

# The GRADE System for Rating Clinical Guidelines

Brian P. Kavanagh<sup>1,2,3\*</sup>

**1** Departments of Critical Care Medicine and Anesthesia, Hospital for Sick Children, Toronto, Canada, **2** Program in Physiology and Experimental Medicine, The Research Institute, Hospital for Sick Children, Toronto, Canada, **3** Departments of Anesthesia, Medicine & Physiology, University of Toronto, Toronto, Canada

## Introduction

Professional medical groups commonly issue clinical practice guidelines. Such guidelines are traditionally the result of consensus conferences or expert panels and represent attempts to synthesize—from the best available evidence and expertise—practical guidance on the best possible care. Beyond issuing a guideline, many organizations have felt the need to provide a grading of each guideline's quality, thereby conveying to the reader a sense of the confidence that might be placed in it. This article addresses only the grading of guidelines, not their use or development.

The idea that evidence in the medical literature should be graded was initially proposed in publications from McMaster University [1–3], with the idea of categorizing individual studies into grades of reliability ranging from randomized controlled trials (most reliable) to case reports with expert opinion (least reliable). Grading of guidelines followed, but this has been besieged with problems. To give one example, a guideline by Ferraris and colleagues gave the use of aprotinin during high-risk cardiac surgery a “high-grade” recommendation [4], but this intervention was subsequently shown to increase mortality [5].

The pursuit of better approaches to grading guidelines has resulted in GRADE (Grades of Recommendation Assessment, Development and Evaluation), introduced in 2004 [6]. GRADE has been adopted “unchanged or with only minor modifications” by national and international professional medical societies, health-related branches of government, health care regulatory bodies, and UpToDate, an on-line medical resource that is accessed by trainees and physicians in most US academic medical centers (Box 1) [7,8].

The developers of the GRADE system emphasized consistency in the rating of

guidelines, as well as a wish to incorporate, and distinguish between, the “strength” of each guideline and the “quality” of the underlying studies (i.e., evidence) upon which it is based. Yet there is a central paradox: while GRADE has evolved through the evidence-based medicine movement, there is no evidence that GRADE itself is reliable.

## Are Different Guidelines Externally Consistent?

GRADE is one of several different systems for grading clinical evidence and creating clinical practice guidelines based on this underlying evidence. How do these systems compare with each other?

Atkins and colleagues, from the GRADE Working Group, compared six different systems (Box 2) [9]. Twelve assessors independently evaluated each system on the basis of 12 criteria to assess the “sensitivity” (overall usefulness) of the different approaches. There was poor agreement between them. In the absence of a proven gold standard, such disagreement signals concern about the inherent validity of any of these grading systems. Commenting on this lack of agreement, the authors wrote that a new system—GRADE—could overcome the problems [9].

But the example of the Surviving Sepsis Campaign (SSC), an important attempt to produce guidelines to improve the care of patients with sepsis or septic shock, suggests that GRADE has not overcome these problems (see Boxes 3 and 4) [10–11]. The endorsement of the SSC by many influential organizations underscores its importance [10–11]. Nonetheless, the SSC illustrates some of the important difficulties with grading in

general and with the GRADE system in particular. There are three reasons why I focus here on the SSC. First, sepsis encompasses all medical and surgical specialties, accounts for over 500,000 emergency visits per year in North America alone [12], and when accompanied by shock has a mortality of over 50% [13]. Second, the SSC may have significant impact: some believe that incorporating the SSC guidelines could save up to 100,000 lives in an 18-mo interval [14]. Third, the SSC is the best known source of advice on managing sepsis and all of its recommendations carry a grading. Finally, because the SSC published two documents 4 y apart (in 2004 and 2008 [10–11]), it presents a unique opportunity to compare interval changes. I focus only on grading (Boxes 3 and 4), not on the controversies surrounding the SSC [15], and I do not express support for—or criticism of—any of its recommendations.

## Is GRADE Internally Consistent?

**Inter-rater agreement of GRADE.** In 2005, the GRADE working group—all experts who themselves developed the GRADE system—published a pilot study of the system [16]. The study found that the kappa value (i.e., the inter-rater agreement beyond chance) for 12 judgments about the quality of evidence was very low (mean  $\kappa = 0.27$ ;  $\kappa < 0$  for four judgments). The authors stated that “with discussion” they were able to considerably improve their system, but provided no supportive data. Furthermore, the presentation of GRADE that had been published a year earlier in 2004 contains neither assessment of reliability, agreement, nor proof of usefulness [6].

**Citation:** Kavanagh BP (2009) The GRADE System for Rating Clinical Guidelines. *PLoS Med* 6(9): e1000094. doi:10.1371/journal.pmed.1000094

**Published:** September 15, 2009

**Copyright:** © 2009 Brian P. Kavanagh. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This perspective was written without funding contributions from any source.

**Competing Interests:** The author has declared that no competing interests exist.

\* E-mail: brian.kavanagh@sickkids.ca

**Provenance:** Not commissioned; externally peer-reviewed.

The Policy Forum allows health policy makers around the world to discuss challenges and opportunities for improving health care in their societies.

### Box 1. Organizations That Have Adopted the Grade System

Agency for Healthcare Research and Quality (USA)  
 Agenzia Sanitaria Regionale (Italy)  
 American College of Chest Physicians (USA)  
 American College of Physicians (USA)  
 American Thoracic Society (USA)  
 Ärztliches Zentrum für Qualität in der Medizin (Germany)  
 British Medical Journal (United Kingdom)  
 BMJ Clinical Evidence (United Kingdom)  
 COMPUS at The Canadian Agency for Drugs and Technologies in Health (Canada)  
 The Cochrane Collaboration (International)  
 EMB Guidelines (Finland/International)  
 The Endocrine Society (USA)  
 European Respiratory Society (Europe)  
 European Society of Thoracic Surgeons (International)  
 Evidence-based Nursing Südtirol (Italy)  
 German Center for Evidence-based Nursing "sapere aude" (Germany)  
 Infectious Diseases Society of America (USA)  
 Japanese Society for Temporomandibular Joint (Japan)  
 Journal of Infection in Developing Countries (International)  
 Kidney Disease: Improving Global Outcome (International)  
 Ministry of Health and Long-Term Care, Ontario (Canada)  
 National Board of Health and Welfare (Sweden)  
 National Institute for Health and Clinical Excellence (United Kingdom)  
 Norwegian Knowledge Centre for the Health Services (Norway)  
 Polish Institute for EBM (Poland)  
 Society for Critical Care Medicine (USA)  
 Society for Vascular Surgery (USA)  
 Spanish Society for Family and Community Medicine (Spain)  
 Surviving Sepsis Campaign (International)  
 University of Pennsylvania Health System  
 Center for Evidence-Based Practice (USA)  
 UpToDate (USA)  
 World Health Organization (International)

**GRADE experts versus content experts.** Comparing expert opinion on sepsis with the result of the GRADE process further suggests that GRADE lacks internal consistency.

First, glucose control in the critically ill is a complex issue [17]. Recent clinical data suggest no benefit to widespread application of "tight" glucose control (i.e., intensive insulin therapy) in most intensive care unit (ICU) patients [18–21]. Brunkhorst and colleagues state that intensive insulin therapy has "no measurable consistent benefit in critically ill patients in a medical ICU regardless of whether the patients have severe sepsis and that such therapy increases the risk of hypoglycemic episodes" [18]. Yet the senior author of that report [18], Konrad Reinhart, is a coauthor of the SSC guidelines that gave a grade 1 ranking (strong recommendation) for "moderate" glucose control and a grade 2 endorsement (a suggestion) for "tight" glucose control [11]. No evidence exists for moderate glucose control in this context, whereas the value of tight control was supported by one single-centre randomized controlled trial (RCT) [22] and opposed by four others [18,20–21,23]. Since the 2008 SSC forum [11], the largest multicentre study, the NICE-Sugar trial, reported that tight glucose control increased ICU mortality by 2.6% (OR 1.14) [24].

Second, the SSC strongly recommends (i.e., grade 1) specific resuscitation targets (blood pressure, urine output, central venous pressure, central venous oxygenation) [11], on the basis of the protocol of a commonly cited single-centre study [25]. In a different forum, the SSC states: "It is impossible to determine from the study which particular facet of the protocol was beneficial for the patients, so the protocol as a whole must be recommended" [26]. But there is considerable debate about the usefulness of this protocol—two ongoing studies are examining if the protocol is effective [27–28]. One of these studies is led by Derek Angus, an author of the SSC guidelines [11]. Thus, I see an inconsistency in a grading system

where the most authoritative expert in the SSC panel is investigating if the protocol is useful versus the aggregate panel decision concluding a strong recommendation that it should be used [11].

### Is GRADE Inherently Logical?

**Strength of recommendation and quality of evidence.** GRADE provides an expression of the strength of the recommendation and also provides a rating on the quality of the evidence upon which the recommendation is based. In terms of strength, GRADE considers evidence to be "strong" or "weak." The GRADE group considers strength to reflect "the degree of confidence that the desirable effects of adherence to a recommendation outweigh the undesirable effects" [7]. This component makes sense, but less so when the strength of the recommendation is dissociated from its foundation (i.e., the quality of the evidence that underpins the recommendation). The group emphasizes the importance of making this dissociation: "Separating the judgments regarding the quality of evidence from judgments about the strength of recommendations is a critical and defining feature of this new grading system" [7]. One can envision having "high-quality" knowledge that points to a small effect (high quality, low strength). The converse, low quality knowledge that yields a high-strength recommendation seems implausible, other than perhaps the avoidance of substances such as potent toxins.

**Combining incommensurate elements.** Another problem is the "leveling" process proposed to determine the quality of the evidence. GRADE ranks the quality of evidence on the basis of the type of study, "quality" issues (e.g., blinding, follow-up, sparseness of data), consistency, directness (generalizability), and effect size. The graders are instructed to raise or lower the level of quality and trade off, for example, the presence of sparse data against demonstration of a dose-response effect [6]; of course these are fundamentally different and can therefore be neither added nor subtracted.

### GRADE Has Not Been Validated

The basis for the GRADE system is articulated in several publications [6–7,9,16,29–31], but none contains supportive data, proof, or logical argument for the system. Rather, there is extensive reference to other papers written largely by the same group but with no data (except a very low kappa value for inter-observer agreement) [16]. Thus, there is no literature-based proof of the validity of the

### Box 2. Systems for Grading Evidence and Issuing Guidelines Based on the Evidence

Atkins and colleagues compared the following six systems [6]:

- The American College of Chest Physicians Evidence-Based Guidelines (<http://www.chestnet.org/education/hsp/guidelinesProducts.php>)
- Australian National Health and Medical Research Council Guidelines (<http://www.nhmrc.gov.au/guidelines/index.htm>)
- Oxford Centre for Evidence-Based Medicine (<http://www.cebm.net/>)
- Scottish Intercollegiate Guidelines Network (<http://www.sign.ac.uk/>)
- US Preventive Services Task Force Recommendations (<http://www.ahrq.gov/CLINIC/uspstfix.htm#Recommendations>)
- US Task Force on Community Preventive Services Recommendations (<http://www.thecommunityguide.org/about/findings.html>)

### Box 3. Antibiotic Use in Sepsis

In 2004 the SSC guidelines recommended that for serious sepsis, intravenous antibiotic therapy should be rapidly instituted [10]; this guideline was given a grade “E.” The grading system that was in use in 2004 was adopted from Sackett’s 1989 description [3]: in both cases an “E” grade corresponded to a recommendation that was supported by so-called level IV or V evidence (nonrandomized, historical controls, uncontrolled studies, expert opinion)—the lowest levels possible [10].

In 2008, the SSC issued almost the identical recommendation but this time assigned to it a grade of 1B (if shock is present) and 1D (if shock is absent), where grade 1 corresponds to a “strong” [7] recommendation [11]. Three studies published between 2004 and 2008 (none of them randomized controlled trials) supported the idea that early antibiotics reduced mortality in sepsis [45–47], exactly the same conclusions reached by at least six others published before 2004 [48–53]. Although all the studies indicated that antibiotic delay has an adverse effect, they told the clinician nothing that was new: once the need for an antibiotic is confirmed, the sooner it is administered the better. Thus it is unclear why the grading went from grade E in 2004 to grade 1B or 1D in 2008. Was the different grading simply due to the use of a different grading system in these two different years? It seems improbable that two systems describing the validity of a recommendation could arrive at such discordant conclusions. While it is easy to see how the recommendation received a meritorious commendation in 2008 [11], it is difficult to see how it did not in 2004 [10].

GRADE system; indeed using approaches for appraising evidence proposed by the Evidence-Based Medicine Working Group [32], I would conclude that there is little basis for GRADE.

The GRADE documents suggested that strong recommendations should require little debate and would be implemented in most circumstances [7,29]. At first glance, this may seem reasonable but there could be unanticipated consequences, such as stifling debate about many important topics, with the result that there is less thought and less research on that topic. High-level recommendations using other grading systems strongly advocated use of beta-blockade (class I, IIa) [33–34] and aprotinin (class 1a) [4] in specific surgical populations. But assuming that the subsequent RCTs were appropriately conducted [5,35], the original high-level recommendations were clearly misguided [4,33–

34]. A major concern about any grading system is that if enshrined, potentially life-saving prospective studies might not be permitted by research ethics boards on the basis that because a guideline has been assigned a “confident” grading, equipoise does not exist.

#### Popularity and Uptake

The GRADE system has been adopted as is, or with minor modifications, by a large number of professional, statutory, and medically related governance organizations (Box 1). It is hard to understand why so many organizations, many of them leading regulatory or professional groups, would adopt a system that has no proof of effectiveness and has demonstrated inconsistency [16]. There are several possible reasons for its popularity: (1) a perceived need to regulate and reduce “unnecessary” and potentially harmful variation in health

care [36]; (2) GRADE uses attractive language (such as “clarity,” “consistency,” “helpfulness,” and “rigor”) [6,37–38]; (3) the attraction of the promise of clinical excellence being obtainable through such a system; (4) influential bodies may adopt GRADE in order not to be left behind what some view as a “state-of-the-art” scientific advance.

#### GRADE: Potential for Bias

The SSC describes in detail how members of the GRADE group interacted with the sepsis experts and influenced the grading decisions [11]. But it is not clear to me why the GRADE group needed to be involved at all in the grading decisions given that all the SSC members are experts. Given also that the GRADE criteria are conveyed as “explicit and clear” [6], there should be little need for intensive methodological consultation from the GRADE group when experts produce guidelines. While grading experts might be helpful to explain technical elements of grading, the above scenario raises the possibility of the grading process shaping the medical message.

#### GRADE: Implications for Practice and Policy

The GRADE group writes that for clinicians, strong recommendations should be seen as a quality criterion or performance indicator, and for policy makers, be adopted as policy [7]. There are similar efforts underway to synthesize studies and implement practice guidelines in several countries, including the UK and the US [39–41]. But knowing which studies and guidelines are best (or are valid) [42] is not straightforward—high-grade recommendations (such as [4,33–34]) have been later proved wrong [5,35].

It is not clear that the opinion of a conscientious, judicious, well-educated, and experienced clinician would necessarily be inferior to a systemized opinion, such as GRADE, especially if GRADE is not valid. Conferring a “strong” rating upon a guideline will constitute a major deterrent to a clinician considering an alternative clinical route, particularly if GRADE recommendations were to be adopted as a policy by regulatory bodies [7]. Indeed warnings have been issued about proposals to convert guidelines into law [43–44].

#### What Should Replace GRADE?

A key question that arises when a system is questioned is: what is the alternative? There is a very good alternative to using the GRADE system to rate

### Box 4. Ventilation in Acute Respiratory Distress Syndrome

In 2004 the SSC guidelines recommended that levels of positive end-expiratory pressure (PEEP) should be set to prevent lung collapse at expiration [10]. Although most clinicians use PEEP, almost none would be able to quantify lung collapse at the end of expiration, given that atelectasis is seldom quantified. Nonetheless, the grade in 2004 was “E” (i.e., very poor) [10]. In 2008, a virtually identical recommendation received a grade of “1” (i.e., strong) [11]. The results of three randomized controlled trials examining the effect of PEEP in acute respiratory distress syndrome (ARDS) [54–56] were made available before the 2008 SSC conference [11]. But none of the trials analyzed PEEP and collapse in end-expiration; rather they addressed higher versus lower levels of PEEP, and broadly showed that as tested, PEEP made little or no difference to outcome [54–56]. Thus, there is no rationale as to how either grading was arrived at, and no basis for the difference in grading from 2004 to 2008.

clinical guidelines: clinicians and organizations should use published guidelines while considering the clinical context, the credentials, and any conflicts of interest among the authors, as well as the expertise, experience, and education of the

practitioner. If in the future a guideline grading system is shown to improve outcome and is without harm, it could usefully be incorporated into clinical practice.

## Author Contributions

ICMJE criteria for authorship read and met: BPK. Wrote the first draft of the paper: BPK.

## References

- Canadian Task Force on the Periodic Health Examination (1979) The periodic health examination. *Can Med Assoc J* 121: 1193–1254.
- Cook DJ, Guyatt GH, Laupacis A, Sackett DL, Goldberg RJ (1995) Clinical recommendations using levels of evidence for antithrombotic agents. *Chest* 108: 227S–230S.
- Sackett DL (1989) Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 95: 2S–4S.
- Ferraris VA, Ferraris SP, Saha SP, Hessel EA 2nd, Haan CK, et al. (2007) Perioperative blood transfusion and blood conservation in cardiac surgery: the Society of Thoracic Surgeons and The Society of Cardiovascular Anesthesiologists clinical practice guideline. *Ann Thorac Surg* 83: S27–86.
- Fergusson DA, Hebert PC, Mazer CD, Frenes S, MacAdams C, et al. (2008) A comparison of aprotinin and lysine analogues in high-risk cardiac surgery. *N Engl J Med* 358: 2319–2331.
- Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, et al. (2004) Grading quality of evidence and strength of recommendations. *BMJ* 328: 1490.
- Schunemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, et al. (2006) An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 174: 605–614.
- UpToDate. Available: <http://www.uptodate.com/home/index.html>. Accessed 15 April, 2009.
- Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, et al. (2004) Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 4: 38.
- Dellinger RP, Carlet JM, Masur H, Gerlach H, Calandra T, et al. (2004) Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Crit Care Med* 32: 858–873.
- Dellinger RP, Levy MM, Carlet JM, Bion J, Parker MM, et al. (2008) Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Crit Care Med* 36: 296–327.
- Wang HE, Shapiro NI, Angus DC, Yealy DM (2007) National estimates of severe sepsis in United States emergency departments. *Crit Care Med* 35: 1928–1936.
- Clermont G, Angus DC, Kalassian KG, Linde-Zwirble WT, Ramakrishnan N, et al. (2003) Reassessing the value of short-term mortality in sepsis: comparing conventional approaches to modeling. *Crit Care Med* 31: 2627–2633.
- 100,000 Lives Campaign - Lives Saved FAQ. Available: <http://www.ihl.org/NR/rdonlyres/0FC36040-53FB-4B06-A95E-7E2D5055A154/0/LivesSavedCalculationFAQ.doc>. Accessed 15 April 2009.
- Eichacker PQ, Natanson C, Danner RL (2006) Surviving sepsis-practice guidelines, marketing campaigns, and Eli Lilly. *N Engl J Med* 355: 1640–1642.
- Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, et al. (2005) Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res* 5: 25.
- Fowler RA, Annane D (2006) The highs and lows of intensive insulin therapy. *Am J Respir Crit Care Med* 173: 367–369.
- Brunkhorst FM, Engel C, Bloos F, Meier-Hellmann A, Ragaller M, et al. (2008) Intensive insulin therapy and pentastarch resuscitation in severe sepsis. *N Engl J Med* 358: 125–139.
- Egi M, Bellomo R, Stachowski E, French CJ, Hart G, et al. (2006) Intensive insulin therapy in postoperative intensive care unit patients: a decision analysis. *Am J Respir Crit Care Med* 173: 407–413.
- Gandhi GY, Nuttall GA, Abel MD, Mullany CJ, Schaff HV, et al. (2007) Intensive intraoperative insulin therapy versus conventional glucose management during cardiac surgery: a randomized trial. *Ann Intern Med* 146: 233–243.
- Van den Berghe G, Wilmer A, Hermans G, Meersseman W, Wouters PJ, et al. (2006) Intensive insulin therapy in the medical ICU. *N Engl J Med* 354: 449–461.
- van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, et al. (2001) Intensive insulin therapy in the critically ill patients. *N Engl J Med* 345: 1359–1367.
- Preiser JC, Devos P, Ruiz-Santana S, Melot C, Annane D, et al. (2009) A prospective randomised multi-centre controlled trial on tight glucose control by intensive insulin therapy in adult intensive care units: the Glucontrol study. *Intensive Care Med*:E-pub ahead of print 28 July 2009. doi:10.1007/s00134-009-1585-2.
- Finfer S, Chittock DR, Su SY, Blair D, Foster D, et al. (2009) Intensive versus conventional glucose control in critically ill patients. *N Engl J Med* 360: 1283–1297.
- Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, et al. (2001) Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 345: 1368–1377.
- SSC Maintain adequate central venous oxygen saturation. Available: [http://ssc.sccm.org/bundles/individual\\_changes/maintaincentralsvo2](http://ssc.sccm.org/bundles/individual_changes/maintaincentralsvo2). Accessed 15 April 2009.
- Australian and New Zealand Intensive Care Society (2008) The Australasian Resuscitation in Sepsis Evaluation (ARISE) Observational Study. Available: <http://www.anzics.com.au/ctg/article.asp?ID=1>. Accessed 15 April 2009.
- National Institute of General Medical Sciences (2006) New study aims to stop sepsis in its tracks. Available: <http://www.nigms.nih.gov/News/Results/10022006e.htm>. Accessed 15 April 2009.
- Guyatt G, Gutterman D, Baumann MH, Adrizzio-Harris D, Hylek EM, et al. (2006) Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American college of chest physicians task force. *Chest* 129: 174–181.
- Guyatt G, Vist G, Falck-Ytter Y, Kunz R, Magrini N, et al. (2006) An emerging consensus on grading recommendations? *ACP J Club* 144: A8–9.
- Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, et al. (2008) What is “quality of evidence” and why is it important to clinicians? *Bmj* 336: 995–998.
- Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, et al. (1995) Users’ guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA* 274: 1800–1804.
- Eagle KA, Berger PB, Calkins H, Chaitman BR, Ewy GA, et al. (2002) ACC/AHA guideline update for perioperative cardiovascular evaluation for noncardiac surgery—executive summary a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1996 Guidelines on Perioperative Cardiovascular Evaluation for Noncardiac Surgery). *Circulation* 105: 1257–1267.
- Fleisher LA, Beckman JA, Brown KA, Calkins H, Chaikof E, et al. (2007) ACC/AHA 2007 guidelines on perioperative cardiovascular evaluation and care for noncardiac surgery: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines on Perioperative Cardiovascular Evaluation for Noncardiac Surgery): developed in collaboration with the American Society of Echocardiography, American Society of Nuclear Cardiology, Heart Rhythm Society, Society of Cardiovascular Anesthesiologists, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, and Society for Vascular Surgery. *Circulation* 116: e418–499.
- Devereaux PJ, Yang H, Yusuf S, Guyatt G, Leslie K, et al. (2008) Effects of extended-release metoprolol succinate in patients undergoing noncardiac surgery (POISE trial): a randomised controlled trial. *Lancet* 371: 1839–1847.
- Morris AH (2004) Iatrogenic illness: a call for decision support tools to reduce unnecessary variation. *Qual Saf Health Care* 13: 80–81.
- Schunemann HJ, Best D, Vist G, Oxman AD (2003) Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *CMAJ* 169: 677–680.
- Swiglo BA, Murad MH, Schunemann HJ, Kunz R, Vigersky RA, et al. (2008) A case for clarity, consistency, and helpfulness: state-of-the-art clinical practice guidelines in endocrinology using the grading of recommendations, assessment, development, and evaluation system. *J Clin Endocrinol Metab* 93: 666–673.
- National Guideline Clearinghouse. Available: <http://www.guidelines.gov/>. Accessed 15 April 2009.
- National institute for Health and Clinical Excellence. Available: <http://www.nice.org.uk>. Accessed 15 April 2009.
- Wadman M (2008) New plan proposed to help resolve conflicting medical advice. *Nat Med* 14: 226.
- Getting with the program. *Nat Med* 14: 223.
- Fein IA, Corrado RR (2008) Clinical practice guidelines: culture eats strategy for breakfast, lunch, and dinner. *Crit Care Med* 36: 1360–1361.
- Jacobson PD (2008) Transforming clinical practice guidelines into legislative mandates: proceed with abundant caution. *Jama* 299: 208–210.
- Houck PM, Bratzler DW, Nsa W, Ma A, Bartlett JG (2004) Timing of antibiotic administration and outcomes for Medicare patients hospitalized with community-acquired pneumonia. *Arch Intern Med* 164: 637–644.
- Kumar A, Haery C, Paladugu B, Kumara D, Symeonides S, et al. (2006) The duration of hypotension before the initiation of antibiotic treatment is a critical determinant of survival in a murine model of *Escherichia coli* septic shock:

- association with serum lactate and inflammatory cytokine levels. *J Infect Dis* 193: 251–258.
47. Proulx N, Frechette D, Toye B, Chan J, Kravcik S (2005) Delays in the administration of antibiotics are associated with mortality from adult acute bacterial meningitis. *QJM* 98: 291–298.
  48. Gacouin A, Le Tulzo Y, Lavoue S, Camus C, Hoff J, et al. (2002) Severe pneumonia due to *Legionella pneumophila*: prognostic factors, impact of delayed appropriate antimicrobial therapy. *Intensive Care Med* 28: 686–691.
  49. Kang CI, Kim SH, Kim HB, Park SW, Choe YJ, et al. (2003) *Pseudomonas aeruginosa* bacteremia: risk factors for mortality and influence of delayed receipt of effective antimicrobial therapy on clinical outcome. *Clin Infect Dis* 37: 745–751.
  50. Larche J, Azoulay E, Fieux F, Mesnard L, Moreau D, et al. (2003) Improved survival of critically ill cancer patients with septic shock. *Intensive Care Med* 29: 1688–1695.
  51. Lodise TP, McKinnon PS, Swiderski L, Rybak MJ (2003) Outcomes analysis of delayed antibiotic treatment for hospital-acquired *Staphylococcus aureus* bacteremia. *Clin Infect Dis* 36: 1418–1423.
  52. Meehan TP, Fine MJ, Krumholz HM, Scinto JD, Galusha DH, et al. (1997) Quality of care, process, and outcomes in elderly patients with pneumonia. *Jama* 278: 2080–2084.
  53. Miner JR, Heegaard W, Mapes A, Biros M (2001) Presentation, time to antibiotics, and mortality of patients with bacterial meningitis at an urban county medical center. *J Emerg Med* 21: 387–392.
  54. Brower RG, Lanken PN, MacIntyre N, Matthay MA, Morris A, et al. (2004) Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. *N Engl J Med* 351: 327–336.
  55. Meade MO, Cook DJ, Guyatt GH, Slutsky AS, Arabi YM, et al. (2008) Ventilation strategy using low tidal volumes, recruitment maneuvers, and high positive end-expiratory pressure for acute lung injury and acute respiratory distress syndrome: a randomized controlled trial. *JAMA* 299: 637–645.
  56. Mercat A, Richard JC, Vielle B, Jaber S, Osman D, et al. (2008) Positive end-expiratory pressure setting in adults with acute lung injury and acute respiratory distress syndrome: a randomized controlled trial. *JAMA* 299: 646–655.