

Guest Editorial

A Primer on Reviewing and Synthesizing Evidence

Donald S. Likosky, PhD

Departments of Surgery and Community and Family Medicine, Dartmouth Medical School, Hanover, New Hampshire

Abstract: In this editorial, I shall describe methods commonly used for synthesizing evidence and review some contemporary examples of how this methodology has been applied to the peer-reviewed literature. It is my hope that the reader will find this

interesting, but more importantly useful as you find yourself reading the evermore voluminous medical literature. **Keywords:** reviewing literature, methodology, study design. *JECT. 2006;38: 112–115*

As we all know, the current financial environment in our health care system is staggering. In 1960, health care spending accounted for 5% of the gross domestic product, in 2004 rose to 16% and in 2014 is projected to account for 18.7% (1). We know, however, that spending more on health care does not necessarily equate with improving quality (2). In fact, in many cases more health care equates with decreasing quality. Variation in the conduct and delivery of care often stems from uncertainty as to the benefit of any given clinical intervention. This should not be surprising. There were 3868 articles added to Index Medicus during calendar year 2003, and another 3587 in calendar year 2004 on the topic of cardiopulmonary bypass (CPB) alone. How can we possibly keep current with the increasingly insurmountable literature? To complicate matters, not all articles are “created equal.” That is, there is a heterogeneity of study methodology and execution that should be accounted for to understand how this new science fits in with previous science.

HETEROGENEITY OF STUDIES

Not all studies are created equally. Simple, yet true. To highlight this point, I have provided a short synopsis of two articles that both focused on predicting neurologic injury after cardiac surgery.

Gardner et al., in 1985, published a study of their 10-year experience of patients undergoing coronary artery bypass grafting (CABG) surgery between 1974 and 1983



Donald S. Likosky, PhD

at Johns Hopkins Hospital, Baltimore, MD (3). They reviewed records of 3279 patients, of whom 56 (1.7%) presented with a stroke during the index admission. Factors found significantly associated with the risk of stroke included age, cerebrovascular disease, severe atherosclerosis of the ascending aorta, considerable hypotension during or immediately after surgery, and extended duration of CPB.

More contemporaneously, Charlesworth et al., in 2003, published a study on this topic (4). These authors undertook a study of 33,062 patients from 1992 to 2001 within the northern New England Cardiovascular Disease Study Group. There were 532 (1.6%) strokes in this population. The authors developed a multiple logistic regression model to identify factors significantly associated with stroke, and included age, sex, presence of diabetes, pres-

Address correspondence to: Donald S. Likosky, PhD, Department of Surgery, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756. E-mail: donald.likosky@dartmouth.edu

ence of vascular disease, renal failure or creatinine ≥ 2 mg/dL, ejection fraction $< 40\%$, and urgent or emergency surgery.

So, why are there different results from these two studies? Both represent large experiences.

1. Is it simply secondary to different time periods (Gardner et al. focused on 1974–1985 and Charlesworth et al. on 1992–2001)? Unlikely, because the base rate of strokes was quite similar (1.7% vs. 1.6%).
2. Is it because of differences in study design? Potentially, because Gardner et al. conducted a case-control study where they reviewed medical records to identify risk factors among patients who experienced a stroke (defined as a case). They chose two patients who did not experience a stroke who immediately preceded each stroke patient during the study period (defined as a control). Charts on both cases and controls were abstracted. Risk factors were identified as present if they were noted in the medical record.

Note on a case-control study design: The goal of a case-control study is to identify individuals who did (defined as cases) or did not (defined as controls) experience the event of interest. Investigators then go back in time to identify risk factors that predispose individuals to experience the events of interest. To isolate the influence of these factors, investigators choose controls that are as similar as possible to cases in every facet except that they do not experience the event. The onus thus resides with the reader to inquire as to whether the investigators chose the most apt controls for each case.

Charlesworth et al. chose a different study design, namely a prospective cohort study. These investigators used a prospective surgical registry. Unlike Gardner et al., risk factors were identified as present or absent before the surgical operation, independent of a patient's presentation of a stroke after surgery.

Note on a prospective cohort study design: The goal of a prospective cohort study is to identify factors that predispose individuals to experience the event of interest (in this case a stroke). Investigators identify which risk factors predispose individuals to have the event. With cohort studies, the onus resides with the reader to inquire as to whether the influence of variables associated with the exposure and event (i.e., confounders) are appropriately adjusted away in the analysis.

As the focus in both the studies of Gardner et al. and Charlesworth et al. were the identification of risk factors, it seems that differences in study design likely led to discrepancies in findings. In the study of Gardner et al., the investigators relied on the medical record review to identify the presence or absence of risk factors, whereas Charlesworth et al. used a prospective registry.

DRAWING INFERENCES FROM THIS HETEROGENEITY

As a reader, what do we do now? We recognize that not all articles are the same, but how can we possibly spend the requisite time necessary to review not only every article, but compare and contrast their methodology? From a reader's standpoint, the goal must be to engage actively rather than passively each and every article. It is through this active engagement that one understands and appreciates the unique aspects of a given article. Additionally, readers must critique an article's conclusions and inferences based on the evidence brought forth in the investigation.

There really are a couple broad types of syntheses that a reader has available to him/her: meta-analyses and review articles.

Meta-Analyses

Meta-analysis is a method for combining the results of two or more research studies focused on a particular topic and subject matter. The first meta-analysis was conducted by Karl Pearson in 1904 (5). Meta-analyses, through their methodology, take into account heterogeneity of sample size and appropriately weight articles that are based on larger samples.

As an example, let us look at a recent meta-analysis by Parolari et al. on the effect of on- vs. off-pump CABG surgery on graft patency (Figure 1). It is critical for a reader to be able to walk through a figure such as the one seen below (called a Forest or Funnel Plot). On the far left column, Parolari et al. (6) show the different studies, here graded by quality. For each study, the authors list the number of patients with the outcome divided by the number of patients in each group ($n =$ number of nonpatent grafts/ $N =$ number of total grafts). The main effect of each study is weighted by the number of patients enrolled in each study. In this manuscript, the study by Khan (7) is given a small weight (3.19) because of a smaller enrollment, whereas Widimsky (8) is given a higher weight (68.83) based on a larger enrollment. An overall effect (listed as an odds ratio with confidence intervals) represents the size of the effect of having on-pump vs. off-pump surgery. For each surgery, a horizontal line is presented, along with a square. The size of the horizontal line represents the confidence level in the findings of a given study. A shorter line suggests a higher level of certainty, whereas a longer line represents less certainty. The size of the square represents the strength of the effect, with a bigger square reflecting a larger effect. For each group, an overall effect is provided and represented with a diamond, the size of which again reflects the strength of the association. A vertical line is drawn, which indicates no difference in treatment effect (i.e., an odds ratio of 1.0). An odds ratio of 1.46 (the effect of the low-quality studies) indicates a

Review: OPCAB vs. CABG. Meta-analysis on graft patency.
 Comparison: 01 OPCAB vs. CABG
 Outcome: 02 Postoperative patency data

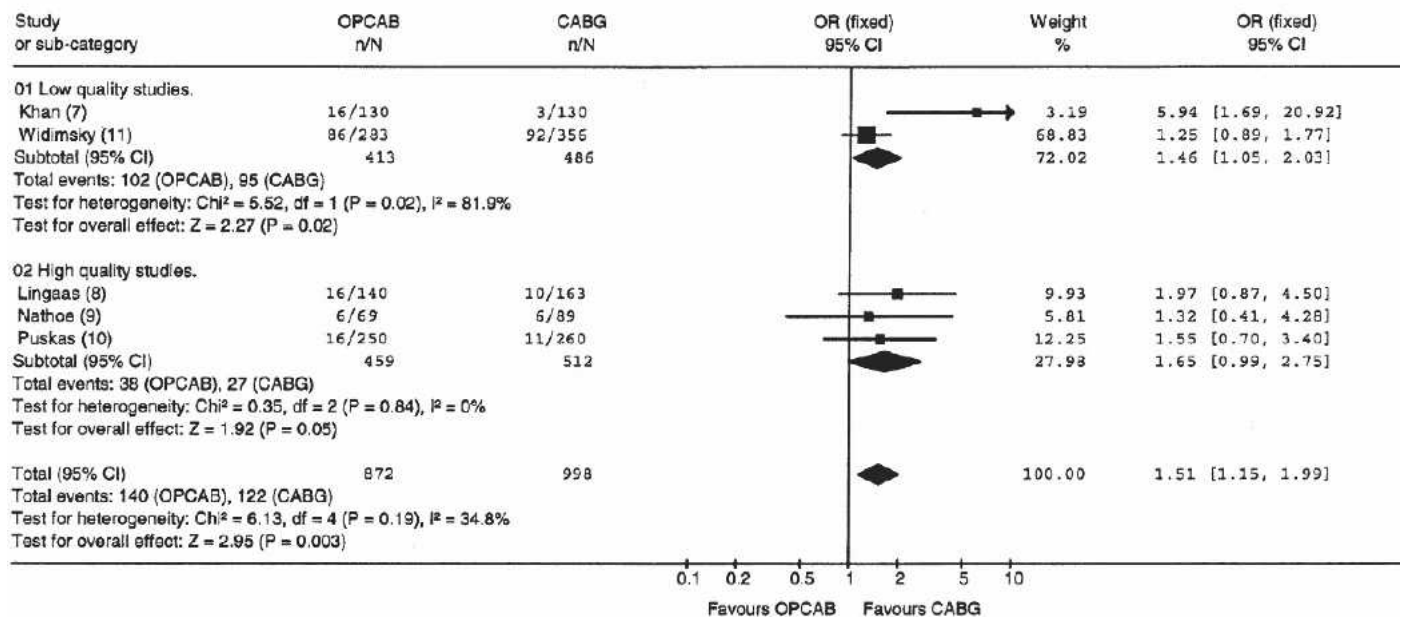


Figure 1. Example of a meta-analysis Reprinted from Parolari A, Alamanni F, Polvani G, et al. Meta-analysis of randomized trials comparing off-pump with on-pump coronary artery bypass graft patency. *Ann Thorac Surg.* 2005;80:2121–5.

46% increased risk of graft occlusion with off-pump CABG.

These types of analyses are extremely helpful for garnering a sense of the magnitude of effect present across multiple research projects. These articles do not account for heterogeneity of study designs. This issue might be accounted for through inclusion or exclusion criteria by the authors of the meta-analysis.

Review Articles

There are a variety of types of review articles that a reader is likely to encounter in the peer-reviewed literature. They may be broadly classified as non-structured and structured.

The non-structured type of review provides the reader with typically a broadly scoped review of the literature concerning a given topic. In many cases, the authors may or may not have used strict criteria for including articles into their manuscript. Additionally, this type of review article does not provide anything more than a broad overview of the literature from a potentially biased standpoint of the authors.

There are at least two types of structured review articles: clinical statements/guidelines and Cochrane Collaboratives. These reviews, when conducted well, offer readers a useful resource for understanding the strength and degree of evidence on a particular topic. There are an enormous, yet growing, number of clinical statements/guidelines that describe a desired clinical practice. While

some of these are written by clinical organizations and societies (9), others may derive from an individual or group's (10,11) synthesis of the peer-reviewed literature. These structured reviews are formulated through an explicit process of reviewing and grading the evidence. Variation in the findings of any one of these structured reviews may be explained by some of the following: choice of topic, inclusion criteria for identifying articles, methodology for critiquing articles, and methodology for grading and synthesizing the evidence base.

As a reader, one of the best ways to likely distinguish the good from the bad of structured reviews is to identify “gold star” attributes of well-done structured reviews. The American College of Cardiology and American Heart Association (ACC/AHA) has published a manual for developing guideline and clinical statements (<http://circ.ahajournals.org/manual/index.shtml>). The American Heart Association has been conducting such work for some time. This manual describes an explicit approach, a cookbook, for such work, including but not limited to developing search criteria and sorting and evaluating the evidence base. Additionally, they provide a schematic for classifying the recommendations (based on the strength of the evidence and agreement in studies) and level of evidence (the types of studies for which the evidence is derived) (Table 1).

Such departure from this “gold standard” explicit and objective methodology should raise the eyebrows from a reader's standpoint. For instance, there are a few guide-

Table 1. Level and class of evidence.

Classification of Recommendations		Level of Evidence	
Class I	Conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective	Level A	Data derived from multiple randomized clinical trials
Class II	Procedure/Treatment should be performed/administered	Level B	Data derived from a single randomized trial, or nonrandomized studies
Class IIA	Additional studies with focused objectives needed		
Class IIB	Additional studies with broad objectives needed; additional registry data would be helpful	Level C	Consensus opinion of experts
Class III	Procedure/Treatment should not be performed/administered since it is not helpful and may be harmful		

lines describing the practice of CPB, both providing contrary recommendations to the reader (10,11). Why might this be so? Both state that they used the methods promulgated by the ACC/AHA. On further review, however, Bartels et al. (10) did not fully implement these methods (both by including in vitro and animal studies, as well as using alternative methodology for deriving recommendation statements). Accordingly, an astute reader should discern these differences based on a careful reading of the methods section of these reviews. Without doing so, a reader may be misled by drawing inferences that are not necessarily grounded in objective criteria.

The Cochrane Collaborative (<http://www.cochrane.org/>) is an international, not-for-profit organization that provides up-to-date systematic reviews of health care-related issues. The Cochrane Collaborative, named after the world-renowned epidemiologist, Archie Cochrane, has, since 1993, provided a periodic and systematic review on >50 areas of medicine. Reviews are updated on a quarterly basis. Each review has a structured format: abstract, background, objectives, selection criteria for studies, search strategy, methods of the review, description of studies, methodological qualities of included studies, results, summary of analyses, conclusions, potential conflict of interest, acknowledgments, characteristics of included studies, characteristics of excluded studies, and references. The value of the Cochrane reviews is not only in their timeliness of material, but in their ability to provide explicit methodology for synthesizing the evidence base.

CONCLUSIONS

I hope that the reader of this article has gained some insight into the complexity of the medical literature and how to traverse through it. The key to success in this endeavor is in part to live according to Socrates' principles:

"I am the wisest man alive, for I know one thing, and that is that I know nothing."

By this, we succeed by knowing what we don't know, yet understanding the important questions to ask. It is a dif-

ficult task to be able to synthesize and keep pace with the growing and demanding literature. However, it is also vitally important to not take studies at their face value—we are obligated to be active and engaged readers and not passively accept the findings and conclusions of a given manuscript. We must recall that the key to science is reproducibility. Journal editors must require that authors make their methodology sufficient to allow for the reproducibility of their findings. Without this information, we cannot truly analyze, synthesize, and draw inferences across studies.

REFERENCES

1. The White House. State of the Union. Accessed 2006 at (<http://www.whitehouse.gov/stateoftheunion/2005/>).
2. Fisher ES, Wennberg, DE, Stukel TA, et al. The implications of regional variations in Medicare spending. Part 2: Health outcomes and satisfaction with care. *Ann Intern Med.* 2003;138:288–98.
3. Gardner TJ, Horneffer PJ, Manolio TA, et al. Stroke following coronary artery bypass grafting: A ten-year study. *Ann Thorac Surg.* 1985;40:574–81.
4. Charlesworth DC, Likosky DJ, Marrin CA, et al. Development and validation of a prediction model for strokes after coronary artery bypass grafting. *Ann Thorac Surg.* 2003;76:436–43.
5. Pearson K. Report on certain enteric fever inoculation statistics. *BMJ.* 1904;3:1243–6.
6. Parolari A, Alamanni F, Polvani G, et al. Meta-analysis of randomized trials comparing off-pump with on-pump coronary bypass graft patency. *Ann Thorac Surg.* 2005;80:2121–5.
7. Khan NE, De Souza A, Mister R, et al. A randomized comparison of off-pump and on-pump multivessel coronary-artery bypass surgery. *N Engl J Med.* 2004;350:21–8.
8. Widimsky P, Straka Z, Stros P, et al. One-year coronary bypass graft patency: a randomized comparison between off-pump and on-pump surgery angiographic results of the PRAGUE-4 trial. *Circulation.* 2004;110:3418–23.
9. Eagle KA, Guyton RA, Davidoff R, et al. ACC/AHA Guidelines for Coronary Artery Bypass Graft Surgery: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1991 Guidelines for Coronary Artery Bypass Graft Surgery). American College of Cardiology/American Heart Association. *J Am Coll Cardiol.* 1999;34:1262–347.
10. Bartels C, Gerdes A, Babin-Ebell J, et al. Cardiopulmonary bypass: Evidence or experience based? *J Thorac Cardiovasc Surg.* 2002;124:20–7.
11. Shann KG, Likosky DS, Murkin JM, et al. An evidence-based review of the practice of cardiopulmonary bypass in adults: A focus on neurological injury, glycemic control, hemodilution and the inflammatory response. *J Thorac Cardiovasc Surg.* (in press).