

Reading between the Lines: A Plastic Surgeon's Guide to Evaluating the Quality of Evidence in Research Publications

Allen Gabriel, MD, FACS*
G. Patrick Maxwell, MD*†

Summary: An important component of practicing evidence-based medicine (EBM) in plastic surgery is staying current with published research, an increasingly difficult task given the rapid growth of the literature. This article reviews the methodological aspects specific to the aesthetic surgery field that should be considered when evaluating the quality of evidence in research publications in the context of the level of evidence (LOE) grading scales that are currently used by plastic surgery journals. Reporting the LOE in a research publication can help to highlight the quality of the research and the potential for bias so that the reader may prioritize information accordingly. However, LOE classifications are not an absolute index of the quality of evidence and do not preclude careful evaluation of the study's methods and results in the context of the authors' conclusions. As the application of evidence-based medicine expands in the plastic surgery community, surgeons must be mindful of how to appropriately interpret research findings and assess the utility of applying results to their practice. (*Plast Reconstr Surg Glob Open* 2019;7:e2311; doi: 10.1097/GOX.0000000000002311; Published online 19 June 2019.)

INTRODUCTION

The evidence-based medicine (EBM) movement in plastic surgery strives to integrate the best research evidence with clinical expertise to enable rational treatment decisions and optimal patient outcomes.¹ An important component of practicing EBM is staying current with published research, an increasingly difficult task given the rapid growth of the literature.¹ To keep pace with the new research, busy clinicians may only have time to skim the abstract and conclusions of a publication without critically evaluating the methodology or the results of the study.² However, before applying research findings to their practice, clinicians must take responsibility for critically appraising the quality of the methods and results and assess how the authors' conclusions are framed in the context of the presented evidence.

There are many important questions to ask when reading a research article. The following are 3 helpful questions to consider³:

1. Does the research address a worthwhile question? Is it the correct question to ask?
2. Are the study's methods appropriate to answer the stated research question(s)?
3. Are the authors' conclusions justified by the data and the methods?

However, questions about the appropriateness of the study's methods and the justification of the authors' conclusions may be difficult to address if one is not profoundly familiar with the specific area of research. This article reviews methodological aspects specific to aesthetic surgery studies that should be considered when evaluating the quality of evidence in research publications and discusses these factors in the context of level of evidence (LOE) grading scales that are currently used by plastic surgery journals.

Assessing Methodological Quality of Plastic Surgery Studies

In general, randomized controlled trials (RCTs) and systematic reviews are considered to be the ideal for obtaining high quality evidence.^{4,5} However, a study cannot automatically be deemed "high-quality" based simply on these designations. There are many methodological factors to consider when evaluating the quality of evidence within a study.^{4,6} For example, studies conducted without a priori power analyses to determine sufficient sample sizes can be problematic and predispose a study to false-negative statistical errors.⁶ Improper study conduct and planning can also result in erroneous conclusions,⁶ as

*From the *Department of Plastic Surgery, Loma Linda University Medical Center, Loma Linda, Calif.; and †Vanderbilt University Medical Center, Nashville, Tenn.*

Received for publication April 12, 2019; accepted April 26, 2019.

Copyright © 2019 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/GOX.0000000000002311

Disclosure: Dr. Gabriel is a consultant for Allergan and KCI. Dr. Maxwell is a consultant for Allergan Medical.

can conclusions from studies with inherent biases, such as test/retest performance bias or recall bias.⁷ Studies that do not report effect sizes cannot be evaluated for the true magnitude of differences between study cohorts, even if *P* values are reported.⁸ The results of meta-analyses warrant critical evaluation based on the heterogeneity of the studies chosen for inclusion; studies with vastly different methodologies may not be suitable for a combined analysis.⁹

Cosmetic surgery is distinct from other medical fields because it is more commonly performed to correct an external defect for aesthetic and psychosocial reasons than it is to address an underlying pathology associated with an individual's health.¹⁰⁻¹² Measuring the effectiveness of an intervention from the patient's perspective is often more important than objective physical measurement in evaluating treatment success.⁶ Consequently, evaluating the use of validated patient-reported outcome measures that assess patient satisfaction and quality of life is an important component of assessing the quality of evidence in a plastic surgery study.¹³ Selecting the appropriate patient-reported outcome measure based on context of use and key concepts to be measured in a study is an important initial step in study design.¹⁴ However, challenges still remain in that patients and surgeons may have discordant perceptions of outcomes after cosmetic surgery, thus making results difficult to interpret.¹⁵

One recent publication illustrates the importance of evaluating a study's methods when interpreting the results. The study was undertaken to evaluate the risk of rare systemic harms associated with silicone breast implants by analyzing publicly available data from long-term postapproval studies of silicone breast implants.¹⁶ Yet, careful reading of the methods and results sections reveals several methodological flaws that lower the quality of evidence in this study. For example, the preponderance of data supporting the authors' conclusions of increased risks of systemic harms was derived from one study of 41,342 patients with a follow-up rate of 20%.¹⁶ The extremely low

follow-up rates resulted in a lack of outcomes data for the majority of the enrolled patients and thereby introduced substantial bias into the population.¹⁷

Level of Evidence Classification Scales Used by Plastic Surgery Journals

The use of hierarchical systems to classify a study's LOE has become a foundation of the EBM movement.¹⁸ The LOE rating scales, which stratify the quality of evidence from strongest to weakest on the basis of the study design and its susceptibility to bias, were first implemented by plastic surgery journals in 2011.^{4,18} Over the past decade, a number of variations of LOE grading scales have been published^{5,6,19} and some are currently used by plastic surgery journals.^{4,5,19} *Plastic and Reconstructive Surgery* and the *Aesthetic Surgery Journal* use the grading recommendations published by the American Society of Plastic Surgeons (ASPS) to assign an LOE rating to eligible publications (Table 1).^{4,18} *JAMA Facial Plastic Surgery* uses a modified version of the LOE ratings that were established by the Oxford Centre for Evidence-Based Medicine (OCEBM) (Fig. 1).^{5,19} Both grading systems are specific to the type of research question being asked (ie, prognostic/risk, diagnostic, or therapeutic questions). In general, both assign the highest LOE to systematic reviews and high-quality cohort studies or RCTs and lower levels to case reports.^{4,5}

Basic science, bench work, and cadaveric and animal studies are deemed nonratable by the ASPS and OCEBM scales because they are not considered to be directly applicable to patient treatment decisions.^{4,19} The OCEBM scale allows for downgrading the LOE rating on the basis of study quality, imprecision, or indirectness; because of inconsistency between studies; or because the absolute effect size is very small.⁵ The ASPS scale gives higher ratings to "high-quality" studies and lower ratings to "lesser quality" studies. However, neither scale clearly defines the factors that comprise a high-quality study.

Table 1. American Society of Plastic Surgeons' Levels of Evidence Grading Recommendations⁴

Level of Evidence	American Society of Plastic Surgeons		
	Therapeutic Studies	Diagnostic Studies	Prognostic/Risk Studies
I	High-quality, multicenter or single-center, randomized controlled trial with adequate power; or systematic review of these studies	High-quality, multicenter or single-center cohort study validating a diagnostic test (with "gold" standard as reference) in a series of consecutive patients; or a systematic review of these studies	High-quality, multicenter or single-center, prospective cohort or comparative study with adequate power; or a systematic review of these studies
II	Lesser-quality, randomized controlled trial; prospective cohort or comparative study; or systematic review of these studies	Exploratory cohort study developing diagnostic criteria (with "gold" standard as reference) in a series of consecutive patients or a systematic review of these studies	Lesser-quality prospective cohort or comparative study; retrospective cohort or comparative study; untreated controls from a randomized controlled trial; or a systematic review of these studies
III	Retrospective cohort or comparative study, case-control study, or systematic review of these studies	Diagnostic study in nonconsecutive patients (without consistently applied "gold" standard as reference) or a systematic review of these studies	Case-control study or systematic review of these studies
IV	Case series with pretest/posttest or only posttest	Case-control study or any of the above diagnostic studies in the absence of a universally accepted "gold" standard	Case series with pretest or posttest or only posttest
V	Expert opinion developed via consensus process; case report or clinical example; or evidence based on physiology, bench research, or "first principles"		

Adapted with permission from *Plast Reconstr Surg*. 2011;128:311-314.



Fig. 1. Levels of evidence established by the OCEBM.⁵ Modified with permission from OCEBM Levels of Evidence Working Group. *Level may be graded down on the basis of study quality, imprecision, indirectness; because of inconsistency between studies; or because the absolute effect size is very small. Level may be graded up if there is a large or very large effect size. †A systematic review is generally better than an individual study.

Table 2. Cosmetic Level of Evidence and Recommendation Scale⁶

Level of Evidence	Description	Recommendation
I	Randomized trial with a power analysis supporting sample sizes	A
II	<ul style="list-style-type: none"> • Prospective study, high inclusion rate (≥80%), and description of eligibility criteria • Objective measuring device (ie, not surgeon’s opinion) or patient-derived outcomes data • Power analysis if treatment effect is compared 	A
III	<ul style="list-style-type: none"> • No control or comparative cohort is needed if effect is profound • Retrospective case–control study using a contemporaneous control group 	B
IV	<ul style="list-style-type: none"> • Prospective clinical study with an inclusion rate <80% • Prospective study without controls or comparison group and a treatment effect that is not dramatic • Retrospective case series of consecutive patients 	C
V	<ul style="list-style-type: none"> • Case/control study using historical controls or controls from other publications • Important confounder that might explain treatment effect Case report, expert opinion, and nonconsecutive case series	D

Adapted with permission from *Plast Reconstr Surg Glob Open*. 2013;1:e66.

Several important indicators of the methodological quality of plastic surgery studies (ie, power analyses, adjustments for confounders, follow-up, and scale reliability) are not explicitly considered in the ASPS and OCEBM LOE scales.⁴⁻⁶ The Cosmetic Level of Evidence and Recommendation

(CLEAR) scale (Table 2) incorporates several of these factors into a modified version of the ASPS LOE scale.⁶ Although the CLEAR scale has not yet been adapted by any journals, it may be a helpful tool for plastic surgeons when assessing quality of evidence in the literature.

LOE in the Real World

Despite the implementation of LOE classification systems by plastic surgery journals, the percentage of level I/II studies remains low.^{20,21} An analysis of studies published in *Plastic and Reconstructive Surgery* in 2013 found that only 3% of the studies were level I and 15% were level II.²⁰ The paucity of level I plastic surgery publications has been attributed to the difficulty of conducting RCTs to address clinical questions in surgery compared with other fields.^{20,22} Although the numbers of RCTs reported in aesthetic surgery journals have increased several-fold since the 1980s,^{23,24} many questions in the practice of plastic surgery remain unamenable to RCTs. One analysis of treatment evaluation questions in the surgical literature estimated that 61% of surgical questions are not suitable for randomized trials.²⁵ In addition, differences in operative technique from surgeon to surgeon are often unavoidable, making it difficult to control for some methodological aspects of an intervention.^{6,23,24} Placebos, sham surgeries, and blinding are usually not feasible, and randomization is not well accepted by patients and surgeons.^{12,23,24,26,27} Surgeons may feel it is unethical to randomize patients into different groups if one of the treatments is known to be inferior.^{6,27} Study designs that introduce a control by involving different interventions on contralateral sides of the face or body may produce asymmetry.⁶

Historically, publications with lower LOE ratings have been relatively more valuable in plastic surgery than in many other specialties.²⁸ A review of the 50 most cited papers published in the 20 highest impact plastic surgery journals between 1945 and 2015 found that none of the top 50 publications were rated level I or II.²⁸ The majority (84%) of the top 50 cited papers was categorized as level IV or V evidence and the average LOE rating for the top 50 papers was 4.28.²⁸ Although there is a need for more level I/II studies in the literature, high-quality observational studies (ie, case series and retrospective cohort studies), which by default receive lower LOE ratings,⁴ play an essential role in evidence-based plastic surgery.²⁹ Observational studies can provide valuable data for patient populations that would likely be excluded from RCTs, such as those with complex, multifaceted conditions and comorbidities.²⁹ Given that the study type and LOE rating do not necessarily reflect the overall importance of a study's results, it is the reader's responsibility to critically appraise the significance and accuracy of study conclusions in the context of the methodology, presented results, and other publications in the field.

CONCLUSIONS

Reporting the LOE in a research publication can help to highlight the quality of the research and the potential for bias, so that the reader may prioritize information accordingly. However, the LOE classification should not be considered an absolute index of the quality of evidence and does not preclude careful evaluation of the study's methods and results in the context of the authors' conclusions. As the plastic surgery community continues to adapt to the concept of EBM, surgeons need to be continually

mindful of how to appropriately interpret research findings and assess the utility of applying the results to their practice.

Allen Gabriel, MD, FACS

Department of Plastic Surgery
Loma Linda University Medical Center
505 NE, 87th Avenue, Suite 250
Vancouver, WA 98664
E-mail: gabriellallen@yahoo.com

ACKNOWLEDGMENTS

Writing and editorial assistance was provided to the authors by Lela Creutz, PhD, of Peloton Advantage, Parsippany, N.J., and was funded by Allergan plc, Dublin, Ireland.

References

1. Chung KC, Swanson JA, Schmitz D, et al. Introducing evidence-based medicine to plastic and reconstructive surgery. *Plast Reconstr Surg*. 2009;123:1385–1389.
2. Govindarajan R, Narayanaswami P. Evidence-based medicine for every day, everyone, and every therapeutic study. *Muscle Nerve*. 2018;58:486–496.
3. Agha RA, Orgill DP. Evidence-based plastic surgery: its rise, importance, and a practical guide. *Aesthet Surg J*. 2016;36:366–371.
4. 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.
5. OCEBM Levels of Evidence Working Group. The Oxford 2011 Levels of Evidence. Available at: <https://www.cebm.net/2016/05/ocebml-levels-of-evidence/>. Accessed August 29, 2018.
6. Swanson E. Levels of Evidence in cosmetic surgery: analysis and recommendations using a new CLEAR classification. *Plast Reconstr Surg Glob Open*. 2013;1:e66.
7. Indrayan A. Varieties of bias to guard against. In: Indrayan A, ed. *Medical Biostatistics*. 3rd ed. Boca Raton, Fla.: CRC Press; 2013: 54–62.
8. Sullivan GM, Feinn R. Using effect size-or why the p value is not enough. *J Grad Med Educ*. 2012;4:279–282.
9. Walker E, Hernandez AV, Kattan MW. Meta-analysis: Its strengths and limitations. *Cleve Clin J Med*. 2008;75:431–439.
10. Barone M, Cogliandro A, Persichetti P. Evidence-based medicine in plastic surgery. *Eur Arch Otorhinolaryngol*. 2017;274:3533–3534.
11. von Soest T, Kvaalem IL, Skolleborg KC, et al. Psychosocial changes after cosmetic surgery: a 5-year follow-up study. *Plast Reconstr Surg*. 2011;128:765–772.
12. 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.
13. Pusic AL, Lemaire V, Klassen AF, et al. Patient-reported outcome measures in plastic surgery: use and interpretation in evidence-based medicine. *Plast Reconstr Surg*. 2011;127:1361–1367.
14. Snyder CF, Watson ME, Jackson JD, et al. Patient-reported outcome instrument selection: designing a measurement strategy. *Value Health*. 2007;10(Suppl 2):S76–85.
15. Breiting LB, Henriksen TF, Kalialis LV, et al. A prospective study of short- and long-term cosmetic outcome after reduction mammoplasty from three different perspectives: the patient, a department surgeon, and an independent private practitioner in plastic surgery. *Plast Reconstr Surg*. 2012;130:273–281.
16. Coroneos CJ, Selber JC, Offodile AC II, et al. US FDA breast implant postapproval studies: long-term outcomes in 99,993 patients. *Ann Surg*. 2019;269:30–36.
17. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: a meta-analysis. *Public Opin Q*. 2008;72:167–189.

18. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. 2011;128:305–310.
19. Rhee JS, Larrabee WF. Facing levels of evidence: the JAMA facial plastic surgery initiative. *JAMA Facial Plast Surg*. 2013;15:174–175.
20. Nguyen A, Mahabir RC. An update on the level of evidence for plastic surgery research published in *Plastic and Reconstructive Surgery*. *Plast Reconstr Surg Glob Open*. 2016;4:e798.
21. Leal DG, Rodrigues MA, Tedesco ACB, et al. Evidence-based medicine in plastic surgery: are we there yet? *Ann Plast Surg*. 2018;80:71–75.
22. Ricci JA, Desai NS. Evidence-based medicine in plastic surgery: where did it come from and where is it going? *J Evid Based Med*. 2014;7:68–71.
23. 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.
24. Loielle F, Mahabir RC, Harrop AR. Levels of evidence in plastic surgery research over 20 years. *Plast Reconstr Surg*. 2008;121:207e–211e.
25. Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? *Surgery*. 1995;118:459–467.
26. 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.
27. McCarthy CM, Collins ED, Pusic AL. Where do we find the best evidence? *Plast Reconstr Surg*. 2008;122:1942–1947; discussion 1948–1951.
28. Joyce KM, Joyce CW, Kelly JC, et al. Levels of evidence in the plastic surgery literature: a citation analysis of the top 50 “Classic” papers. *Arch Plast Surg*. 2015;42:411–418.
29. Kowalski E, Chung KC. The outcomes movement and evidence-based medicine in plastic surgery. *Clin Plast Surg*. 2013;40:241–247.