



Published in final edited form as:

Plast Reconstr Surg. 2011 July ; 128(1): 305–310. doi:10.1097/PRS.0b013e318219c171.

The Levels of Evidence and their role in Evidence-Based Medicine

Patricia B. Burns, MPH¹, Rod J. Rohrich, MD², and Kevin C. Chung, MD, MS³

¹Research Associate, Section of Plastic Surgery, Department of Surgery, The University of Michigan Health System

²Professor of Surgery, Department of Plastic Surgery, University of Texas Southwestern Medical Center

³Professor of Surgery, Section of Plastic Surgery, Department of Surgery, The University of Michigan Health System

Abstract

As the name suggests, evidence-based medicine (EBM), is about finding evidence and using that evidence to make clinical decisions. A cornerstone of EBM is the hierarchical system of classifying evidence. This hierarchy is known as the levels of evidence. Physicians are encouraged to find the highest level of evidence to answer clinical questions. Several papers published in Plastic Surgery journals concerning EBM topics have touched on this subject.^{1–6} Specifically, previous papers have discussed the lack of higher level evidence in PRS and need to improve the evidence published in the journal. Before that can be accomplished, it is important to understand the history behind the levels and how they should be interpreted. This paper will focus on the origin of levels of evidence, their relevance to the EBM movement and the implications for the field of plastic surgery as well as the everyday practice of plastic surgery.

Keywords

Evidence-based medicine; levels of evidence

History of Levels of Evidence

The levels of evidence were originally described in a report by the Canadian Task Force on the Periodic Health Examination in 1979.⁷ The report's purpose was to develop recommendations on the periodic health exam and base those recommendations on evidence in the medical literature. The authors developed a system of rating evidence (Table 1) when determining the effectiveness of a particular intervention. The evidence was taken into account when grading recommendations. For example, a Grade A recommendation was given if there was good evidence to support a recommendation that a condition be included in the periodic health exam. The levels of evidence were further described and expanded by Sackett⁸ in an article on levels of evidence for antithrombotic agents in 1989 (Table 2). Both

Corresponding Author: Kevin C. Chung, MD, MS, Section of Plastic Surgery, University of Michigan Health System, 2130 Taubman Center, SPC 5340, 1500 E. Medical Center Drive, Ann Arbor, MI, 48109-5340, kecchung@umich.edu, Phone 734-936-5885, Fax 734-763-5354.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

systems place randomized controlled trials (RCT) at the highest level and case series or expert opinions at the lowest level. The hierarchies rank studies according to the probability of bias. RCTs are given the highest level because they are designed to be unbiased and have less risk of systematic errors. For example, by randomly allocating subjects to two or more treatment groups, these types of studies also randomize confounding factors that may bias results. A case series or expert opinion is often biased by the author's experience or opinions and there is no control of confounding factors.

Modification of levels

Since the introduction of levels of evidence, several other organizations and journals have adopted variation of the classification system. Diverse specialties are often asking different questions and it was recognized that the type and level of evidence needed to be modified accordingly. Research questions are divided into the categories: treatment, prognosis, diagnosis, and economic/decision analysis. For example, Table 3 shows the levels of evidence developed by the American Society of Plastic Surgeons (ASPS) for prognosis⁹ and Table 4 shows the levels developed by the Centre for Evidence Based Medicine (CEBM) for treatment.¹⁰ The two tables highlight the types of studies that are appropriate for the question (prognosis versus treatment) and how quality of data is taken into account when assigning a level. For example, RCTs are not appropriate when looking at the prognosis of a disease. The question in this instance is: "What will happen if we do nothing at all"? Because a prognosis question does not involve comparing treatments, the highest evidence would come from a cohort study or a systematic review of cohort studies. The levels of evidence also take into account the quality of the data. For example, in the chart from CEBM, poorly designed RCTs have the same level of evidence as a cohort study.

A grading system that provides strength of recommendations based on evidence has also changed over time. Table 5 shows the Grade Practice Recommendations developed by ASPS. The grading system provides an important component in evidence-based medicine and assists in clinical decision making. For example, a strong recommendation is given when there is level I evidence and consistent evidence from Level II, III and IV studies available. The grading system does not degrade lower level evidence when deciding recommendations if the results are consistent.

Interpretation of levels

Many journals assign a level to the papers they publish and authors often assign a level when submitting an abstract to conference proceedings. This allows the reader to know the level of evidence of the research but the designated level of evidence does always guarantee the quality of the research. It is important that readers not assume that level 1 evidence is always the best choice or appropriate for the research question. This concept will be very important for all of us to understand as we evolve into the field of EBM in Plastic Surgery. By design, our designated surgical specialty will always have important articles that may have a lower level of evidence due to the level of innovation and technique articles which are needed to move our surgical specialty forward.

Although RCTs are the often assigned the highest level of evidence, not all RCTs are conducted properly and the results should be carefully scrutinized. Sackett⁸ stressed the importance of estimating types of errors and the power of studies when interpreting results from RCTs. For example, a poorly conducted RCT may report a negative result due to low power when in fact a real difference exists between treatment groups. Scales such as the Jadad scale have been developed to judge the quality of RCTs.¹¹ Although physicians may not have the time or inclination to use a scale to assess quality, there are some basic items that should be taken into account. Items used for assessing RCTs include: randomization,

blinding, a description of the randomization and blinding process, description of the number of subjects who withdrew or drop out of the study; the confidence intervals around study estimates; and a description of the power analysis. For example, Bhandari et al.¹² published a paper assessing the quality of surgical RCTs. The authors evaluated the quality of RCTs reported in the *Journal of Bone and Joint Surgery (JBJS)* from 1988–2000. Papers with a score of > 75% were deemed high quality and 60% of the papers had a score < 75%. The authors identified 72 RCTs during this time period and the mean score was 68%. The main reason for the low quality score was lack of appropriate randomization, blinding, and a description of patient exclusion criteria. Another paper found the same quality score of papers in *JBJS* with a level 1 rating compared to level 2.¹³ Therefore, one should not assume that level 1 studies have higher quality than level 2.

A resource for surgeons when appraising levels of evidence are the users' guides published in the *Canadian Journal of Surgery*^{14, 15} and the *Journal of Bone and Joint Surgery*.¹⁶ Similar papers that are not specific to surgery have been published in the *Journal of the American Medical Association (JAMA)*.^{17, 18}

Plastic surgery and EBM

The field of plastic surgery has been slow to adopt evidence-based medicine. This was demonstrated in a paper examining the level of evidence of papers published in PRS.¹⁹ The authors assigned levels of evidence to papers published in PRS over a 20 year period. The majority of studies (93% in 1983) were level 4 or 5, which denotes case series and case reports. Although the results are disappointing, there was some improvement over time. By 2003 there were more level 1 studies (1.5%) and fewer level 4 and 5 studies (87%). A recent analysis looked at the number of level 1 studies in 5 different plastic surgery journals from 1978–2009. The authors defined level 1 studies as RCTs and meta-analysis and restricted their search to these studies. The number of level 1 studies increased from 1 in 1978 to 32 by 2009.²⁰ From these results, we see that the field of plastic surgery is improving the level of evidence but still has a way to go, especially in improving the quality of studies published. For example, approximately a third of the studies involved double blinding, but the majority did not randomize subjects, describe the randomization process, or perform a power analysis. Power analysis is another area of concern in plastic surgery. A review of the plastic surgery literature found that the majority of published studies have inadequate power to detect moderate to large differences between treatment groups.²¹ No matter what the level of evidence for a study, if it is under powered, the interpretation of results is questionable.

Although the goal is to improve the overall level of evidence in plastic surgery, this does not mean that all lower level evidence should be discarded. Case series and case reports are important for hypothesis generation and can lead to more controlled studies. Additionally, in the face of overwhelming evidence to support a treatment, such as the use of antibiotics for wound infections, there is no need for an RCT.

Clinical examples using levels of evidence

In order to understand how the levels of evidence work and aid the reader in interpreting levels, we provide some examples from the plastic surgery literature. The examples also show the peril of medical decisions based on results from case reports.

An association was hypothesized between lymphoma and silicone breast implants based on case reports.^{22–27} The level of evidence for case reports, depending on the scale used, is 4 or 5. These case reports were used to generate the hypothesis that a possible association existed. Because of these results, several large retrospective cohort studies from the United States, Canada, Denmark, Sweden and Finland were conducted.^{28–32} The level of evidence

for a retrospective cohort is 2. All of these studies had many years of follow-up for a large number of patients. Some of the studies found an elevated risk and others no risk for lymphoma. None of the studies reached statistical significance. Therefore, higher level evidence from cohort studies does not provide evidence of any risk of lymphoma. Finally, a systematic review was performed that combined the evidence from the retrospective cohorts.²⁷ The results found an overall standardized incidence ratio of 0.89 (95% CI 0.67–1.18). Because the confidence intervals include 1, the results indicate there is no increased incidence. The level of evidence for the systematic review is 1. Based on the best available evidence, there is no association between lymphoma and silicone implants. This example shows how low level evidence studies were used to generate a hypothesis, which then led to higher level evidence that disproved the hypothesis. This example also demonstrates that RCTs are not feasible for rare events such as cancer and the importance of observational studies for a specific study question. A case-control study is a better option and provides higher evidence for testing the prognosis of the long-term effect of silicone breast implants.

Another example is the injection of epinephrine in fingers. Based on case reports prior to 1950, physicians were advised that epinephrine injection can result in finger ischemia.³³ We see in this example in which level 4 or 5 evidence was accepted as fact and incorporated into medical textbooks and teaching. However, not all physicians accepted this evidence and are performing injections of epinephrine into the fingers with no adverse effects on the hand. Obviously, it was time for higher level evidence to resolve this issue. An in-depth review of the literature from 1880 to 2000 by Denkler,³³ identified 48 cases of digital infarction of which 21 were injected with epinephrine. Further analysis found that the addition of procaine to the epinephrine injection was the cause of the ischemia.³⁴ The procaine used in these injections included toxic acidic batches that were recalled in 1948. In addition, several cohort studies found no complications from the use of epinephrine in the fingers and hand.^{35,36, 37} The results from these cohort studies increased the level of evidence. Based on the best available evidence from these studies, the hypothesis that epinephrine injection will harm fingers was rejected. This example highlights the biases inherent in case reports. It also shows the risk when spurious evidence is handed down and integrated into medical teaching.

Obtaining the best evidence

We have established the need for RCTs to improve evidence in plastic surgery but have also acknowledged the difficulties, particularly with randomization and blinding. Although RCTs may not be appropriate for many surgical questions, well designed and conducted cohort or case-control studies could boost the level of evidence. Many of the current studies tend to be descriptive and lack a control group. The way forward seems clear. Plastic surgery researchers need to consider utilizing a cohort or case-control design whenever an RCT is not possible. If designed properly, the level of evidence for observational studies can approach or surpass those from an RCT. In some instances, observation studies and RCTs have found similar results.³⁸ If enough cohort or case-control studies become available, this increases the prospect of systematic reviews of these studies that will increase overall evidence levels in plastic surgery.

Conclusion

The levels of evidence are an important component of EBM. Understanding the levels and why they are assigned to publications and abstracts helps the reader to prioritize information. This is not to say that all level 4 evidence should be ignored and all level 1 evidence accepted as fact. The levels of evidence provide a guide and the reader needs to be cautious when interpreting these results.

Acknowledgments

Supported in part by a Midcareer Investigator Award in Patient-Oriented Research (K24 AR053120) from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (to Dr. Kevin C. Chung).

References

1. McCarthy CM, Collins ED, Pusic AL. Where do we find the best evidence? *Plast Reconstr Surg.* 2008; 122:1942–1947. [PubMed: 19050548]
2. Chung KC, Swanson JA, Schmitz D, Sullivan D, Rohrich RJ. Introducing evidence-based medicine to plastic and reconstructive surgery. *Plast Reconstr Surg.* 2009; 123:1385–1389. [PubMed: 19337107]
3. Chung KC, Ram AN. Evidence-based medicine: the fourth revolution in American medicine? *Plast Reconstr Surg.* 2009; 123:389–398. [PubMed: 19116577]
4. Rohrich RJ. So you want to be better: the role of evidence-based medicine in plastic surgery. *Plast Reconstr Surg.* 2010; 126:1395–1398. [PubMed: 20885263]
5. Burns PB, Chung KC. Developing good clinical questions and finding the best evidence to answer those questions. *Plast Reconstr Surg.* 2010; 126:613–618. [PubMed: 20679843]
6. Sprague S, McKay P, Thoma A. Study design and hierarchy of evidence for surgical decision making. *Clin Plast Surg.* 2008; 35:195–205. [PubMed: 18298992]
7. The periodic health examination. Canadian Task Force on the Periodic Health Examination. *Can Med Assoc J.* 1979; 121:1193–1254. [PubMed: 115569]
8. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest.* 1989; 95:2S–4S. [PubMed: 2914516]
9. American Society of Plastic Surgeons. Available at: http://www.plasticsurgery.org/Medical_Professionals/Health_Policy_and_Advocacy/Health_Policy_Resources/Evidence-based_GuidelinesPractice_Parameters/Description_and_Development_of_Evidence-based_Practice_Guidelines/ASPS_Evidence_Rating_Scales.html
10. Centre for Evidence Based Medicine. Available at <http://www.cebm.net>
11. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials.* 1996; 17:1–12. [PubMed: 8721797]
12. Bhandari M, Richards RR, Sprague S, Schemitsch EH. 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. [PubMed: 11886908]
13. Poolman RW, Struijs PA, Krips R, Siersevelt IN, Lutz KH, Bhandari M. Does a "Level I Evidence" rating imply high quality of reporting in orthopaedic randomised controlled trials? *BMC Med Res Methodol.* 2006; 6:44. [PubMed: 16965628]
14. Urschel JD, Goldsmith CH, Tandan VR, Miller JD. Users' guide to evidence-based surgery: how to use an article evaluating surgical interventions. Evidence-Based Surgery Working Group. *Can J Surg.* 2001; 44:95–100. [PubMed: 11308245]
15. Thoma A, Farrokhyar F, Bhandari M, Tandan V. Users' guide to the surgical literature. How to assess a randomized controlled trial in surgery. *Can J Surg.* 2004; 47:200–208. [PubMed: 15264385]
16. Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about prognosis. *J Bone Joint Surg Am.* 2001; 83-A:1555–1564. [PubMed: 11679610]
17. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA.* 1993; 270:2598–2601. [PubMed: 8230645]
18. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. *JAMA.* 2000; 284:1290–1296. [PubMed: 10979117]
19. Loiselle F, Mahabir RC, Harrop AR. Levels of evidence in plastic surgery research over 20 years. *Plast Reconstr Surg.* 2008; 121:207e–211e.

20. McCarthy JE, Chatterjee A, McKelvey TG, Jantzen EM, Kerrigan CL. 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. [PubMed: 21042137]
21. Chung KC, Kalliainen LK, Spilson SV, Walters MR, Kim HM. The prevalence of negative studies with inadequate statistical power: an analysis of the plastic surgery literature. *Plast Reconstr Surg.* 2002; 109:1–6. [PubMed: 11786783]
22. Newman MK, Zimmel NJ, Bandak AZ, Kaplan BJ. Primary breast lymphoma in a patient with silicone breast implants: a case report and review of the literature. *J Plast Reconstr Aesthet Surg.* 2008; 61:822–825. [PubMed: 17509956]
23. Gaudet G, Friedberg JW, Weng A, Pinkus GS, Freedman AS. Breast lymphoma associated with breast implants: two case-reports and a review of the literature. *Leuk Lymphoma.* 2002; 43:115–119. [PubMed: 11908714]
24. Sahoo S, Rosen PP, Feddersen RM, Viswanatha DS, Clark DA, Chadburn A. Anaplastic large cell lymphoma arising in a silicone breast implant capsule: a case report and review of the literature. *Arch Pathol Lab Med.* 2003; 127:e115–e118. [PubMed: 12653596]
25. Keech JA Jr, Creech BJ. Anaplastic T-cell lymphoma in proximity to a saline-filled breast implant. *Plast Reconstr Surg.* 1997; 100:554–555. [PubMed: 9252643]
26. Duvic M, Moore D, Menter A, Vonderheid EC. Cutaneous T-cell lymphoma in association with silicone breast implants. *J Am Acad Dermatol.* 1995; 32:939–942. [PubMed: 7751462]
27. Lipworth L, Tarone RE, McLaughlin JK. Breast implants and lymphoma risk: a review of the epidemiologic evidence through 2008. *Plast Reconstr Surg.* 2009; 123:790–793. [PubMed: 19319041]
28. Lipworth L, Tarone RE, Friis S, et al. Cancer among Scandinavian women with cosmetic breast implants: a pooled long-term follow-up study. *International Journal of Cancer.* 2009; 124:490–493.
29. Deapen DM, Hirsch EM, Brody GS. Cancer risk among Los Angeles women with cosmetic breast implants. *Plast Reconstr Surg.* 2007; 119:1987–1992. [PubMed: 17519689]
30. Brisson J, Holowaty EJ, Villeneuve PJ, et al. Cancer incidence in a cohort of Ontario and Quebec women having bilateral breast augmentation. *Int J Cancer.* 2006; 118:2854–2862. [PubMed: 16381020]
31. Pukkala E, Boice JD Jr, Hovi SL, et al. Incidence of breast and other cancers among Finnish women with cosmetic breast implants, 1970–1999. *J Long Term Eff Med Implants.* 2002; 12:271–279. [PubMed: 12627789]
32. Brinton LA, Lubin JH, Burich MC, Colton T, Brown SL, Hoover RN. Cancer risk at sites other than the breast following augmentation mammoplasty. *Ann Epidemiol.* 2001; 11:248–256. [PubMed: 11306343]
33. Denkler K. A comprehensive review of epinephrine in the finger: to do or not to do. *Plast Reconstr Surg.* 2001; 108:114–124. [PubMed: 11420511]
34. Thomson CJ, Lalonde DH, Denkler KA, Feicht AJ. A critical look at the evidence for and against elective epinephrine use in the finger. *Plast Reconstr Surg.* 2007; 119:260–266. [PubMed: 17255681]
35. Lalonde D, Bell M, Benoit P, Sparkes G, Denkler K, Chang P. A multicenter prospective study of 3,110 consecutive cases of elective epinephrine use in the fingers and hand: the Dalhousie Project clinical phase. *J Hand Surg Am.* 2005; 30:1061–1067. [PubMed: 16182068]
36. Chowdhry S, Seidenstricker L, Cooney DS, Hazani R, Wilhelmi BJ. Do not use epinephrine in digital blocks: myth or truth? Part II. A retrospective review of 1111 cases. *Plast Reconstr Surg.* 2010; 126:2031–2034. [PubMed: 20697319]
37. Wilhelmi BJ, Blackwell SJ, Miller JH, et al. Do not use epinephrine in digital blocks: myth or truth? *Plast Reconstr Surg.* 2001; 107:393–397. [PubMed: 11214054]
38. 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. [PubMed: 10861325]

Table 1

Canadian Task Force on the Periodic Health Examination's Levels of Evidence*

| Level | Type of evidence |
|-------|---|
| I | At least 1 RCT with proper randomization |
| II.1 | Well designed cohort or case-control study |
| II.2 | Time series comparisons or dramatic results from uncontrolled studies |
| III | Expert opinions |

* Adapted from Canadian Task Force on the Periodic Health Examination. The periodic health examination. *Can Med Assoc J* 1979;121:1193-254

Table 2

Levels of Evidence from Sackett*

| Level | Type of evidence |
|-------|---|
| I | Large RCTs with clear cut results |
| II | Small RCTs with unclear results |
| III | Cohort and case-control studies |
| IV | Historical cohort or case-control studies |
| V | Case series, studies with no controls |

* Adapted from Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1989;95:2S-4S

Table 3

Levels of Evidence for Prognostic Studies*

| Level | Type of evidence |
|-------|--|
| I | High quality prospective cohort study with adequate power or systematic review of these studies |
| II | Lesser quality prospective cohort, retrospective cohort study, untreated controls from an RCT, or systematic review of these studies |
| III | Case-control study or systematic review of these studies |
| IV | Case series |
| V | Expert opinion; case report or clinical example; or evidence based on physiology, bench research or "first principles" |

* Adapted from the American Society of Plastic Surgeons,
http://www.plasticsurgery.org/Medical_Professionals/Health_Policy_and_Advocacy/Health_Policy_Resources/Evidence-based_GuidelinesPractice_Parameters/Description_and_Development_of_Evidence-based_Practice_Guidelines/ASPS_Evidence_Rating_Scales.html.

Table 4

Levels of Evidence for Therapeutic Studies*

| Level | Type of evidence |
|-------|--|
| 1A | Systematic review (with homogeneity) of RCTs |
| 1B | Individual RCT (with narrow confidence intervals) |
| 1C | All or none study |
| 2A | Systematic review (with homogeneity) of cohort studies |
| 2B | Individual Cohort study (including low quality RCT, e.g. <80% follow-up) |
| 2C | “Outcomes” research; Ecological studies |
| 3A | Systematic review (with homogeneity) of case-control studies |
| 3B | Individual Case-control study |
| 4 | Case series (and poor quality cohort and case-control study) |
| 5 | Expert opinion without explicit critical appraisal or based on physiology bench research or “first principles” |

* From the Centre for Evidence-Based Medicine, <http://www.cebm.net>.

Table 5

Grade Practice Recommendations*

| Grade | Descriptor | Qualifying Evidence | Implications for Practice |
|-------|-----------------------|--|---|
| A | Strong recommendation | Level I evidence or consistent findings from multiple studies of levels II, III, or IV | Clinicians should follow a strong recommendation unless a clear and compelling rationale for an alternative approach is present |
| B | Recommendation | Levels II, III, or IV evidence and findings are generally consistent | Generally, clinicians should follow a recommendation but should remain alert to new information and sensitive to patient preferences |
| C | Option | Levels II, III, or IV evidence, but findings are inconsistent | Clinicians should be flexible in their decision-making regarding appropriate practice, although they may set bounds on alternatives; patient preference should have a substantial influencing role |
| D | Option | Level V evidence: little or no systematic empirical evidence | Clinicians should consider all options in their decision making and be alert to new published evidence that clarifies the balance of benefit versus harm; patient preference should have a substantial influencing role |

* From the American Society of Plastic Surgeons. Evidence-based clinical practice guidelines. Available at: http://www.plasticsurgery.org/Medical_Professionals/Health_Policy_and_Advocacy/Health_Policy_Resources/Evidence-based_GuidelinesPractice_Parameters/Description_and_Development_of_Evidence-based_Practice_Guidelines/ASPS_Grade_Recommendation_Scale.html. Accessed March 3, 2011