control or comparative cohort is needed if effect is profound	A
3	Retrospective case-control study using a contemporaneous control group Prospective clinical study with an inclusion rate <80% Prospective study without controls or comparison group and a treatment effect that is not dramatic	B
4	Retrospective case series of consecutive patients Case/control study using historical controls or controls from other publications Important confounder that might explain treatment effect	C
5	Case report, expert opinion, and nonconsecutive case series	D

Table 2. Grade of Recommendation

A: Conclusion strongly supported by the evidence, likely to be conclusive
B: Conclusion strongly supported by the evidence
C: Moderate support based on the evidence
D: Inconclusive based on the evidence presented

Table 3. Comparison of Level of Evidence and CLEAR Criteria

Parameter	PRS* Level of Evidence	CLEAR
Study design		
Randomization	✓	✓
Prospective vs retrospective	✓	✓
Control or comparative cohort	✓	✓
Methodology		
Consecutive patients		✓
Power analysis		✓
Eligibility criteria		✓
Inclusion rate		✓
Important confounder		✓
Dramatic effect		✓

PRS, *Plastic and Reconstructive Surgery*

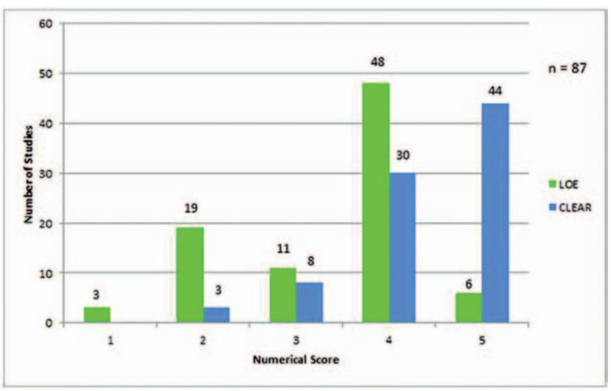

Fig 1. Comparison of the assigned Level of Evidence (LOE) and CLEAR Grade for 87 consecutive studies published in the Cosmetic Section of *Plastic and Reconstructive Surgery* from July 2011 to June 2013. Two studies were unratable by the CLEAR classification because of study error.

clusion. The CLEAR Grade rates the overall quality of the study itself, regardless of conventional wisdom. A low-quality study that concludes, for example, that smoking increases the complication rate may receive a low grade of recommendation, despite support in the literature. Because methodology is considered in the CLEAR numerical rating (1–5), the grade tends to be closely linked. In this study, the CLEAR level and grade always matched (2A, 3B, 4C, and 5D). The traditional Level of Evidence rating does not correlate well with the recommendation grade ($\rho = 0.11$, not significant) because it does not consider several important quality parameters (Table 3).

Level 1 Studies

Only 3 studies were designated level 1. The first study to be assigned a level 1 in the Cosmetic Section of *Plastic and Reconstructive Surgery* was titled: "A multicenter, prospective, randomized, single-blind, controlled clinical trial comparing VASER-assisted lipoplasty and suction-assisted lipoplasty."[22] This corporate-funded study's title promises a high level of quality. Unfortunately, however, this article's methodological deficiencies, including error in calculations,[23] make it unratable.

In their level 1 study, Costa-Ferreira et al[24] did not control for an important confounding variable—electrodissection.[25] Their study points to the value of medical experience and physiological understanding on the part of the reviewer.[7] The third level 1 article concludes that 35% of patients undergoing cosmetic rhinoplasty suffer from body dysmorphic disorder, using an expanded and inaccurate definition of this syndrome[26] in addition to other methodological deficiencies.[27]

Paradoxically, all 3 level 1 studies arrive at unreliable conclusions that encourage the reader to needlessly (1) purchase a 6-figure instrument,[22] (2) compromise the esthetic result of an abdominoplasty,[24] and (3) deny surgery to one third of prospective cosmetic rhinoplasty patients.[26] These 3 level 1 studies represent just 3% of the total, equal to the percentage of level 1 studies published in 3 major plastic surgery journals from 1998 to 2007.[11] Their frequency does not seem to be increasing as hoped.[11,17] It is reasonable to ask whether a randomized trial (the additional descriptors, "controlled" and "prospective" are redundant) is the ideal model.

Randomized Trials and Cosmetic Surgery

Randomized trials balance both known and unknown confounders and avoid selection bias[3,18,28]—at least theoretically.[29] In drug testing, the need to identify a true benefit from a medication, without the influence of other factors, is well-known. However, surgery is a much different discipline.[11,18,30–33]

Unlike a pill, a procedure is not identical from patient to patient,[11,18] placebos and blinding are usually not possible, and randomization is not well accepted by patients,[11,16,28] surgeons,[16,28,33] or referral sources.[31] Patients are particularly averse to randomization when the choice involves an operation with irreversible consequences.[16,17] Solomon and McLeod[34] report that most (60%) surgical questions would not be suitable for randomized trials, citing patient resistance, uncommon conditions, and lack of clinical equipoise as the most common reasons. Other shortcomings include a lack of external validity (generalizability),[3,18,28,35] the fact that

Table 4. Publications and Quality of Evidence Criteria

No.	Authors	Month, Year	Title	Consecutive	Power Analysis	Inclusion Criteria	Inclusion Rate (%)	Prospective/Retrospective	Control Group	Confounders	Sample Sizes	Measuring Device	Methodological Issues	Discussion Limits	Commercial Bias	Level of Evidence	CLEAR Rating*
1	Jewell et al[36]	July, 2011	Randomized sham-controlled trial to evaluate the safety and effectiveness of a high-intensity focused ultrasound device for noninvasive body sculpting	N	N	Y	N	Prospective	Sham group	Different operators and measurers	58/59/63	Waist circumference, patient surveys, investigator reviews of photographs	Precision of measuring tape; commercial bias; nonconsecutive patients	N	Y	2T	5D*
2	Maffi et al[37]	July, 2011	Traditional lower blepharoplasty: is additional support necessary? A 30-year review	Y	N	Y	66.6	Retrospective	N	N	2007	Chart review	Nonstandardized photographs; no measurements; no data on length of follow-up or missing photographs	Y	N	4T	4C*
3	Mojallal[38] et al	July, 2011	Dorsal aesthetic lines in rhinoplasty: a quantitative outcome-based assessment of the component dorsal reduction technique	Y	N	Y	N	Retrospective	N	N	100	Comparison of anatomical points on photographs with computer assistance	Subjectivity in point assignments; lighting differences	Y	N	4T	4C
4	Picavet et al[39]	August, 2011	High prevalence of body dysmorphic disorder symptoms in patients seeking rhinoplasty	Y	N	Y	72.9	Prospective	Y	Controls not cosmetic patients	226/65	Patient surveys, scoring of photographs by 2 investigators	Expanded definition of body dysmorphic disorder	N	N	3R	3B*
5	Coriddi et al[40]	August, 2011	Changes in quality of life and functional status following abdominal contouring in the massive weight loss population	Y	N	Y	94	Retrospective	N	Concomitant procedures, rectus plication, bariatric surgery	48	Phone surveys	Small sample size, number of patients with plication not given, type 2 error possible with regard to diastasis repair, no adjustment of α level for multiple comparisons	Y	N	4T	4C*

6	Sforza et al[41]	August, 2011	Transversus abdominis plane block anesthesia in abdominoplasties	N	N	N	N	Prospective†	Y	Controls had no local anesthesia (block may act as local anesthetic)	14/14	Subjective pain scores, morphine use	Small sample sizes, no power analysis, no inclusion rate, no comparison of block vs local anesthetic infusion to confirm superiority of block	N	N	2T	5D
7	Cárdenas-Camarena et al[42]	August, 2011	Buttocks fat grafting: 14 years of evolution and experience	Y	N	Y	100	Retrospective	N	Different surgeons, different times	492/132/165	Complication rates	Chronology bias, no patient-derived data or measurements of results	N	N	3T	4C
8	Trussler et al[43]	September, 2011	The viscoelastic properties of the SMAS and its clinical translation: firm support for the high SMAS rhytidectomy	Y	N	Y	100	Prospective	Y	Applied tension varies	30	Tensiometer, biomechanical tissue studies	SMAS tension likely to change after suturing; tension likely to be less and duration much longer in vivo	Y	N	4T	2A
9	von Soest et al[44]	September, 2011	Psychosocial changes after cosmetic surgery: a 5-year follow-up study	N	N	N	48.7/41.9	Prospective	Y	Controls did not have any surgery	201/838	Mailed surveys	Nonconsecutive patients, no inclusion rate, low response rate, women only, disproportionately breast surgery, no follow-up of control group	Y (excellent)	N	2T	5D*
10	Rohrich et al[45]	September, 2011	The five-step lower blepharoplasty: blending the eyelid-cheek junction	Y	N	Y	N	Retrospective	N	Effects of simultaneous face-lifts	50	Computer-assisted photograph measurements of ratios	Combining right- and left-sided data; no control or comparison group	Y	N	4T	4C

Complete data are provided in **Supplemental Table** (**Supplemental Digital Content 1**, which shows the publications in cosmetic surgery, 2011–2013, and quality of evidence criteria, ***://links.lww.com/PRSGO/A12***). Articles not indicating whether the patients were consecutive were deemed to include nonconsecutive patients.
*Discussion accompanied article.
†Randomized group included.
N, no; Y, yes

Table 5. Study Characteristics by CLEAR Rating

Study Parameter	2A (%)	3B (%)	4C (%)	5D (%)	All Studies (%)
No. studies	3	8	30	44	85
Design					
Randomized	0 (0)	0 (0)	1 (3.3)	3 (6.8)	4 (4.7)
Prospective	3 (100)	5 (62.5)	2 (6.7)	17 (38.6)	27 (31.8)
Comparative cohort	1 (33.3)	5 (62.5)	5 (16.7)	10 (22.7)	21 (24.7)
Control	1 (33.3)	2 (25.0)	1 (3.3)	9 (20.5)	13 (15.3)
Methodology					
Consecutive patients	3 (100)	8 (100)	30 (100)	0 (0)	41 (48.2)
Power analysis	1 (33.3)	1 (12.5)	0 (0)	1 (2.3)	3 (3.5)
Description of inclusion criteria	3 (100)	8 (100)	29 (96.7)	19 (43.2)	59 (69.4)
Inclusion rate provided	3 (100)	7 (87.5)	21 (70.0)	11 (25.0)	42 (49.4)
Confounders	1 (33.3)	7 (87.5)	24 (80.0)	33 (75.0)	65 (76.5)
Inclusion rate, %					
Mean	89.4	78.9	81.9	54.5	75.1
SD	10.0	14.9	26.4	42.3	30.9
Range	80–100	65.3–100	23.6–100	1.5–100	1.5–100
Sample sizes					
Mean	150.3	612.8	371.1	332.1	361.8
SD	105.6	962.0	761.4	759.8	754.2
Range	30–225	20–2971	9–3636	5–3800	5–3800
Other					
Discussion of limitations	3 (100)	3 (37.5)	16 (53.3)	19 (43.2)	41 (48.2)
Commercial bias	0 (0)	0 (0)	4 (13.3)	8 (18.2)	12 (14.1)
Discussion accompanying article	0 (0)	5 (62.5)	9 (30.0)	8 (18.2)	22 (25.9)

surgeons are rarely equally proficient in and enthusiastic about 2 different techniques[32] and cost.[18,28,35] Funding is an issue for cosmetic surgeons in practice.[16] Such studies need to be cost-effective.[46] Lack of funding can lead to methodological compromises.[47] Randomized trials suffer from low inclusion rates and recruitment biases and may be underpowered.[35] In surgery, by the time a randomized trial is conducted, the novel procedure has often been improved.[31] Techniques evolve quickly, particularly in plastic surgery.[32] Fortunately, well-performed nonrandomized studies can still provide accurate and clinically useful information.[32]

Two rigorous reviews published in the same issue of *The New England Journal of Medicine* in 2000 reveal that randomized trials and observational studies usually produce similar results.[48,49] Observational studies may be more consistent and less prone to reporting contradictory results.[48] Concato et al[48] conclude that research design should not be considered a rigid hierarchy. They attribute the greater homogeneity of observational studies to their broader representation of the general population.

Randomized trials are inflexible and disallow modifications that might better suit individual patients. Inadequate concealment of randomization and treatment assignments can cause serious bias that may exceed the magnitude of the treatment effect.[50–52] Bhandari et al[47] report that two thirds of randomized orthopedic trials did not use proper techniques of randomization or concealment. Reviews of randomized trials in plastic surgery uniformly report low quality.[17,21,53–56]

Equipoise

Ethical considerations prohibit randomization of patients into 2 groups, one of which constitutes a known inferior treatment.[28] Different operations on contralateral sides of the face or body may produce asymmetry. Predictably, such studies tend to find no difference in treatment effects.[57,58] Cognitive dissonance may inhibit a surgeon from finding that one half of his or her randomized patients received an inferior treatment.[29]

This discussion leads to a catch-22. If investigators compare one operation with another, they already believe one treatment is superior or they would not be conducting the study. If the difference is so slight that there is no consistent evidence one way or the other, the study is probably pointless. Fortunately, most clinical questions in plastic surgery do not concern whether a procedure is superior to nothing.[46] Therefore, shams are usually not needed.[46]

Most randomized controlled trials in plastic surgery evaluate nonsurgical interventions.[59–69] Surgical trials may compare adjunctive techniques or products.[36,70–72] Such issues (eg, the use of drains) do not substantially affect the long-term result and are therefore more appropriate for a randomized model.

Limitations of Historical Controls

Studies using historical controls are predisposed to find that the newer therapy is superior to its predecessor.[51,73] Similar to randomized trials, the conclusions are usually more dependent on the method of selection of control groups than on the therapy, and the majority differ from the results of randomized trials of the same therapy.[73] Methodological standards are commonly violated in case-control studies.[74] Chronology bias is difficult to avoid.[75] Matched cohort groups are notoriously difficult in plastic surgery, especially cosmetic surgery.[76] Recent guidelines assign a level 4 to such studies, no better than a case series.[77] Contemporaneous controls are preferred.

If the treatment effect is dramatic (eg, breast self-consciousness after augmentation), a control group is unnecessary (eg, a control group of women not electing to have a breast augmentation). A prospective study with a dramatic effect, but no control group, can qualify as a CLEAR level 2 study if other requirements are met (Table 1).[77]

Prospective vs Retrospective Study Design

The literature consistently supports the superiority of a prospective study. A prospective study is always preferred over a retrospective study if it is feasible.[78] Some investigators may question this distinction because data are always collected prospectively. The difference is the vantage point—literally looking forward vs looking backward. The outcome of a prospective study is unknown when it is undertaken, making the investigator less prejudiced. A review of a "prospective" database does not qualify because the investigator is looking back to interpret data. By definition, in a prospective study, the study is conceived before the data are collected.

Selection bias and confounders are reduced by specifying eligibility criteria, encouraging follow-up appointments, standardizing and calibrating photographs and measurements, and administering contemporaneous surveys (rather than years later). An example would be a study to determine whether patient gender affects seroma rates after body contouring surgery. A prospective study would take care to record patient weights on the same scales, preoperative weight loss, intraoperative use of electrodissection, and tissue resection weights. Some of these important details might be missing in a retrospective study. Prospective studies usually disclose more realistic complication rates than retrospective studies. Unavoidable confounders (eg, a difference in mean body mass indices) may be managed using an analysis of covariance or other statistical adjustment.[79]

Markers of Success in Cosmetic Surgery

Patient satisfaction and improved quality of life,[80,81] assessed using patient-derived outcome measures, are the hallmarks of successful plastic surgery. Morbidity and mortality measures are less relevant to plastic surgery than other surgical disciplines.[28,81] Reoperation rates are unreliable markers of quality in cosmetic surgery.[82]

Consecutive Patients

Over 2 decades ago, Goldwyn[83] cautioned that selectively reporting better results does nothing to advance the specialty. Nevertheless, a requirement for consecutive patients is conspicuously absent from the existing Level of Evidence rating (likely because of its nonsurgical origins). This scale does not penalize the investigator for "cherry picking" patients; nor does it reward the investigator for reporting both good and bad results. Both series receive the same catchall level 4 designation. Insisting on consecutive patients (1) sends a message to investigators to report all results and (2) prevents studies of selected patients that include higher level design characteristics from receiving undeserved higher rankings. Like a framework built on a weak foundation, no other study attribute can compensate for an unrepresentative patient sample.[84]

When discussing consecutive patients, it is important to be precise. A study that reports 1-year postoperative photographic findings in 100 "consecutive patients" would be unlikely because not all patients are likely to return for photographs in 1 year; the authors more likely mean "consecutive patients returning for 1-year follow-up" and the inclusion rate should be provided. Many studies would improve from a CLEAR 5 to a CLEAR 4 ranking, or higher, simply by including consecutive patients (eg, clinical studies) or consecutive patients subject to reasonable inclusion criteria that usually include sufficient time for resolution of swelling (eg, measurement and outcome studies). A nonconsecutive case series is just a plural form of a case report and is therefore no more deserving of a higher rank. It is not difficult to report consecutive patients. Goldwyn[85] observes that "it is amazing how easy it is to be truthful if one wants to be." Correction of this bad habit represents the single most important change to increase the overall level of evidence in plastic surgery publications. Although level 1 studies will continue to be rare, it is realistic to expect a more balanced distribution of articles between levels 2 and 5.

Statistical Power and α Level

Sample size calculation is an important part of any prospective study, whether randomized or not,[28,56,86] but is infrequently performed (3.5% of studies).[56,86] Small sample sizes predispose to type II false-negative statistical errors. Although an α level of 0.05 is the standard (ie, 5% false positives), most investigators prefer an α level of 0.01 or a Bonferroni correction to reduce the risk of type I error when multiple comparisons are made.

Eligibility Criteria

Eligibility criteria are necessary to preserve the integrity of the data, avoid confounders, and respect patient privacy (eg, for face-lift studies, a minimum follow-up time, no makeup, no additional surgery or injections, and patient consent for photographs).[84]

Inclusion Rate

Every effort should be made to avoid losing patients to follow-up (ideally, <20%).[87] If the outcome of nonresponders is missing (eg, dissatisfied patients may seek follow-up elsewhere, or alternatively, satisfied patients may see no reason to return), the reliability of the conclusion is jeopardized.[87]

Confounders

Most of the studies (76.5%) included extraneous factors that might correlate with the study variables. If a confounder was judged important enough to undermine the conclusion, a study was given a CLEAR level of 4, provided it still met the requirement for consecutive patients. Plastic surgeons need to take part in evaluating levels of evidence and not delegate this task.[88] There is no substitute for clinical experience and judgment in assessing a study's validity.[2]

Measuring Device

The missing link in the application of the scientific method to plastic surgery is frequently a reliable measuring device.[89] Most studies feature subjective assessments or arbitrary metrics.[11] Direct measurements on standardized, calibrated photographs are preferred. Photographs should include at least one view accompanied by a ruler or measuring tape for calibration, avoiding the need for less intuitive devices such as ratios or pixel counts (eg, rhinoplasty). Computer-assisted photographic standardization and calibration greatly facilitate such measurements.[89]

Discussion of Limitations

All studies had limitations (**Supplemental Table, Supplemental Digital Content 1**, which shows the publications in cosmetic surgery, 2011–2013, and quality of evidence criteria, ***http://links.lww.com/PRS-GO/A12***). However, over half (52%) did not discuss limitations. Such discussions reflect well on the investigators and improve credibility.

Commercial Bias

Corporate sponsorship affects conclusions.[90] Hall-Findlay[91] expresses a concern familiar to many experienced plastic surgeons: "We listen to the manufacturer's claims and then years later we find that we have been misled—both by the manufacturers themselves and by those surgeons who are burdened by a conflict of interest." The willingness to resist marketing pressures and prioritize science over marketing is a sign of professionalism.[89]

Systematic Reviews

A limitation of systematic reviews is that their validity depends on the quality of the reviewed material.[92] As overall study quality improves, systematic reviews become feasible.

Clinical Relevance to Plastic Surgeon Investigators

How might these principles be put to use? A timely example might be an investigation of the effectiveness of buttock fat injection. An investigator may set out to assess the results using a measuring device (standardized photographic measurements) on consecutive patients meeting appropriate eligibility criteria (a minimum follow-up period to allow resolution of swelling), with a power analysis supporting sample size, and a comparative cohort of patients to serve as controls (eg, breast augmentation patients who agree to have their lower body photographed) with simultaneous measurement of body weights to rule out a possible confounder. These principles are highly practical and are likely to meet with a high level of patient compliance. Such a level 2 study would serve to answer an important clinical question.

Recommendations

The CLEAR classification preserves the common language of the original 5-level scale. A level 1 study remains a randomized controlled trial. The CLEAR system differs in adding important methodological considerations (Table 3). Levels 1 and 2 are considered equals, so that the top 2 levels of the pyramid may be colored just one color. Such modifications are permissible and encouraged to keep up with the latest knowledge.[77] Classifying articles as "therapeutic," "risk," or "diagnostic" does not affect quality assessment. One set of guidelines is simpler than 3.

Limitations of the Study

This review consists only of articles appearing in the Cosmetic Surgery section of *Plastic and Reconstructive Surgery*. All grade assignments were made by the author, although the opinions of the discussants were considered when available. One might consider whether a committee would make a more valid determination. The fact that all articles passed peer review would suggest that a consensus opinion is not always reliable or objective either. It is impossible to fully objectify this process.

Strengths of the Study

This analysis uses the concepts of evidence-based medicine to evaluate its own guidelines as applied to cosmetic surgery. Using the same reviewer (E.S.) for each study eliminates interobserver variation. Recommendations are made based on analysis of the data, a review of the literature, and the particular needs of this subspecialty.

CONCLUSIONS

The vestiges of an artistic perspective are evident in plastic surgery publications. Plastic surgeons need to recommit to scientific scrutiny of their results.[89] Practical improvements in study design and methodology are possible. A randomized controlled trial is unlikely to be feasible or even desirable. A prospective study among consecutive patients meeting eligibility criteria, with a reported inclusion rate, and the use of contemporaneous controls when indicated, is a realistic goal. Objective measurements and consideration of patient-derived data are most useful. With attention to such basic steps, an improvement in study quality is inevitable.

Supplemental Table. See Supplemental Table, Supplemental Digital Content 1, which shows publications in cosmetic surgery, 2011–2013, and quality of evidence criteria, ***http://links.lww.com/PRSGO/A12***.

Eric Swanson, MD
Swanson Center
11413 Ash Street
Leawood
KS 66211
E-mail: eswanson@swansoncenter.com

ACKNOWLEDGMENT

The author thanks Jane Zagorski, PhD, for statistical analyses.

REFERENCES

1. Sullivan D, Chung KC, Eaves FF III, et al. The level of evidence pyramid: indicating levels of evidence in Plastic and Reconstructive Surgery articles. *Plast Reconstr Surg.* 2011;128:311–314.
2. Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490–1494.
3. Chung KC, Swanson JA, Schmitz D, et al. Introducing evidence-based medicine to plastic and reconstructive surgery. *Plast Reconstr Surg.* 2009;123:1385–1389.
4. Swanson JA, Schmitz D, Chung KC. How to practice evidence-based medicine. *Plast Reconstr Surg.* 2010;126:286–294.
5. Matarasso A. Discussion: prospective study of lidocaine, bupivacaine, and epinephrine levels and blood loss in patients undergoing liposuction and abdominoplasty. *Plast Reconstr Surg.* 2012;130:723–725.
6. Guyatt GH. Evidence-based medicine. *Ann Intern Med.* 1991;114(ACP *J Club.* Suppl 2): A-16.
7. Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 1992;268:2420–2425.
8. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1986;89(Suppl.)2S–3S.
9. Canadian Task Force on the Periodic Health Examination: the periodic health examination. *Can Med Assoc J.* 1979;121:1193–1254.
10. Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71–72.
11. Chang EY, Pannucci CJ, Wilkins EG. Quality of clinical studies in aesthetic surgery journals: a 10-year review. *Aesthet Surg J.* 2009;29:144–147; discussion 147–149.
12. Sackett DL, Rosenberg WM. The need for evidence-based medicine. *J R Soc Med.* 1995;88:620–624.
13. Rohrich RJ, Eaves FF III. So you want to be an evidence-based plastic surgeon? A lifelong journey. *Plast Reconstr Surg.* 2011;127:467–472.
14. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999;281:1900–1905.
15. Shekelle PG, Ortiz E, Rhodes S, et al. Validity of the Agency for Healthcare Research and Quality clinical practice guidelines: how quickly do guidelines become outdated? *JAMA* 2001;286:1461–1467.
16. Offer GJ, Perks AG. In search of evidence-based plastic surgery: the problems faced by the specialty. *Br J Plast Surg.* 2000;53:427–433.
17. Momeni A, Becker A, Antes G, et al. Evidence-based plastic surgery: controlled trials in three plastic surgical journals (1990 to 2005). *Ann Plast Surg.* 2009;62:293–296.
18. Loiselle F, Mahabir RC, Harrop AR. Levels of evidence in plastic surgery research over 20 years. *Plast Reconstr Surg.* 2008;121:207e–211e.
19. Hanzlik S, Mahabir RC, Baynosa RC, et al. Levels of evidence in research published in *The Journal of Bone and Joint Surgery (American Volume)* over the last thirty years. *J Bone Joint Surg Am.* 2009;91:425–428.
20. Sinno H, Neel OF, Lutfy J, et al. Level of evidence in plastic surgery research. *Plast Reconstr Surg.* 2011;127:974–980.
21. Chuback JE, Yarascavitch BA, Eaves F III, et al. Evidence in the aesthetic surgical literature over the past decade: how far have we come? *Plast Reconstr Surg.* 2012;129:126e–134e.
22. Nagy MW, Vanek PF Jr. A multicenter, prospective, randomized, single-blind, controlled clinical trial comparing VASER-assisted lipoplasty and suction-assisted lipoplasty. *Plast Reconstr Surg.* 2012;129:681e–689e.
23. Swanson E. Improved skin contraction after VASER-assisted lipoplasty: is it a change we can believe in? *Plast Reconstr Surg.* 2012;130:754e–756e.
24. Costa-Ferreira A, Rebelo M, Silva A, et al. Scarpa fascia preservation during abdominoplasty: randomized clinical study of efficacy and safety. *Plast Reconstr Surg.* 2013;131:644–651.

25. Swanson E. Scarpa fascia preservation during abdominoplasty: randomized clinical study of efficacy and safety. *Plast Reconstr Surg.* 2013;132:871e–873e.
26. Picavet VA, Gabriëls L, Grietens J, et al. Preoperative symptoms of body dysmorphic disorder determine postoperative satisfaction and quality of life in aesthetic rhinoplasty. *Plast Reconstr Surg.* 2013;131:861–868.
27. Swanson E. Preoperative symptoms of body dysmorphic disorder determine postoperative satisfaction and quality of life in aesthetic rhinoplasty. *Plast Reconstr Surg.* 2013; In Press.
28. McCarthy CM, Collins ED, Pusic AL. Where do we find the best evidence? *Plast Reconstr Surg.* 2008;122:1942–1947; discussion 1948.
29. Swanson E. Randomized controlled trial comparing health-related quality of life in patients undergoing vertical scar versus inverted T-shaped reduction mammaplasty. *Plast Reconstr Surg.* 2013; In Press.
30. Love JW. Drugs and operations. Some important differences. *JAMA* 1975;232:37–38.
31. Bonchek LI. Sounding board. Are randomized trials appropriate for evaluating new operations? *N Engl J Med.* 1979;301:44–45.
32. Khan AA, Murthy AS, Ali N. Randomized controlled trials in plastic surgery. *Plast Reconstr Surg.* 2006;117:2080–2081.
33. Gupta DM, Panetta NJ, Longaker MT. Quality of clinical studies in aesthetic surgery journals: a 10-year review (commentary). *Aesthet Surg J.* 2009;29:147–149.
34. Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? *Surgery* 1995;118:459–467.
35. Chung KC, Ram AN. Evidence-based medicine: the fourth revolution in American medicine? *Plast Reconstr Surg.* 2009;123:389–398.
36. Jewell ML, Baxter RA, Cox SE, et al. Randomized sham-controlled trial to evaluate the safety and effectiveness of a high-intensity focused ultrasound device for noninvasive body sculpting. *Plast Reconstr Surg.* 2011;128:253–262; discussion 263–264.
37. Maffi TR, Chang S, Friedland JA. Traditional lower blepharoplasty: Is additional support necessary? A 30-year review. *Plast Reconstr Surg.* 2011;128:265–273: discussion 274–279.
38. Mojallal A, Ouyang D, Saint-Cyr M, Bui N, Brown SA, Rohrich RJ. Dorsal aesthetic lines in rhinoplasty: A quantitative outcome-based assessment of the component dorsal reduction technique. *Plast Reconstr Surg.* 2011;128:280–288.
39. Picavet VA, Prokopakis EP, Gabriëls L, Jorissen M, Hellings PW. High prevalence of body dysmorphic disorder symptoms in patients seeking rhinoplasty. *Plast Reconstr Surg.* 2011;128:509–517; discussion 518–519.
40. Coriddi MR, Koltz PF, Chen R, Gusenoff JA. Changes in quality of life and functional status following abdominal contouring in the massive weight loss population. *Plast Reconstr Surg.* 2011;128:520–526; discussion 527–528.
41. Sforza M, Andjelkov K, Zaccheddu R, Nagi H, Colic M. Transversus abdominis plane block anesthesia in abdominoplasties. *Plast Reconstr Surg.* 2011;128:529–535.
42. Cárdenas-Camarena L, Arenas-Quintana R, Robles-Cervantes J-A. Buttocks fat grafting: 14 years of evolution and experience. *Plast Reconstr Surg.* 2011;128:545–555.
43. Trussler AP, Hatef D, Broussard GB, Brown S, Barton FE. The viscoelastic properties of the SMAS and its clinical translation: Firm support for the high-SMAS rhytidectomy. *Plast Reconstr Surg.* 2011;128:757–764.
44. von Soest T, Kvalem IL, Skolleborg KC, Roald HE. Psychosocial changes after cosmetic surgery: A 5-year follow-up study. *Plast Reconstr Surg.* 2011;128:765–772; discussion 773–774.
45. Rohrich RJ, Ghavami A, Mojallal A. The five-step lower blepharoplasty: Blending the eyelid-cheek junction. *Plast Reconstr Surg.* 2011;128:775–783.
46. McCarthy JE, Chatterjee A, McKelvey TG, et al. A detailed analysis of level I evidence (randomized controlled trials and meta-analyses) in five plastic surgery journals to date: 1978 to 2009. *Plast Reconstr Surg.* 2010;126:1774–1778.
47. Bhandari M, Richards RR, Sprague S, et al. The quality of reporting of randomized trials in the *Journal of Bone and Joint Surgery* from 1988 through 2000. *J Bone Joint Surg Am.* 2002;84-A:388–396.
48. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med.* 2000;342:1887–1892.
49. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med.* 2000;342:1878–1886.
50. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–412.
51. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185–1190.
52. Chalmers TC, Celano P, Sacks HS, et al. Bias in treatment assignment in controlled clinical trials. *N Engl J Med.* 1983;309:1358–1361.
53. Taghinia AH, Liao EC, May JW Jr. Randomized controlled trials in plastic surgery: a 20-year review of reporting standards, methodologic quality, and impact. *Plast Reconstr Surg.* 2008;122:1253–1263.
54. Veiga Filho J, Castro AA, Veiga DF, et al. Quality of reports of randomized clinical trials in plastic surgery. *Plast Reconstr Surg.* 2005;115:320–323.
55. Karri V. Randomised clinical trials in plastic surgery: survey of output and quality of reporting. *J Plast Reconstr Aesthet Surg.* 2006;59:787–796.
56. Ayeni O, Dickson L, Ignacy TA, et al. A systematic review of power and sample size reporting in randomized controlled trials within plastic surgery. *Plast Reconstr Surg.* 2012;130:78e–86e.
57. Rees TD, Aston SJ. A clinical evaluation of the results of submusculo-aponeurotic dissection and fixation in face lifts. *Plast Reconstr Surg.* 1977;60:851–859.
58. Ivy EJ, Lorenc ZP, Aston SJ. Is there a difference? A prospective study comparing lateral and standard SMAS face lifts with extended SMAS and composite rhytidectomies. *Plast Reconstr Surg.* 1996;98:1135–1143; discussion 1144.
59. Collis N, Elliot LA, Sharpe C, et al. Cellulite treatment: a myth or reality: a prospective randomized, controlled trial of two therapies, endermologie and aminophylline cream. *Plast Reconstr Surg.* 1999;104:1110–1114; discussion 1115.
60. Oliver DW, Hamilton SA, Figle AA, et al. A prospective, randomized, double-blind trial of the use of fibrin sealant for face lifts. *Plast Reconstr Surg.* 2001;108:2101–2105, discussion 2106.
61. Carruthers JD, Lowe NJ, Menter MA, et al; Botox Glabellar Lines II Study Group. Double-blind, placebo-controlled study of the safety and efficacy of botulinum toxin type A for patients with glabellar lines. *Plast Reconstr Surg.* 2003;112:1089–1098.

62. Cohen SR, Holmes RE. Artecoll: a long-lasting injectable wrinkle filler material: report of a controlled, randomized, multicenter clinical trial of 251 subjects. *Plast Reconstr Surg.* 2004;114:964–976; discussion 977.
63. Lindqvist C, Tveten S, Bondevik BE, et al. A randomized, evaluator-blind, multicenter comparison of the efficacy and tolerability of Perlane versus Zyplast in the correction of nasolabial folds. *Plast Reconstr Surg.* 2005;115: 282–289.
64. Jones BM, Grover R, Hamilton S. The efficacy of surgical drainage in cervicofacial rhytidectomy: a prospective, randomized, controlled trial. *Plast Reconstr Surg.* 2007;120:263–270.
65. Totonchi A, Guyuron B. A randomized, controlled comparison between arnica and steroids in the management of postrhinoplasty ecchymosis and edema. *Plast Reconstr Surg.* 2007;120:271–274.
66. Kazmier FR, Henry SL, Christiansen D, et al. A prospective, randomized, double-blind, controlled trial of continuous local anesthetic infusion in cosmetic breast augmentation. *Plast Reconstr Surg.* 2008;121:711–715.
67. McCarthy CM, Pusic AL, Hidalgo DA. Efficacy of pocket irrigation with bupivacaine and ketorolac in breast augmentation: a randomized controlled trial. *Ann Plast Surg.* 2009;62:15–17.
68. Larson JD, Gutowski KA, Marcus BC, et al. The effect of electroacustimulation on postoperative nausea, vomiting, and pain in outpatient plastic surgery patients: a prospective, randomized, blinded, clinical trial. *Plast Reconstr Surg.* 2010;125:989–994.
69. Pannucci CJ, Reavey PL, Kaweski S, et al. A randomized controlled trial of skin care protocols for facial resurfacing: lessons learned from the Plastic Surgery Educational Foundation's Skin Products Assessment Research study. *Plast Reconstr Surg.* 2011;127:1334–1342.
70. Collis N, Coleman D, Foo IT, et al. Ten-year review of a prospective randomized controlled trial of textured versus smooth subglandular silicone gel breast implants. *Plast Reconstr Surg.* 2000;106:786–791.
71. Prado A, Andrades P, Danilla S, et al. A prospective, randomized, double-blind, controlled clinical trial comparing laser-assisted lipoplasty with suction-assisted lipoplasty. *Plast Reconstr Surg.* 2006;118:1032–1045.
72. Andrades P, Prado A, Danilla S, et al. Progressive tension sutures in the prevention of postabdominoplasty seroma: a prospective, randomized, double-blind clinical trial. *Plast Reconstr Surg.* 2007;120:935–946; discussion 947.
73. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med.* 1982;72:233–240.
74. Horwitz RI, Feinstein AR. Methodologic standards and contradictory results in case-control research. *Am J Med.* 1979;66:556–564.
75. Swanson E. Chemoprophylaxis for venous thromboembolism prevention: concerns regarding efficacy and ethics. *Plast Reconstr Surg—Global Open* 2013;1:e23. doi: 10.1097/ GOX.0b013e318299fa26.
76. Rohrich RJ. So you want to be better: the role of evidence-based medicine in plastic surgery. *Plast Reconstr Surg.* 2010;126:1395–1398.
77. Oxford centre for evidence-based medicine, 2011 levels of evidence. Available at: http://www.cebm.net/index.aspx?o=5653. Accessed July 10, 2013.
78. Hess DR. Retrospective studies and chart reviews. *Respir Care* 2004;49:1171–1174.
79. Swanson E. Prospective photographic measurement study of 196 cases of breast augmentation, mastopexy, augmentation/mastopexy, and breast reduction. *Plast Reconstr Surg.* 2013;131:802e–819e.
80. Ching S, Thoma A, McCabe RE, et al. Measuring outcomes in aesthetic surgery: a comprehensive review of the literature. *Plast Reconstr Surg.* 2003;111:469–480; discussion 481.
81. Chung KC, Rohrich RJ. Measuring quality of surgical care: is it attainable? *Plast Reconstr Surg.* 2009;123:741–749.
82. Pollock H, Pollock T. Is reoperation rate a valid statistic in cosmetic surgery? *Plast Reconstr Surg.* 2007;120:569.
83. Goldwyn RM. Consecutive patients. *Plast Reconstr Surg.* 1990;86:962.
84. Swanson E. Objective assessment of face lifts. *Plast Reconstr Surg.* 2013;131:915e–916e.
85. Goldwyn RM. Wanted: real clinical results. *Plast Reconstr Surg.* 2004;114:1000–1001.
86. Chung KC, Kalliainen LK, Spilson SV, et al. The prevalence of negative studies with inadequate statistical power: an analysis of the plastic surgery literature. *Plast Reconstr Surg.* 2002;109:1–6; discussion 7.
87. Sackett DL, Straus SE, Richardson WS, et al. Therapy. In: *Evidence-based Medicine.* 2nd ed. Toronto, ON: Churchill Livingstone; 2000:105–153.
88. Kuzon WM Jr, Urbanchek MG, McCabe S. The seven deadly sins of statistical analysis. *Ann Plast Surg.* 1996;37:265–272.
89. Swanson E. The plastic surgeon: artist or scientist? *Plast Reconstr Surg.* 2013;131:182–184.
90. Lexchin J, Bero LA, Djulbegovic B, et al. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167–1170.
91. Hall-Findlay EJ. Discussion: late seromas and breast implants: theory and practice. *Plast Reconstr Surg.* 2012;130:436–438.
92. Lee MR, Unger JG, Rohrich RJ. Management of the nasal dorsum in rhinoplasty: a systematic review of the literature regarding technique, outcomes, and complications. *Plast Reconstr Surg.* 2011;128:538e–550e.