The Medical Literature

Users' Guides to the Medical Literature

XI. How to Use an Article About a Clinical Utilization Review

C. David Naylor, MD, DPhil; Gordon H. Guyatt, MD, MSc; for the Evidence-Based Medicine Working Group

CLINICAL SCENARIO

You are a general internist attending a medical advisory committee meeting as the newly appointed chief of staff in a large community hospital affiliated with a major health maintenance organization. A junior administrator presents data showing that the hospital's utilization of percutaneous transluminal coronary angioplasty (PTCA) is high relative to similar-sized centers with similar numbers of interventional cardiologists. He insinuates that unnecessary PTCAs are being done. The cardiologists present are infuriated, and the meeting degenerates into a shouting match. After the hospital chief executive officer brings the meeting back to order, you and the chief of cardiology agree to research the matter independently and report back in 1 week.

THE SEARCH

Raw utilization data are insufficient to assess whether cardiologists at your hospital are using PTCA inappropriately. You need to review their practice in light of criteria for deciding whether each application of PTCA was likely, given a balance of risks and benefits, to be in the patient's best interest. Using MEDLINE on CD-ROM, you search from January 1991 to November 1995. The medical subject heading (MeSH), "angioplasty, transluminal, percutaneous coronary" yields 2052 citations even after the search is limited to "human" and "English language" with an abstract on file. You then try "guideline" or "practice guideline" as key words. The relevant guideline references look useful for informing a practitioner's decisions, but you cannot readily see how to translate them into criteria for auditing individual charts.

Finally you combine PTCA with "utilization review" as a MeSH heading, and 2 references turn up. The abstract of 1 article looks directly relevant. Carried out by researchers at RAND, the study used explicit criteria to assess the appropriateness of PTCA for 1306 randomly selected patients in 15 randomly selected New York State hospitals.[1] A retrospective medical record audit was performed—similar to what you envisage may be necessary for your hospital. However, you also note that the records were drawn from 1990, raising a concern that the criteria may be outdated. Fifty-eight percent of PTCAs were rated appropriate; 38%, uncertain; and 4%, inappropriate. The inappropriate rate varied by hospital from 1% to 9% (P=.12), while the uncertain rate ranged from 26% to 50% (P=.02). Judging from this article, your hospital would have a defensible profile if its rate of apparently inappropriate PTCA were under 10%. But are the criteria developed by the RAND investigators valid or easily applied?

INTRODUCTION

Evidence on a particular clinical topic is often usefully compiled in published meta-analyses, decision analyses, or practice guidelines. These integrative reports synthesize multiple research studies to help define what a practitioner ought to do when confronted with a particular clinical situation. However, actual practice sometimes differs from what the evidence suggests ought to be done, raising concerns about quality of care. Quality concerns, together with the omnipresent focus on cost containment, have led a growing cadre of researchers, insurers, administrators, and policymakers to examine what clinicians do. Their examinations may focus on outcomes, but as the previous Users' Guide showed,[2] it is not easy to determine whether an adverse outcome was due to some aspect of the care provided or attributable to the patient's clinical situation. Indeed, even exemplary care may be associated with bad outcomes if the patient's prognosis is inherently poor. Thus, it is often more straightforward and valid to assess processes of care—the topic of this article.

In assessing clinical processes, researchers and managers seek to determine whether the right service is provided to the right type of patient for the right reasons at the right time and place. This can be done by implicit reviews, relying on the individualized judgments of expert clinicians. Practitioners then have the comfort of knowing that their work is being appraised by someone who understands the clinical world and its exigencies. Unfortunately, lack of stan-

From the Institute for Clinical Evaluative Sciences in Ontario, University of Toronto (Ontario) Clinical Epidemiology and Health Care Research Program (Sunnybrook Unit), and the Departments of Medicine and Surgery, University of Toronto (Dr Naylor); and McMaster University Health Sciences Centre, Hamilton, Ontario (Dr Guyatt).

The original list of members (with affiliations) of the Evidence-Based Medicine Working Group appears in the first article of this series (*JAMA*. 1993;270:2093-2095). A list of new members appears at the end of this article. The following members contributed to this article: Eric Bass, MD, MPH; Hertzel Gerstein, MD, MSc; Daren Heyland, MD, MSc; Ann Holbrook, MD, PharmD, MSc; Virginia Moyer, MD, MPH; Thomas B. Newman, MD; Andrew Oxman, MD, MSc; W. Scott Richardson, MD; Peter Tugwell, MD, MSc; and John Williams, Jr, MD, MHS.

Corresponding author: C. David Naylor, MD, DPhil, FRCPC, Institute for Clinical Evaluative Sciences in Ontario, G-106, Sunnybrook Health Science Centre, 2075 Bayview Ave, North York, Ontario, Canada M4N 3M5.

Reprints: Gordon Guyatt, MD, MSc, McMaster University Health Sciences Centre, 1200 Main St W, Room 2C12, Hamilton, Ontario, Canada L8N 3Z5.

Users' Guides to the Medical Literature section editor: Drummond Rennie, MD, Deputy Editor (West), *JAMA*.

dardization renders implicit reviews unreliable.[3,4] Explicit criteria, which form the basis for most process-of-care analyses in the literature, have the advantages of standardization and consistency, as well as transparency. Where necessary, trained staff can apply them retrospectively to medical records without a major time commitment from clinicians. Such criteria may nonetheless have a weak basis in evidence, or be applied in a biased or imprecise fashion, or be impractical for use in your particular practice setting. This Users' Guide will accordingly assist you in either of 2 related goals: to critique an article purporting to measure the quality of the process of care delivered in a particular setting, and to decide whether, in conducting your own utilization review, you should emulate the methods or borrow the tools used in a published study.

In the following discussion, we shall use the American term "utilization review" and the British term "clinical audit" interchangeably to describe this type of process-of-care assessment. We shall refer to "panelists" as members of the group of clinical experts that helps establish the explicit review criteria and "auditors" as those who review patient charts or interview patients and/or physicians to obtain the clinical information needed to apply the criteria.

GUIDES FOR REVIEWING A CLINICAL AUDIT

We have modified the basic questions used in earlier Users' Guides to consider 3 issues: Are the criteria valid? Were the criteria applied appropriately? Can you use the criteria in your own practice setting (Table)?

Are the Criteria Valid?

For process-of-care criteria to be valid, they must have a direct link either to improving health or to lowering resource use without compromising health outcomes. These criteria constitute guidelines for others to use in assessing whether a practitioner made the right decision, as opposed to guidelines aimed at helping a practitioner actually make clinical decisions. Despite this different focus, the questions for appraising the validity of criteria for a utilization review are similar to those presented earlier for practice guidelines.[5,6]

Was an Explicit and Sensible Process Used to Identify, Select, and Combine Evidence for the Criteria?—If you review the articles earlier in this series that addressed overviews[7] and practice guidelines,[5,6] you will find guides for deciding whether the authors used explicit and rigorous methods to identify, select, and combine available evidence. How does the PTCA audit mentioned in our opening scenario measure up? Reading the full article, you see at once that some of the methods are described in a companion article on coronary artery bypass graft (CABG) surgery.[8] The investigators undertook a systematic literature review, with a comprehensive search and analysis of risks and benefits of PTCA in various patient subgroups.[1,8] The full literature review on PTCA is a separate background document, with explicit inclusion and exclusion criteria.[9] Like an iceberg, guidelines and clinical audit criteria often represent a "visible tip," supported by a large literature review that most journals don't wish to publish, and most clinicians won't want to read. Thus, as is the case here, you will sometimes have to rely on a description of how the literature was assembled and distilled.

User's Guides to Applying the Results of a Process-of-Care Audit

Are the criteria valid?
- Was an explicit and sensible process used to identify, select, and combine evidence for the criteria?
- What is the quality of the evidence used in framing the criteria?
- If necessary, was an explicit, systematic, and reliable process used to tap expert opinion?
- Was an explicit and sensible process used to consider the relative values of different outcomes?
- If the quality of the evidence used in originally framing the criteria was weak, have the criteria themselves been correlated with patient outcomes?

Were the criteria applied appropriately?
- Was the process of applying the criteria reliable, unbiased, and likely to yield robust conclusions?
- What is the impact of uncertainty associated with evidence and values on the criteria-based ratings of process of care?

Can you use the criteria in your own practice setting?
- Are the criteria relevant to your practice setting?
- Have the criteria been field-tested for feasibility of use in diverse settings, including settings similar to yours?

What Is the Quality of the Evidence Used in Framing the Criteria?—After assessing the methods for search and synthesis of the evidence, you must still decide on the quality of the evidence itself. Are the criteria based on evidence from high-quality studies, preferably definitive randomized trials or meta-analyses of multiple trials? Are most of the key indications for the service covered by trial evidence, or must observational evidence, inference, and expert opinion be brought frequently into play? If the latter is required, the validity of the audit criteria is reduced.

The PTCA example is germane here. The RAND group highlights that, at the time they conducted their work, no randomized trial evidence of PTCA vs alternative therapies existed for stable angina.[1] However, their literature review runs only to 1990.[1,9] You recall seeing trials of PTCA vs CABG in the literature and undertake another literature search that turns up abstracts reporting on 1 randomized trial of PTCA vs medical therapy in stable single-vessel disease[10] and 4 reporting on PTCA vs CABG.[11-14] This new evidence highlights that any audit criteria must be up-to-date, since what is optimal practice at one time may be malpractice a short time later. Investigators could now create stronger criteria based on the higher-quality evidence available from these recent randomized trials.

If Necessary, Was an Explicit, Systematic, and Reliable Process Used to Tap Expert Opinion?—To the extent expert opinion is tapped in framing criteria, there should be an explicit process for selecting panelists, and a sensible, systematic method for collating their judgments. The RAND group uses an original[15] and widely emulated multispecialty panel process that is clearly outlined in the PTCA report and companion article on CABG.[1,8] Specifically, for PTCA, a group of 9 panelists was assembled based on nominations of recognized experts by national specialty societies; the panelists were chosen from different geographic areas of the United States, from academic and private practice, and from different specialties (eg, cardiac surgeons, interventional and noninterventional cardiologists, and internists).[8] Each panelist independently rates hundreds of different case scenarios on a risk-benefit scale; each scenario describes a potential indication for the procedure or clinical service in question. Scenarios are rerated at a panel meeting after patterns of interpanelist agreement and disagreement are shown anonymously. The final set of panelists' ratings then determines whether a given indication is deemed potentially appropriate, uncertain, or inappropriate. Given the limited evidence from randomized trials, it seems very reasonable that the appropriateness of PTCA was graded as "uncertain" for 38% of the patients whose records were audited.[1]

A weakness of this method is that for any given clinical indication the researchers never make clear whether the appropriateness ratings rested primarily on research evidence or inference, extrapolation, and opinion. On the other hand, the RAND methods compare favorably with those used to create several utilization review tools now in widespread use. For example, various sets of diagnosis- and procedure-independent criteria are applied to hospital records to determine whether initial or continued stay in an acute care setting is necessary. These criteria are usually de-

rived in the first instance from implicit judgments of clinicians and utilization managers. One study found that from 28% to 74% of the verdicts reached by utilization review nurses using 3 of these instruments were rejected by physician panels.[16] Nonetheless, with the diffusion of managed care, criteria such as these have an enormous and continuing impact on the lives of patients, families, and health professionals.

Was an Explicit and Sensible Process Used to Consider the Relative Values of Different Outcomes?—The confusion of facts and values in expert judgments is a recurrent issue in these exercises. Most treatment decisions involve trade-offs. The randomized trials of CABG vs PTCA highlight this issue. Percutaneous transluminal coronary angiography has a slightly lower early mortality, along with lower initial costs and more rapid recovery from the procedure. Longer-term mortality data are similar, but CABG patients appear to achieve better symptom relief, have decreased use of medication, and require fewer subsequent procedures.[11-14] Panelists' ratings in the RAND study presumably reflected these types of trade-offs, but we cannot be sure that patients themselves would make the same choices. This issue is especially important for "uncertain" indications, where patients' preferences must be given special weight. However, chart audits and concurrent reviews using explicit criteria do not lend themselves to capturing patients' preferences and values.

Indeed, studies of expert panels show that surgeons' ratings of surgical options are more favorable than physicians and that medical generalists are more negative in procedural appropriateness ratings than medical specialists who do the procedure.[17-20] This again emphasizes that you should look for a clear description of how the panel was assembled along with the members' specialties and any organizations they are representing. Even when panels have similar practitioner profiles, the nationality of the panel markedly affects the criteria and the results of applying them to actual cases.[20,21] Perceptions of the values of different oucomes will continue to vary, but researchers should try to elucidate these issues whenever possible.

If the Quality of the Evidence Used in Originally Framing the Criteria Was Weak, Have the Criteria Themselves Been Correlated With Patient Outcomes?—When audit criteria follow directly from evidence from randomized trials, a link to outcomes can be assumed. For example, because the medications have been shown by systematic overviews of randomized trials to lower mortality,[22] substandard practice would be strongly suggested if an audit of prescribing practices after myocardial infarction showed that patients without contraindications were not receiving aspirin or β-blockers.

When weaker evidence and expert opinion form the basis for criteria, investigators (and users) can add strength to the criteria by determining how outcomes correlate with adherence to the criteria. Are outcomes improved or are outcomes similar despite decreased costs? These studies are tantamount to assessing a therapeutic intervention and could be critically appraised using criteria we have suggested in prior Users' Guides.[23,24] For example, researchers might randomly allocate practices or practitioners to usual care vs a program of concurrent audit, focusing on the service(s) of interest.

Although the design is much weaker, the impact of utilization review criteria can also be assessed using so-called historical controls. Here one would compare patient experience before and after a program of audit or prospective case management is implemented. Yet another option is to determine whether patients meeting the criteria who do not undergo a procedure have poorer outcomes than those who receive the procedure as indicated. As an example, the RAND group assembled a cohort of 671 subjects sampled from patients undergoing coronary angiography in 6 Los Angeles, Calif, hospitals, and followed them for a median of 2 years.[25] Patients meeting panel criteria for revascularization were examined: those who did not undergo revascularization had significantly worse outcomes than those who received either PTCA or CABG.[25] In general, we suggest that clinicians should seek outcomes-based evidence to support the safety and/or effectiveness of various utilization review tools and managed care programs.

Were the Criteria Applied Appropriately?

Audit criteria based on sound evidence can be poorly applied. This section may help you either to critique the published results of a utilization review undertaken for research purposes or to apply audit criteria to your own practice setting.

Was the Process of Applying the Criteria Reliable, Unbiased, and Likely to Yield Robust Conclusions?—Application of explicit process-of-care criteria often rests on data derived from retrospective chart reviews by professional auditors. Your confidence in their findings should be strengthened by evidence for reliability, eg, if 2 or more auditors generate the same data from the same patients' records or if the findings agree with those of a reference auditor with proven expertise. Such reproducibility demands very explicit definitions of the clinical variables incorporated into the criteria, eg, if PTCA is deemed appropriate for refractory unstable angina with single-vessel coronary disease, then there should be a clear definition of refractory unstable angina.

In the RAND study of PTCA in New York State hospitals,[1,8] the interauditor reliability of the chart review process is not described, and there is no mention of agreement with a criterion-standard abstractor. However, the process they used is well established, with good interabstractor reliability for other services.[26] A particular strength of the RAND process is a series of checks, wherein the auditors' work is reviewed by a nurse-specialist, and information on key clinical details is copied verbatim from the medical record for interpretation by trained physicians.[1,8]

Standardization of explicit audit criteria and the drive for reliable work by abstractors does exist in tension with a potential lack of responsiveness to mitigating clinical factors. Most utilization reviews, including the RAND PTCA study,[1,8] apply audit criteria as a screening test. If the explicit review shows potential problems with the appropriateness of a service, the case is assessed by experienced clinicians to preclude false-positive results. However, this introduces more subjectivity into the audit and raises the question as to why a sample of supposedly appropriate charts is not also reviewed for false-negative results. There is no easy resolution of this tension.

As to potential biases in practice audits, these are of more concern when implicit reviews are undertaken. Blinding as to institutional or practitioner identity is then desirable, and patient outcomes should also ideally be masked, as physicians are more likely to rate identical cases and care processes as inappropriate when there are severe adverse outcomes.[27] In this respect, it is unfortunate that some licensing and discipline bodies respond to complaints with unblinded implicit audits of the "problem practice" without comparison samples from other practices. However, in explicit criteria-based audits, biases can also be introduced through skewed sampling of practitioners, hospitals, and patients. The RAND investigators appropriately selected a random sample of both hospitals and patients for their PTCA study.

Last, it is crucial that enough cases be reviewed to draw robust conclusions. In the PTCA study, about 1500 charts were

reviewed. Institutions had from 1% to 9% inappropriate procedures, but the investigators could not exclude the play of chance as an explanation for the differences. Differences of this magnitude, if real, would be important to patients, payers, and policymakers. Thus, this sample size may have been insufficient for the investigators to detect important differences in quality among hospitals.

What Is the Impact of Uncertainty Associated With Evidence and Values on the Criteria-Based Ratings of Process of Care?—Limitations of evidence and uncertainty about values may suggest different criteria for appropriateness, and investigators should examine the impact of these different criteria. This may be done in a number of ways. If panelists have disagreed, investigators might present alternative results based on ratings from both the harsher and more lenient raters. Alternatively, one could look at the implications of assuming that ratings of "uncertain" represent adequate or inadequate care. This examination of alternative ratings is a form of sensitivity analysis as discussed in our Users' Guides to decision analysis.[28,29] The RAND report on PTCA in New York[1] offers extensive sensitivity analyses, including an exploration of how cases were placed in the uncertain category (eg, by explicit ratings of uncertain risk-benefit ratio; by being rated appropriate for revascularization rather than medical therapy, but with CABG preferred to PTCA; and by panelist disagreement).

Can You Use the Criteria in Your Own Practice Setting?

Even if the criteria are adequate in terms of their validity and you are satisfied with your understanding of how they should ideally be applied, it may not be reasonable or feasible to use them in your own practice setting.

Are the Criteria Relevant to Your Practice Setting?—Medical practice is always shaped by an amalgam of evidence, values, and circumstances. We noted earlier that expert panels generate rather different sets of audit criteria in different countries. Although the task is difficult, you should consider intangibles such as your local medical culture and practice circumstances before importing a particular set of audit criteria that may not be relevant. The stronger the evidence on which the criteria are based, the less you need to consider local factors; for example, few medical cultures would reject a practical intervention that was definitively proven in a randomized trial to yield major reductions in all-cause mortality. With weaker evidence, however, the judgments are less straightforward. For example, it is unlikely that US patterns of PTCA utilization could be readily transplanted to the United Kingdom, with its tradition of comparative restraint in adopting invasive cardiovascular procedures.[20]

Have the Criteria Been Field-Tested for Feasibility of Use in Diverse Settings, Including Settings Similar to Yours?—Even if criteria are sufficiently valid and relevant, there are still feasibility issues to be faced. The RAND criteria-based assessments of PTCA were applied successfully in diverse hospital settings in New York,[1] but the work was done by a highly skilled team of researchers and auditors. You will want to know how long it takes to train staff to use the criteria and the costs of available training programs. Costs per case for the audit must include training and labor charges, as well as any purchase charges for special audit forms. Consideration must also be given to whether the criteria are going to be applied for concurrent case management. Errors associated with use of the criteria will have immediate consequences for individual patients and physicians in a managed care program, and the logistics of concurrent review can be daunting. Nonetheless, many busy hospitals already apply a wide range of concurrent utilization review criteria, as most practitioners know to their occasional frustration.

CONCLUSION AND RESOLUTION

This Users' Guide provides an approach to critically appraising quality-of-care studies that focus on the process of delivering a service. We have focused on methods that involve a blend of evidence and expert opinion or judgment, as these are widely applied in deriving utilization review criteria. However, on occasion, more straightforward approaches will be possible. As noted above, one can draw on randomized control trial or overview evidence in isolation and derive indications where the service is either highly effective or definitively proven to be inferior to alternatives. Other indications can be set aside as resting in the "gray zone" of uncertainty where reasonable persons can disagree.[21] While this approach is simpler and less controversial, there are 2 problems with streamlined criteria. The first problem is that randomized trial evidence is often limited and may never become available for some procedures and clinical situations.[21,30] A commitment to evidence-based practice cannot preclude the reasonable use of clinical judgment, inference, and extrapolation.[21] The second problem is that trials are better at helping us decide what to do than what not to do. Expert panels, with all their limitations, do permit detailed assessments of inappropriate and uncertain indications.

Currently, however, the proliferation of quality-of-care assessments has greatly outstripped the credible research in the field.[3,4] Despite the eager embrace of managed care, the measurement of quality of care remains difficult. Reliability of implicit assessments is low, while the available evidence for derivation of explicit criteria is often limited. Furthermore, the overall impact of these criteria on clinical behaviors, system costs, and patients' health outcomes is difficult to know as they are seldom evaluated in formal prospective studies and are often coupled with changes in practice organization and/or reimbursement that in themselves may change behavior.

The resolution of our scenario has you revisiting the library to obtain the new randomized studies of PTCA vs medical therapy and PTCA vs CABG that have appeared since 1990. These articles are digested with lunch at your desk in the following few days. At the next medical advisory committee meeting, you are prepared to discuss the RAND study on PTCA, as well as the new randomized trials. However, the chief of cardiology speaks first. She informs the committee that she has been to the health records department and has visited colleagues at the 2 area hospitals with different utilization statistics. She presents data showing that the discrepant utilization profile is almost completely attributable to acute PTCA for myocardial infarction, which your hospital's cardiologists offer as an alternative to thrombolysis for patients presenting early after the onset of symptoms. Her literature search shows 4 relevant randomized trials.[31-34] The chief of cardiology rightly claims: "The trial evidence supports direct PTCA as a safe and effective alternative to intravenous thrombolysis when patients present early and are suitable candidates for emergency angioplasty." The meeting briefly degenerates into a squabble over whether the administrator should apologize to the hospital's cardiologists, but the hospital chief executive officer rescues his junior colleague by questioning whether the hospital can be cost-competitive if it relies more on PTCA than its neighboring institutions. Amid grumbles about "the eternal bottom line" and "the economic oath" from the other physicians present, you and the chief of cardiology volunteer each other to research the comparative costs of PTCA and thrombolysis for acute myocardial infarction.

The following individuals are new members of the Evidence-Based Medicine Working Group and will be active contributors to the remainder of the series: Antonio L. Dans, MD, MSc, Department of Internal Medicine, University of the Philippines, Manila; Leonila F. Dans, MD, MSc, Department of Pediatrics, University of the Philippines, Manila; Paul Glasziou, MB, PhD, Department of Social and Preventive Medicine, University of Queensland, Herston, Queensland, Australia; Lee Green, MD, MPH, Department of Family Practice, University of Michigan, Ann Arbor; Daren Heyland, MD, Department of Medicine, McMaster University, Hamilton, Ontario; Les Irwig, MBBCh, PhD, FFPHM, Department of Epidemiology and Biostatistics, School of Public Health, University of Sydney, Sydney, New South Wales, Australia; Alejandro Jadad, MD, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario; and Thomas B. Newman, MD, Department of Laboratory Medicine, University of California at San Francisco.

We are grateful for critical commentary on the manuscript by 2 anonymous referees, Adrienne G. Randolph, MD, fellow at the Institute for Health Policy Studies at the University of California at San Francisco, and Geoffrey M. Anderson, MD, PhD, senior scientist at the Institute for Clinical Evaluative Sciences, University of Toronto (Ontario).

References

1. Hilborne LH, Leape LL, Bernstein SJ, et al. The appropriateness of use of percutaneous transluminal coronary angioplasty in New York State. *JAMA.* 1993;269:761-765.
2. Naylor CD, Guyatt GH, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, X: how to use an article reporting variations in the outcomes of health services. *JAMA.* 1996;275:554-558.
3. Health Services Research Group (Basinski ASH, Naylor CD, Ferris LE, Williams JI, Llewellyn-Thomas HA, Cohen MM). Quality of care, 1: what is quality and how can it be measured? *Can Med Assoc J.* 1992;146:2153-2158.
4. Health Services Research Group (Basinski ASH, Naylor CD, Ferris LE, Williams JI, Llewellyn-Thomas HA, Cohen MM). Quality of care, 2: quality of care studies and their consequences. *Can Med Assoc J.* 1992;147:163-167.
5. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt G, for the 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;274:570-574.
6. Wilson MC, Hayward RSA, Tunis SR, Bass EB, Guyatt G, for the Evidence-Based Medicine Working Group. Users' guide 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;274:1630-1632.
7. Oxman AD, Cook DJ, Guyatt GH, for the Evidence-Based Medicine Working Group. Users' guide to the medical literature, VI: how to use an overview. *JAMA.* 1994;272:1367-1371.
8. Leape LL, Hilborne LH, Park RE, et al. The appropriateness of use of coronary artery bypass graft surgery in New York State. *JAMA.* 1993;269:753-760.
9. Hilborne LH, Leape LL, Kahan JP, Park RE, Kamberg CJ, Brook RH. *Percutaneous Transluminal Coronary Angioplasty: A Literature Review and Ratings of Appropriateness and Necessity.* Santa Monica, Calif: RAND; 1991. Document JRA-01.
10. Parisi AF, Folland ED, Hartigan P, for the Veterans Affairs ACME Investigators. A comparison of angioplasty with medical therapy in the treatment of single-vessel coronary artery disease. *N Engl J Med.* 1992;326:10-16.
11. Hamm CW, Reimers J, Ischinger T, Rupprecht HJ, Berger J, Bleifeld W. A randomized study of coronary angioplasty compared with bypass surgery in patients with symptomatic multivessel coronary disease: German Angioplasty Bypass Surgery Investigation (GABI). *N Engl J Med.* 1994;331:1037-1043.
12. King SB III, Lembo NJ, Weintraub WS, et al. A randomized trial comparing coronary angioplasty with coronary bypass surgery: Emory Angioplasty Versus Surgery Trials (EAST). *N Engl J Med.* 1994;331:1044-1050.
13. Coronary angioplasty versus coronary artery bypass surgery: the Randomized Intervention Treatment of Angina (RITA) trial. *Lancet.* 1993; 341:573-580.
14. Rodriguez A, Boullon F, Perez-Balino N, Paviotti C, Liprandi MI, Palacios IF, for the ERACI Group. Argentine randomized trial of percutaneous transluminal coronary angioplasty versus coronary artery bypass surgery in multivessel disease (ERACI): in-hospital results and 1-year follow-up. *J Am Coll Cardiol.* 1993;22:1060-1067.
15. Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care.* 1986; 2:53-63.
16. Strumwasser I, Paranjpe NV, Ronis DL, Share D, Sell LJ. Reliability and validity of utilization review criteria: appropriateness evaluation protocol, standardized medreview instrument, and intensity-severity-discharge criteria. *Med Care.* 1990; 28:95-111.
17. Leape LL, Park RE, Kahan JP, Brook RH. Group judgments of appropriateness: the effect of panel composition. *Qual Assur Health Care.* 1992; 4:151-159.
18. Fraser GM, Pilpel D, Hollis S, Kosecoff JB, Brook RH. Indications for cholecystectomy: the results of a consensus panel approach. *Qual Assur Health Care.* 1993;5:75-80.
19. Kahn KL, Park RE, Brook RH, et al. The effect of comorbidity on appropriateness ratings for two gastrointestinal procedures. *J Clin Epidemiol.* 1988; 41:115-122.
20. Brook RH, Kosecoff JB, Park RE, Chassin MR, Winslow CM, Hampton JR. Diagnosis and treatment of coronary artery disease: comparison of doctors' attitudes in the USA and the UK. *Lancet.* 1988;1:750-753.
21. Naylor CD. Grey zones of clinical practice: some limits to evidence-based medicine. *Lancet.* 1995; 345:840-842.
22. Yusuf S, Wittes J, Friedman L. Overview of results of randomized clinical trials in heart disease, I: treatments following myocardial infarction. *JAMA.* 1988;260:2088-2093.
23. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention, A: are the results of the study valid? *JAMA.* 1993;270:2598-2601.
24. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention, B: what were the results and will they help me in caring for my patients? *JAMA.* 1994;271:59-63.
25. Kravitz RL, Laouri M, Kahan JP, et al. Validity of criteria used for detecting underuse of coronary revascularization. *JAMA.* 1995;274:632-638.
26. Kahn KL, Kosecoff J, Chassin MR, et al. Measuring the clinical appropriateness of the use of a procedure: can we do it? *Med Care.* 1988;26:415-422.
27. Caplan RA, Posner KL, Cheney FW. Effect of outcome on physician judgments of appropriateness of care. *JAMA.* 1991;265:1957-1960.
28. Richardson WS, Detsky AS, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, VII: how to use a clinical decision analysis, A: are the results of the study valid? *JAMA.* 1995;273:1292-1295.
29. Richardson WS, Detsky AS, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, VII: how to use a clinical decision analysis, B: what are the results and will they help me in caring for my patients? *JAMA.* 1995;273:1610-1613.
30. Fink A, Brook RH, Kosecoff J, Chassin MR, Solomon DH. Sufficiency of clinical literature on the appropriate uses of six medical and surgical procedures. *West J Med.* 1987;147:609-614.
31. Grines CL, Browne KF, Marco J, et al, for the Primary Angioplasty in Myocardial Infarction Study Group. A comparison of immediate angioplasty with thrombolytic therapy for acute myocardial infarction. *N Engl J Med.* 1993;328:673-679.
32. Zijlstra F, de Boer MJ, Hoorntje JC, Reiffers S, Reiber JH, Suryapranata H. A comparison of immediate coronary angioplasty with intravenous streptokinase in acute myocardial infarction. *N Engl J Med.* 1993;328:680-684.
33. Gibbons RJ, Holmes DR, Reeder GS, Bailey KR, Hopfenspirger MR, Gersh BJ. Immediate angioplasty compared with the administration of a thrombolytic agent followed by conservative treatment for myocardial infarction: the Mayo coronary care unit and catheterization laboratory groups. *N Engl J Med.* 1993;328:685-691.
34. Ribeiro EE, Silva LA, Carneiro R, et al. Randomized trial of direct coronary angioplasty versus intravenous streptokinase in acute myocardial infarction. *J Am Coll Cardiol.* 1993;22:376-380.