

Analysis of Overall Level of Evidence Behind Infectious Diseases Society of America Practice Guidelines

Dong Heun Lee, MD; Ole Vielemeyer, MD

Background: Clinical practice guidelines are developed to assist in patient care. Physicians may assume that following such guidelines means practicing evidence-based medicine. However, the quality of supporting literature can vary greatly.

Methods: We analyzed the strength of recommendation and overall quality of evidence behind 41 Infectious Diseases Society of America (IDSA) guidelines released between January 1994 and May 2010. Individual recommendations were classified based on their strength of recommendation (levels A through C) and quality of evidence (levels I through III). Guidelines not following this format were excluded from further analysis. Evolution of IDSA guidelines was assessed by comparing 5 recently updated guidelines with their earlier versions.

Results: In the 41 analyzed guidelines, 4218 individual recommendations were found and tabulated. Fourteen

percent of the recommendations were classified as level I, 31% as level II, and 55% as level III evidence. Among class A recommendations (good evidence for support), 23% were level I (≥ 1 randomized controlled trial) and 37% were based on expert opinion only (level III). Updated guidelines expanded the absolute number of individual recommendations substantially. However, few were due to a sizable increase in level I evidence; most additional recommendations had level II and III evidence.

Conclusions: More than half of the current recommendations of the IDSA are based on level III evidence only. Until more data from well-designed controlled clinical trials become available, physicians should remain cautious when using current guidelines as the sole source guiding patient care decisions.

Arch Intern Med. 2011;171(1):18-22

FOR CENTURIES, MEDICINE WAS taught in a purely authoritarian manner and was practiced by following expert advice. Testing of treatment modalities for their efficacy was reported as early as the 11th century (see *The Canon of Medicine* by Ibn Sīnā¹); nevertheless, interventions of clear benefit to patients were rarely endorsed. For example, despite Semmelweis' observation in 1847 that hand washing could abort outbreaks of puerperal fever, most physicians at the time did not adhere to the practice.² Only through the pioneering efforts of A. Cochrane and a research group at McMaster's University led by David Sackett, MD, FRCP, and Gordon. Guyatt, MD, FRCP, in the 1950s has medicine since embraced evidence-based practices.

clinical decision making and encumbered it owing to the tremendous volume of new information. Clinical practice guidelines were developed to aid clinicians in improving patient outcomes and streamlining health care delivery by analyzing and summarizing data from all relevant publications.³⁻⁵ Lately, these guidelines have also been used as tools for educational purposes, performance measures, and policy making.⁶

Interest has been growing in critically appraising not only individual clinical practice guidelines but also entire guideline sets of different medical (sub)specialties.⁷⁻¹⁰ We assessed the overall quality of evidence underlying recommendations outlined in existing guidelines from the Infectious Diseases Society of America (IDSA).

*For editorial comment
see page 15*

During the past half century, a deluge of publications addressing nearly every aspect of patient care has both enhanced

Author Affiliations: Division of Infectious Diseases and HIV Medicine, Department of Medicine, Drexel University College of Medicine, Philadelphia, Pennsylvania.

METHODS

The IDSA guidelines use the IDSA-US Public Health Service grading system (henceforth referred to as the IDSA evidence-grading system).¹¹ In this system, each recommendation is graded according to its strength and the

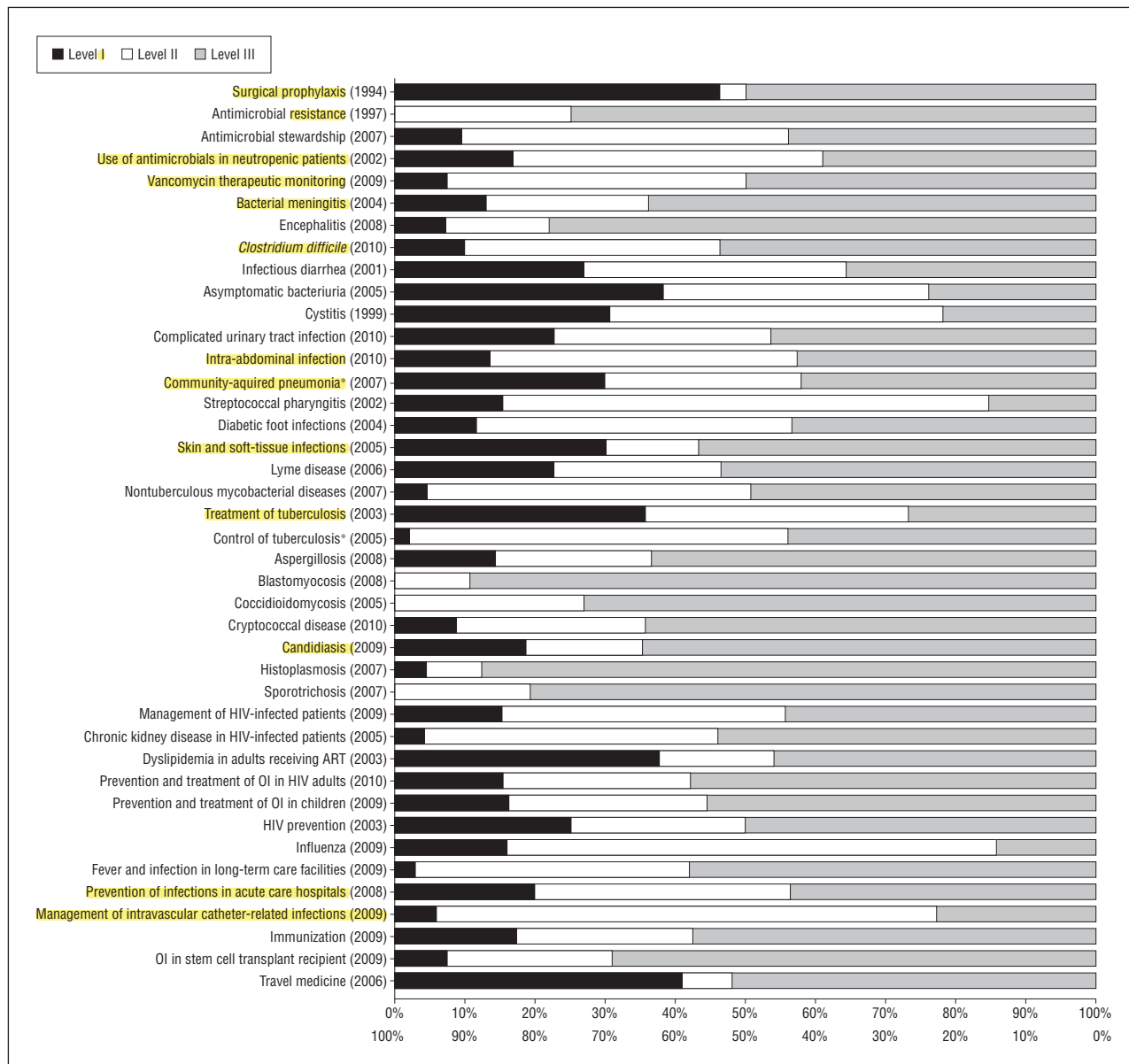


Figure 1. Comparison of 41 guidelines using percentage distribution of quality of evidence underlying individual recommendations. ART indicates antiretroviral therapy; HIV, human immunodeficiency virus; and OI, opportunistic infection. *Used a grading system that constituted a modification of the standard Infectious Diseases Society of America evidence-grading system.

underlying quality of evidence. Strength of recommendation includes levels A through C (ie, A indicates good evidence to support recommendation for use; B, moderate evidence to support recommendation; and C, poor evidence to support recommendation); some guidelines also included levels D (moderate evidence to support recommendation against use) and E (good evidence to support recommendation against use). These guidelines were mostly released before 2008. Quality of evidence ranges from level I through III (I indicates evidence from ≥ 1 properly randomized controlled trial; II, evidence from ≥ 1 well-designed clinical trial, without randomization, from cohort or case-controlled analytical studies or from dramatic results from uncontrolled experiments; and III, evidence from opinions of respected authorities based on clinical experience, descriptive studies, or reports of expert committees). For ease of analysis, we merged all recommendations labeled D and E (found mainly in pre-2008 guidelines) into categories B and A, respectively. We read and tabulated all current guidelines

posted on the IDSA Web site (<http://www.idsociety.org>) as of May 2010. Designations of strength of recommendation and quality of evidence linked to each recommendation were extracted. Special attention was given to avoiding duplication in the counting of recommendations listed more than once (ie, listed in the summary section or in table format in addition to the body of a given guideline). Guidelines cosponsored by other societies were included in the analysis if the IDSA evidence-grading system was used with only minor modifications. Publications not adhering to the IDSA evidence-grading system were excluded from subsequent analysis to maintain objectivity of data compilation. For comparison of the overall quality of evidence between individual guidelines, we used relative proportions rather than absolute numbers because the latter varied greatly between guidelines.

To study the evolution of IDSA guidelines across time, we compared 5 recently published guidelines, updated between January 2008 and May 2010, with their respective older ver-

Table. Distribution of Individual Recommendations From Current Infectious Diseases Society of America Guidelines According to Strength of Recommendation and Quality of Evidence^a

Strength of Recommendation ^b	Quality of Evidence			Total
	Level I	Level II	Level III	
Level A	414 (23)	715 (40)	667 (37)	1796 (100)
Level B	143 (8)	544 (30)	1132 (62)	1819 (100)
Level C	24 (4)	48 (8)	531 (88)	603 (100)
Total	581 (14)	1307 (31)	2330 (55)	4218 (100)

^aData are given as number (percentage).

^bA indicates good evidence to support recommendation for use; B, moderate evidence to support recommendation; and C, poor evidence to support recommendation; I, evidence from 1 or more properly randomized controlled trial; II, evidence from 1 or more well-designed clinical trial, without randomization, from cohort or case-controlled analytical studies or from dramatic results from uncontrolled experiments; and III, evidence from opinions of respected authorities based on clinical experience, descriptive studies, or reports of expert committees.

sions. We compiled data on the percentage of new references (ie, citations with a publication date after release of the earlier guidelines) and on the total number of individual recommendations and the number of level I, II, and III quality-of-evidence designations for each of the 5 guideline pairs.

RESULTS

Between January 1994 and May 2010, the IDSA released 90 guidelines covering a wide range of topics. As of May 2010, fifty-two current guidelines were listed on the IDSA Web site (<http://www.idsociety.org>). Of these guidelines, 41 (79%) used the IDSA evidence-grading system for individual recommendations,¹¹ some with minor modifications (2 guidelines), and could, therefore, be analyzed in more detail (**Figure 1**).

Twenty-one of the 41 guidelines (51%) cover a new topic and 20 (49%) are updates of earlier publications. Two guidelines had been updated twice. The mean time between original and updated versions was 6.7 years (range, 1-15 years). The total number of individual recommendations per guideline ranged from 4 to 864, with a median of 48. A mean of 13 authors (range, 4-66) contributed to each guideline.

We identified 4218 individual recommendations, which we charted according to strength of recommendation and quality of evidence (**Table**). Forty-three percent of recommendations (n=1796) were designated as strength A, 43% (n=1819) as strength B, and 14% (n=603) as strength C recommendations. Among all level A strength recommendations ("good evidence to support a recommendation for use") less than one-quarter were supported by level I quality of evidence (ie, designated as A-I recommendation). A global look at the overall distribution of the quality of evidence underlying all 4218 recommendations showed that only 14% were linked to level I, whereas more than half were supported by level III quality of evidence only (Table).

We then compared the distribution of recommendation grading according to quality of evidence between individual guidelines (Figure 1). Again, level I quality of

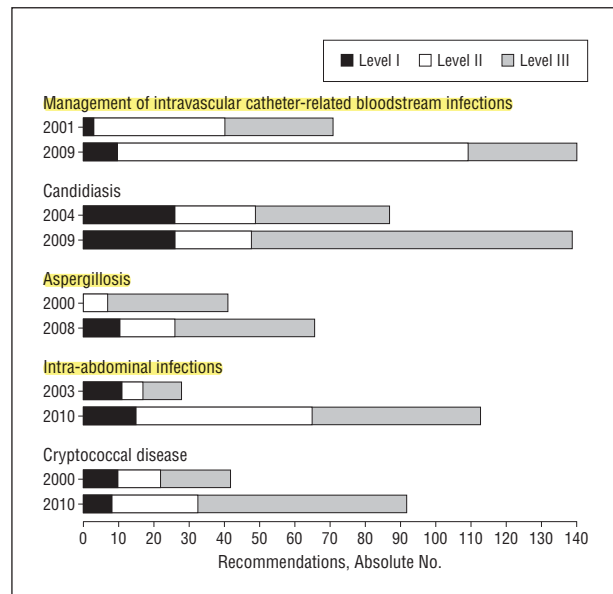


Figure 2. Comparison of 5 recently updated guidelines with their respective previous versions. The total number of individual recommendations found is graphically depicted for each guideline pair and according to the quality of underlying evidence. See the Table footnote for an explanation of levels.

evidence was seen in only 1 of 6 recommendations per guideline (median, 15%; interquartile range, 6%-24%), whereas half of the recommendations (50%; 43%-64%) were supported by level III quality of evidence only. Guidelines on surgical prophylaxis (published in 1994) had the highest percentage of level I recommendations (46%), followed by guidelines on travel medicine (41%) (published in 2006) and asymptomatic bacteriuria (38%) (published in 2005). In contrast, for blastomycosis (published in 2008) and sporotrichosis (published in 2007), more than 80% of all recommendations were based on level III evidence only, and level I support was lacking entirely.

Finally, we looked at the extent to which the quality of evidence supporting IDSA guidelines improved across time. We selected 5 guidelines that had recently been updated and compared these with their respective earlier versions. In all but 1 guideline pair, the total number of cited articles increased in the newer guidelines, in 1 case 5-fold. On average, new publications constituted 53% (range, 34%-65%) of all referenced citations in the updated guidelines. For each guideline pair, the total number of recommendations increased with the update, ranging from 20% to 400% (**Figure 2**). However, only 2 updated guidelines had a significant increase in the number of level I quality-of-evidence recommendations; most additional recommendations were supported by level II or III quality of evidence only.

COMMENT

In 1990, the Institute of Medicine proposed the development of guidelines to **reduce** inappropriate **variation** in the provision of health care by assisting patient and practitioner decision making.¹² Since then, for the field of **infectious diseases** alone, **90** guidelines summarizing

a large body of published data have been sponsored by the IDSA, and as of May 2010, fifty-two current guidelines can be found on their Web site.

In daily clinical work, practitioners sometimes assume that adhering to practice guidelines means practicing evidence-based medicine. However, individual recommendations within published guidelines may not be supported by high-quality evidence. We examined all 41 current IDSA guidelines that followed the IDSA evidence-grading system.¹¹ Of the 4218 individual recommendations found, only 14% were supported by the strongest (level I) quality of evidence; more than half were based on level III evidence only. Although on average updated guidelines contained newly published studies to 50% in their reference lists and although updates contained substantially more individual recommendations, only 2 of 5 new guidelines had a significant increase in level I recommendations.

There are several possible reasons to explain these findings. Recently, Tricoci et al¹⁰ assessed the scientific evidence underlying the clinical practice guidelines in cardiology. They mainly analyzed how the American College of Cardiology and the American Heart Association guidelines have evolved across time and found results that are similar to the present data. In their study, only 11% of recommendations were supported by the highest level of evidence, whereas 48% were based on expert opinion, case studies, or standards of care only. Tricoci et al speculate that “guidelines in other medical areas in which large clinical trials are performed less frequently, may have an even weaker evidence-based foundation.”^{10(p835)} Indeed, in contrast to other subspecialties of internal medicine, in the field of infectious diseases, relatively few large multicenter randomized controlled trials (RCTs) have been conducted, with the notable exception of antiretroviral therapy trials in human immunodeficiency virus (HIV) care. Many infectious diseases occur infrequently, present in a heterogeneous manner, or are difficult to diagnose with certainty. For others, an RCT would be impractical or wasteful or might be deemed unethical. Such examples might include the study of the usefulness of tick bite avoidance through physical distance for the prevention of Lyme disease, the utility of hand hygiene to reduce nosocomial infections, or the use of cesarean delivery over vaginal delivery to reduce the risk of vertical HIV transmission. Also, RCTs are costly, and the lack of resources and funding in the field clearly poses an obstacle to obtaining a level I recommendation for many management decisions. Travel medicine guidelines had one of the highest percentage points of recommendations supported by level I quality of evidence. This could be because vaccines and medications for prophylaxis and treatment of travel-related illnesses require RCTs before Food and Drug Administration approval and that there are global efforts to combat many of the diseases highly prevalent in the tropics. Conversely, guidelines for the management of endemic fungal infections lacked level I recommendations entirely. The infrequent occurrence of these diseases and the difficulty of an accurate diagnosis make it difficult to conduct RCTs.

A second reason for the scarcity of level I quality-of-evidence recommendations may be the use of the IDSA evidence-grading system.^{11,13} This system was originally

proposed to evaluate the effectiveness of preventive health care interventions in Canada.¹⁴ It requires at least 1 supporting RCT for a level I recommendation. Many IDSA recommendations, however, address questions about diagnosis or prognosis (neither of which can be studied using an RCT and, thus, could never receive the highest-level recommendation). Other recommendations endorse obvious interventions, such as hand hygiene, for which no RCT will ever be conducted. Finally, not all RCTs leading to a level I designation are of the same quality. Some may have used surrogate markers as an outcome measure, some may have had small sample sizes, and others may have been poorly conducted. Well-designed nonrandomized studies, on the other hand, may yield solid information but nonetheless cannot lead to a level I recommendation using the current evaluation system. The GRADE (Grading of Recommendations Assessment, Development, and Evaluation) system,¹⁵⁻¹⁹ first proposed in 2004, offers a potential solution. It has 4 quality-of-evidence categories (high, moderate, low, and very low) that are based solely on the likelihood of further research being able to challenge the confidence in the estimate of effect. The label “high” could, thus, be allocated in the absence of an RCT. The grade system was recently applied for grading recommendations in the HIV management guidelines released by the World Health Organization in November 2009.²⁰ We believe that standardization and use of a single grading system across all fields of medicine would make it easier for the busy clinician to interpret individual recommendations. For the current 52 guidelines endorsed by the IDSA, 6 different grading systems were used.

Ideally, clinical guidelines should state clearly and concisely all important decision options and outcomes; include information about diagnostic tests, prognosis, treatment, harm, and economic analyses whenever applicable; identify, validate, and correctly display all relevant underlying evidence; and be resistant to clinically sensible variations in practice.²¹⁻²³ Publishers of guidelines themselves, however, do not universally follow guidelines on how to publish guidelines.^{9,24} Although quality analysis was not part of this study, we came across imprecisions on more than 1 occasion and for more than 1 guideline, including illogical, erroneous, or missing references for recommendations and their associated grades, including A-I recommendations. Inaccurate or wrong citations can be found relatively frequently in the medical literature, as has been reported previously.^{25,26} The creation of practice guidelines is labor and time intensive, and the final product becomes a frequently used reference guide. To improve the overall quality and, thus, usefulness of practice guidelines, we ponder how, in addition to having a more standardized grading system, a more accurate proofreading mechanism especially for cited articles underlying the recommendations' grading could be implemented.

This study has several limitations. First, we did not include all current IDSA guidelines in the analysis. Excluded guidelines, however, did not use the IDSA evidence-grading system; hence, their analysis would have introduced subjectivity into the analysis. We believe that their inclusion would not have changed the conclusions. Second, the analysis was purely statistical and de-

scriptive, and the primary cited literature was not evaluated. Third, although several national and international organizations have, likewise, published guidelines pertinent to the field of infectious diseases, we limited this study to guidelines put forth by the IDSA, considered the preeminent organization in the field worldwide.

In the era of managed care, guidelines are used increasingly not only for decision making in clinical practice but also as **benchmarks** in the **appraisal of quality** of health care provision. A thorough understanding of their **inherent limitations** is, therefore, critical. This analysis of IDSA guidelines showed a **relative paucity** of recommendations supported by **level I** quality of evidence and a rate of **more than 50%** of all recommendations supported by **low-level** evidence only. What are the implications of these findings for busy clinicians who manage patients with infectious diseases? We believe that the current clinical practice guidelines released by the IDSA constitute a great and reliable source of information that should be used. However, in circumstances when patient outcome is less than desirable, or when colleagues use diagnostic or therapeutic choices not included in the recommendations, it is prudent to remember that many of the individual recommendations are not supported by solid evidence. In such cases, **we encourage reviewing the primary literature and using one's clinical judgment rather than relying solely on recommendations.**

Guidelines can only summarize the best available evidence, which often may be weak. Thus, even more than 50 years since the inception of evidence-based medicine, following guidelines cannot always be equated with practicing medicine that is founded on robust data. To improve patient outcomes and minimize harm, future research efforts should focus on areas where only low-level quality of evidence is available. Until more data from such research in the form of well-designed and controlled clinical trials emerge, physicians and policy makers should remain cautious when using current guidelines as the sole source guiding decisions in patient care.

Accepted for Publication: June 14, 2010.

Correspondence: Ole Vielmeyer, MD, Division of Infectious Diseases and HIV Medicine, Department of Medicine, Drexel University College of Medicine, 245 N 15th St, NCB 6306, Mailstop 461, Philadelphia, PA 19102 (ole.vielmeyer@drexelmed.edu).

Author Contributions: *Study concept and design:* Lee and Vielmeyer. *Acquisition of data:* Lee and Vielmeyer. *Analysis and interpretation of data:* Lee and Vielmeyer. *Drafting of the manuscript:* Lee and Vielmeyer. *Critical revision of the manuscript for important intellectual content:* Vielmeyer. *Statistical analysis:* Lee. *Study supervision:* Vielmeyer.

Financial Disclosure: None reported.

Previous Presentation: This study was presented in part at the 47th IDSA Annual Meeting; October 31, 2009; Philadelphia, Pennsylvania.

Additional Contributions: Diana Winters, BA, Drexel University College of Medicine Academic Publishing Ser-

VICES, provided invaluable input during the editing process of the manuscript.

REFERENCES

1. Sinā I. *The Cannon of Medicine*. Cited by: Smith RD. Avicenna and the *Cannon of Medicine*: a millennial tribute. *West J Med*. 1980;133(4):367-370.
2. The first report on Semmelweis' discovery of the cause and prevention of child-bed fever. Most important experiences about the etiology of puerperal fevers epidemic in birthing institutions. *Infect Control*. 1982;3(6):478-479.
3. Chassin MR, Koseoff J, Park RE, et al. Does inappropriate use explain geographic variations in the use of health care services? a study of three procedures. *JAMA*. 1987;258(18):2533-2537.
4. Field MJ, Lohr KN, eds. *Guidelines for Clinical Practice: From Development to Use*. Washington, DC: National Academy Press; 1992.
5. Woolf SH. Practice guidelines: a new reality in medicine, I. recent developments. *Arch Intern Med*. 1990;150(9):1811-1818.
6. Berg AO, Atkins D, Tierney W. Clinical practice guidelines in practice and education. *J Gen Intern Med*. 1997;12(Suppl 2):S25-S33.
7. Browman GP. Evidence-based clinical practice guideline development: principles, challenges, and accountability to evidence. *J Surg Oncol*. 2010;101(1):1-2.
8. Gupta S, Bhattacharyya OK, Brouwers MC, et al. Canadian Thoracic Society: Presenting a new process for clinical practice guideline production. *Can Respir J*. 2009;16(6):e62-e68.
9. Shaneyfelt TM, Centor RM. Reassessment of clinical practice guidelines: go gently into that good night. *JAMA*. 2009;301(8):868-869.
10. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA*. 2009;301(8):831-841.
11. Kish MA; Infectious Diseases Society of America. Guide to development of practice guidelines. *Clin Infect Dis*. 2001;32(6):851-854.
12. Field MJ, Lohr KN, eds. *Clinical Practice Guidelines: Directions for a New Program*. Washington, DC: National Academy Press; 1990.
13. Gross PA, Barrett TL, Dellinger EP, et al; Infectious Diseases Society of America. Purpose of quality standards for infectious diseases. *Clin Infect Dis*. 1994; 18(3):421.
14. Canadian Task Force on the Periodic Health Examination. The periodic health examination. *Can Med Assoc J*. 1979;121(9):1193-1254.
15. GRADE Working Group Web site. <http://www.gradeworkinggroup.org/index.htm>. Accessed February 26, 2010.
16. Atkins D, Eccles M, Flottorp S, et al; GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches. *BMC Health Serv Res*. 2004;4(1):38. doi:10.1186/1472-6963-4-38.
17. Guyatt GH, Oxman AD, Vist GE, et al; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926.
18. Jaeschke R, Guyatt GH, Dellinger P, et al; GRADE Working Group. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. *BMJ*. 2008;337:a744. doi:10.1136/bmj.a744.
19. Atkins D, Best D, Briss PA, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490. doi:10.1136/bmj.328.7454.1490.
20. World Health Organization Rapid advice: antiretroviral therapy for HIV infection in adults and adolescents. <http://www.who.int/hiv/pub/arv/advice/en/index.html>. 2009. Accessed October 16, 2010.
21. Gross PA. Practice guidelines for infectious diseases: rationale for a work in progress. *Clin Infect Dis*. 1998;26(5):1037-1041.
22. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt G; Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature, VIII. how to use clinical practice guidelines: A. are the recommendations valid? *JAMA*. 1995; 274(7):570-574.
23. Wilson MC, Hayward RS, Tunis SR, Bass EB, Guyatt G; Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature, VIII. how to use clinical practice guidelines: B. what are the recommendations and will they help you in caring for your patients? *JAMA*. 1995;274(20):1630-1632.
24. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? the methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA*. 1999;281(20):1900-1905.
25. de Lacey G, Record C, Wade J. How accurate are quotations and references in medical journals? *Br Med J (Clin Res Ed)*. 1985;291(6499):884-886.
26. Siebers R, Holt S. Accuracy of references in five leading medical journals. *Lancet*. 2000;356(9239):1445.