Quality of Care and Technology Assessment: Report of a Forum of the Council on Health Care Technology (1988)

Pages
160

Size
5 x 9

ISBN
0309319676

Kathleen N. Lohr and Richard A. Rettig, Editors; Institute of Medicine

Report of a Forum of the
Council on Health Care Technology

Quality of Care and Technology Assessment

Kathleen N. Lohr and
Richard A. Rettig, editors

Institute of Medicine

NATIONAL ACADEMY PRESS • WASHINGTON, D.C.

THE INSTITUTE OF MEDICINE was chartered in 1970 by the National Academy of Sciences to enlist distinguished members of appropriate professions in the examination of policy matters pertaining to the health of the public. In this, the Institute acts under both the Academy's 1863 congressional charter responsibility to be an adviser to the federal government, and its own initiative in identifying issues of medical care, research, and education.

THE COUNCIL ON HEALTH CARE TECHNOLOGY was established in 1986 by the Institute of Medicine of the National Academy of Sciences as a public-private entity to address issues of health care technology and technology assessment. The council is committed to the well-being of patients as the fundamental purpose of technology assessment. In pursuing that goal, the council draws on the services of the nation's experts in medicine, health policy, science, engineering, and industry.

The Forum and this monograph were supported in part by a grant to the Council on Health Care Technology of the Institute of Medicine from the National Center for Health Services Research of the U.S. Department of Health and Human Services (grant 5 R09 HS055 26 02). The opinions and conclusions expressed here are those of the authors and do not necessarily represent the views of the Department of Health and Human Services, the National Academy of Sciences, or any of their constituent parts.

Library of Congress Catalog Card Number: 88-062514.

ACKNOWLEDGMENTS

The Forum on Quality of Care and Technology Assessment would not have been possible without the dedicated efforts of Leslie Hardy, M.H.S., who, as a staff officer of the Council on Health Care Technology at the time, devoted long hours to the logistics of the meeting. Similarly, the publication of this monograph would not have been possible without the many hours of skillful and cheerful preparation of several drafts of this manuscript by Brenda Patterson of the council staff. The original version of these papers was most competently edited by Marcia Feinleib. On behalf of the members of the council and the authors in this volume, the editors thank them all for their contributions.

The editors wish to thank Enriqueta Bond, Ph.D., Director of the Division of Health Promotion and Disease Prevention, for her substantial contributions to the council and its activities. They are grateful as well to Samuel O. Thier, M.D., President of the Institute of Medicine, for inspired leadership and support of council initiatives. Finally, they acknowledge a considerable debt to William N. Hubbard, M.D., Chairman of the council, and to Jeremiah A. Barondess, M.D., Co-Chairman; their unique blend of guidance and support through the past year has made this forum and the entire forum series possible.

Council on Health Care Technology

PREFACE

In the recent past the interests of different groups concerned with health care have focused on the use of medical technologies—their impacts on safety, efficacy, and effectiveness; cost-effectiveness and cost-benefit; quality; and their social, legal, and ethical implications. The sum of these varied interests is the field of health care technology assessment.

The Council on Health Care Technology was created to promote the development and application of technology assessment in health care and the review of health care technologies for their appropriate use. The council was established as a public-private enterprise at the Institute of Medicine, a component of the National Academy of Sciences, through the Health Promotion and Disease Prevention Amendments of 1984 (P.L. 98-551, later amended by P.L. 99-117). In 1987 the U.S. Congress extended support for the council as a public-private venture for an additional three years (by P.L. 100-177).

The goals and objectives of the council, as stated in the report of its first two years of operations, are "to promote the development and application of technology assessment in medicine and to review medical technologies for their appropriate use. The council is guided in its efforts by the belief that the fundamental purpose of technology assessment is to improve patient well-being and the quality of care." In pursuing these goals the council seeks to improve the use of medical technology by developing and evaluating the measurement criteria and the methods used for assessment; to promote education and training in assessment methods; and to provide technical assistance in the use of data from published assessments.

The council conducts its activities through several working and liaison panels. Members of these panels reflect a broad set of interested constituencies—physicians and other health professionals, patients and their families, payers for care, biomedical and health services researchers, manufacturers of health-related products, managers and administrators throughout the health care system, and public policymakers. In addition, it carries out councilwide activities that utilize the specific assignments of more than one panel.

The Forum Series is one of the councilwide activities. The Forum Series seeks to facilitate an exchange of views among the many and diverse groups concerned with health care technology. Themes addressed

by these symposia include the interdependence among medical technology assessment, quality assessment, and assurance of the quality of patient care; the performance of existing technology assessment activities; the appropriate use of technology to minimize risks posed to special or vulnerable patient populations; and the development, application, and appropriate use of technology.

A guiding principle of the council is a special focus on outcome measures that coincide with patient well-being, quality of health care, and quality of life. To highlight this principle, the first forum dealt with quality of care and technology assessment. This report is the record of that symposium.

William N. Hubbard, Jr., Chairman

Jeremiah A. Barondess, Co-Chairman

Overview of the Forum

Kathleen N. Lohr and
Richard A. Rettig

The first volume in the Council on Health Care Technology's Forum Series focuses on the intersection of technology assessment (TA) and quality of care. Two premises dictated this topic. First, the council observed that two fields—medical TA and quality assessment and assurance (QA)—have developed as separate endeavors; the disciplines evolved with different vocabularies, different methods, and under the leadership of different theoreticians and practitioners. Second, this separation, however understandable in historical terms, is regarded as unproductive, because the two fields cannot reach their full potential alone. To explore the implications of these premises, the council invited two groups to describe their TA or QA programs and to reflect on how their efforts might improve patient well-being and foster progress in the practice of medicine.

This overview gives a synopsis of the major presentations, which appear in full in the remainder of this monograph. It concludes with a brief discussion of the major themes, issues, and recommendations that emerged from the forum.

FRAMING THE ISSUES

In the opening presentation Saul Farber reflects on the proposition that medicine is a learned profession. Its goals are those of continuing professional competence, dedicated public service, unwavering responsibility to individual patients, and adherence to scientific and scholarly principles of problem analysis. Its methods are those of logical problem solving, self-examination, and self-discipline. Critical to these goals and these methods is the knowledge that selected practices and procedures have demonstrated benefits in defined circumstances (TA) and that they are provided by physicians with the appropriate capabilities to utilize them (QA). In the present times of economic strain and scientific revolution, achievement of these goals can contribute to the traditional percep-

tion that "the profession of medicine serves the public interest and serves it well."

Robert Brook discusses the links between TA and QA by emphasizing two major points: both quality and technology assessments are vital to a successful health system, and each type of assessment depends on the other's success. "[A]mple evidence suggests that quality of care can be assessed, that technologies can be assessed, and that both types of assessments can dramatically change the practice of medicine for the better."

Dr. Brook offers examples of how the two fields have not been integrated and discusses why this should change. He calls for study of the epidemiology of quality of care, for detailed analysis of the TA outputs, and for more complete sharing of the findings of both enterprises. He concludes with specific recommendations for a mature TA and QA system that would provide contemporary guides to good medical practice and produce "a more active patient, a more satisfied physician, and, ultimately, better health for all at a price we would be willing to pay."

THE FIELD OF QUALITY ASSURANCE

The examination of QA is not an academic exercise. Regardless of their limitations, a variety of QA programs are being deployed. Moreover, the field is moving away from the question of whether quality of care can be measured and toward the question of how to assess and assure quality in a comprehensive yet efficient way. Major issues involve the elaboration and refinement of methods. To aid our understanding, several papers in quality measurement and assurance are presented.

The range of programs they describe is deliberately broad and underscores the variety and adaptability of QA methods. Included are QA efforts in the prepaid group practice setting, responses of New York State hospitals to comprehensive state regulation, the outcome-oriented initiatives of the Joint Commission on Accreditation of Healthcare Organizations, voluntary self-assessment efforts by family practitioners in Canada, and the activities of the professional peer review organizations that review care provided through the Medicare program.

Donald Berwick discusses current QA work at the Harvard Community Health Plan (HCHP) to develop methods applicable to certain segments of the "managed care sector" of the health care industry. Several forces make quality measurement an "urgently felt need": the transition from a profession-dominant to a purchaser-dominant industry; large variations across patients and physicians in use of health care resources; improved data management capabilities; and managed care systems that

provide opportunities to gather and use quality-related information. A major obstacle to QA remains: "Despite a deep intellectual tradition [in the quality assessment field], we almost completely lack a useful applied technology for the measurement of quality in health care."

After briefly discussing five key characteristics of a quality measurement system, Dr. Berwick describes the initial HCHP efforts to devise a quality measurement system to help the plan's managers preserve and enhance the quality of care delivered. The system is predicated on methods of "industrial quality control" and principles of "continual improvement" applied in the assessment of eight critical attributes of quality: outcomes, technical process of care, access, interpersonal aspects of care, integration of care, physical environment, staff attitudes, and reliability of the system of care. He concludes that an appropriate research agenda in this arena would have "three cornerstones: clarifying values, enhancing design and process, and measurement . . . unified by the goal of continual improvement."

Dr. Berwick's paper focuses on a QA effort developed by and for the benefit of a single health care plan. In contrast, Thomas Morris discusses the new QA efforts of the Hospital Association of New York State (HANYS) and the impact of the New York State regulatory system on a large hospital community. Sweeping state regulations promulgated early in 1986 necessitated a nearly complete revamping of existing QA procedures in the state's hospitals. Two other factors contributed to the climate of reform: the release of hospital-specific mortality data (both by the Health Care Financing Administration and by the state's Department of Health) and a new "cosurvey process" to augment surveys normally conducted by the Joint Commission on Accreditation of Healthcare Organizations.

In 1986 HANYS established a Quality Assurance Task Force charged with developing quality indicators, standards of care, and a model plan. Among its early achievements was the publication of *Quality Assurance Plan Guidelines*, which emphasizes that the responsibility for QA within an institution lies with its chief executive officer. HANYS is now turning to indices of quality care, especially ones sensitive to severity of illness and to important characteristics of the populations that New York's hospitals serve. Despite these advances, a crucial concern is that these changes have occurred in an "unnecessarily adversarial setting." This theme—hostile relations between the medical community and health regulators, and the worry it provokes about the future—is echoed more than once in the forum.

Historically, the best-known group committed to monitoring and ensuring the quality of care received in hospitals has been the Joint Commission on Accreditation of Hospitals (now, of Healthcare Organizations).

Traditionally, the Joint Commission approached its QA mission as a voluntary, private sector effort. James Roberts describes the Joint Commission's new directions—its "Agenda for Change."

At the heart of the Joint Commission's multiyear research and development program lie five key efforts: selecting clinical performance indicators, identifying appropriate severity-of-illness or case-mix adjustors, selecting organizational performance indicators, improving the procedures for monitoring between scheduled surveys, and improving the survey process. This is an ambitious program—one that must overcome several societal and technical obstacles that may impede the Joint Commission's work and, by extension, all QA efforts. Dr. Roberts concurs with Dr. Berwick's conclusion that the key to quality of care may be an organization's commitment to constant improvement in performance.

Another private sector, voluntary QA effort—one directed to ambulatory practice—is described by Alexander Borgiel of the College of Family Physicians of Canada. The college has developed a practice assessment strategy combining three elements: physician questionnaire and interview, medical chart abstract, and patient questionnaire. Diagnosis-specific criteria for the chart review were developed through an extensive pilot project. They are applied through personal-computer-based procedures (described by Dr. Borgiel) that use portable computers that permit chart abstracting to be done in the physician's office. Despite its comparative newness, the practice assessment program has proven to be "a positive, constructive method of practice analysis in ambulatory care [that] . . . provides the reassurance that all family physicians need from time to time."

In the United States, peer review of ambulatory, and later inpatient, care had its origins in the "foundations for medical care" movement pioneered in California. As the movement spread eastward it evolved (in Medicare) into Professional Standards Review Organizations and, later, Peer Review Organizations. The history of this phenomenon is outlined by Kenneth Platt, from his position as a nationally recognized leader of statewide peer review efforts.

Dr. Platt reflects that "quality is a poorly defined entity, subject to varying definitions from differing perspectives . . . [and] the technique of quality assurance is in its infancy." Notwithstanding these realities, the tools of utilization review and QA can be used in TA: "[I]t would appear prudent to assess [new and expensive] technologies on a limited, trial basis prior to widespread implementation." Thus, he echoes earlier papers in two ways, emphasizing the importance of the TA/QA link and stressing that being a physician "must be seen as a sacred and honored trust to be used humbly and for the public good."

TECHNOLOGY ASSESSMENT

Five authors describe significant TA programs that cover a spectrum as broad as that for QA. The key distinguishing feature of the TA efforts is the diversity of sponsors—professional societies, investor-owned provider systems, third-party payers, manufacturers, and the government. Similarly diverse are the motivations for TA: for instance, defining "optimal strategies [to be applied] to individual patient problems" or identifying "truly valuable technology . . . [to] be considered for coverage" by "prudent purchasers" in the health insurance industry.

The American College of Physicians (ACP) sponsors perhaps the most active TA program within the medical community. Its Clinical Efficacy Assessment Project (CEAP) attempts to improve "the quality of medical practice by providing practitioners with guidelines and recommendations regarding the appropriate use of tests, procedures, and therapies and the rationale for those recommendations." Earl Steinberg describes how CEAP assessments rely principally on literature syntheses, secondary data analysis, and extensive review by ACP members and other interested parties. Assessments are published as lengthy background papers by the assessment authors, with accompanying statements of clinical practice recommendations by the ACP.

Other activities discussed by Dr. Steinberg include developing guidelines for the use of common diagnostic tests; establishing a "national network of physicians who would systematically collect information regarding the use and impact of selected tests and procedures in . . . ambulatory practice"; and writing statements about the knowledge, skills, and training required to establish and maintain competence in numerous internal medicine procedures. The ACP aims to extend its TA program to evaluate the relationship between processes of care and patient outcomes—a goal precisely in line with one major objective of quality-of-care assessment and research.

Multihospital systems have also been rapidly revising TA and QA activities. John Moxley, in a presentation that emphasizes practical realities, describes the experiences of American Medical International, Inc. (AMI). Until quite recently, major technology questions at AMI centered on whether physicians wanted a new technology and whether it had appropriate regulatory approvals; if the answers were yes, the technology was acquired. Now, by contrast, structural changes in the health care sector dictate that expenditures for technology be markedly reduced. Hence, AMI reviews alternative types and brands of a given technology (even conducting "blinded" field trials) and engages in complex negotiations with manufacturers of the preferred choice.

Numerous pressures on the hospital community will likely rein-

force this increasingly conservative attitude toward the acquisition of new or replacement technology and prompt better TA. In the process, however, they raise different issues: (1) how new technology will affect existing technology and the management of disease; (2) where new technology should be sited; (3) how technologies can be disseminated (e.g., from tertiary centers to community hospitals;) (4) what the useful life cycle of a technology is; (5) how payment processes by the Health Care Financing Administration can be improved.

Following these TA efforts are new QA initiatives. AMI plans to relate these two programs to each other more fully, as broader and more precise information from the former serves to improve the quality of patient care throughout their system. Dr. Moxley emphasizes the need "to stop creating adversarial relationships between the various constituencies involved in the provision of health care [and] to stop increasing the fear level . . . among physicians and hospital managers"; he restates the views of several speakers about the importance of the public and private sectors, and the professional and lay communities, working together on technology and quality-of-care issues.

In an interesting parallel to the comments of Dr. Platt about the origins of professional peer review efforts in California, Ralph Schaffarzick describes the history and development of TA efforts of the first nonprofit statewide physicians' service plan, California Blue Shield, California Physicians' Service. Its assessment efforts, like those of the ACP's CEAP process, are grounded in traditional expert-panel or group-judgment techniques. Because California Blue Shield is a third-party payer, however, the recommendations derived from the TA procedures are more directly related to coverage decisions.

The criteria, developed by the national Blue Cross and Blue Shield Association for coverage decision making, are instructive because they could just as easily be constructed for QA purposes:

- The technology must have final approval from the appropriate government regulatory bodies.
- The scientific evidence must permit conclusions concerning the effect of the technology on health outcomes.
- The technology must improve net health outcome.
- The technology must be as beneficial as any established alternative.
- Improved health outcomes from the technology must be attainable outside investigational settings.

Finally, Dr. Schaffarzick emphasizes that selective coverage of procedures (i.e., coverage limited to indications, medical centers, or qualified practi-

tioners) can and will enhance the quality of health care, especially as TAs evolve in more rigorous, structured, and sophisticated ways.

Morton Paterson, in reporting the experiences of one pharmaceutical company (Smith, Kline and French Laboratories), highlights rigorous quantitative aspects of TA and, in particular, use of cost-effectiveness analysis. Using cimetidine as his example,[1] Dr. Paterson describes the range of investigations (clinical trials, analysis of statistical series, retrospective cost studies) conducted to clarify the net health and cost impact of introducing this drug.

Although the evidence undeniably supports the cost-effectiveness of cimetidine, Dr. Paterson uses this fact to make two cautionary points relative to TA. First, for a technology to affect national statistics in a detectable way, it probably must be dramatically effective and appear in a relatively "quiet" environment. Second, clinical trial data probably overstate the impact of a new product or procedure; thus, they do not furnish appropriate expectations for cost reductions or gains in health and quality of life when the technology is more widely dispersed. To overcome some of these drawbacks, Dr. Paterson advances an argument for a new type of cost-effectiveness trial—one that combines "randomized prescribing of approved drugs and hands-off follow-up with recording of medical outcomes and determination of costs from routinely generated, computerized patient records."

A second limitation to cost-effectiveness analysis is that although costs of different technologies can be counted in dollars, the outcomes of using different technologies cannot; what is needed is a standard unit of effectiveness that permits the effects of different treatments to be compared on equivalent grounds. The "quality-adjusted life-year [QALY]" is such a tool, and its use in studies of auranofin[2] illustrates the power of this concept. Many instruments used to evaluate quality of life are based on measures related to patient outcomes, including activities of daily living and various global assessments of functioning and well-being; it is not a difficult transition to outcome measurement in QA terms.

Dr. Paterson's conclusions thus pertain both to QA and TA: "We need to overcome the problem of the artificiality of data from carefully controlled clinical trials. We need routinely generated and computerized data on patients' health care and work loss over time. Finally, we need to understand what cost-effectiveness . . . can mean for practical decision making. These are not insignificant hurdles, but they can probably be overcome."

[1] Cimetidine is a drug used in peptic ulcer disease that dramatically reduces the need for surgery and lowers the total cost of treating the illness.

[2] Auranofin is an oral gold agent for treatment of rheumatoid arthritis and would normally supplement standard regimens based on nonsteroidal anti-inflammatory agents.

John Gronvall of the Veterans Administration makes the final TA presentation as a representative of an agency that acts as provider, as technology assessor, and as quality assurer. Part of Dr. Gronvall's paper describes the very complex VA system, which provides care "for 1.4 million hospital patients and 20 million outpatients . . . through a network of 172 hospitals, 117 nursing homes, 228 outpatient clinics, and 16 domiciliaries." He emphasizes the interconnectedness of the VA's efforts in health services research, TA (including a major task force), and QA (including a systemwide peer review program devoted exclusively to quality-of-care issues).

In pursuing a philosophy of "excellence in health care" the VA tries to base its TA-QA relationship on objective critera. TA "creates the objective basis for assessment of each diagnostic or therapeutic step in a patient's care"; QA determines "what actual effect those individual steps have on accomplishing the ultimate outcome of all medical endeavors—improved health of the patient." In closing with these sentiments, Dr. Gronvall returns to many of the ideas enunciated earlier in the volume.

CHALLENGES AND OPPORTUNITIES FOR TECHNOLOGY ASSESSMENT AND QUALITY ASSURANCE

Summary of the Forum

Sheldon Greenfield in a formal summary asserts that "these two fields desperately need each other," and he cites several forum papers that overtly or implicitly document the interdependence of the two disciplines. To the question of how technology assessment can influence quality assurance, he proposes two answers: TA information is critical in establishing quality-of-care criteria and standards, and TA methods (e.g., rigorous cost-effectiveness and probabilistic techniques) will enhance quality assessments. To the parallel question of how quality assessment can help technology assessment, he notes the virtue of taking a broad view (as evaluations of quality, especially those based on patient outcomes, often do): i.e., quality assessments at least potentially deal with many aspects of patient care simultaneously (diagnosis, therapy, complications, quality of life, and patient preferences), and this perspective needs to be applied to technology assessment.

This line of reasoning supports a focus on patient well-being (such as, that adopted by the Council on Health Care Technology and endorsed by many quality-assessment efforts), but it also calls for extreme caution. "[O]utcomes have the potential of being measured and adjusted adequately to make meaningful statements about quality of care and technology use . . . [but these] adjustments or controlling factors must be dealt

with carefully." Well-validated outcome measures occupy a key place in both quality and technology assessments. Realizing the objective of a primary focus on patient well-being, however, requires a stable financial mechanism for conducting the necessary research into methods and for pursuing the assessments themselves. It also calls for patience and perseverance: "[A]void trying to judge things that we cannot judge right now, and avoid assuming that the methods available today will give us perfect measurements." If the TA and QA fields work together in these areas, the achievable benefit is improved health for the American people, according to Dr. Greenfield.

Implications of the Forum

Following the forum, many participants shared their reactions to the presentations and discussion in letters to council staff. That material and the content of the forum itself enabled us to draw several major conclusions about the TA and QA sectors. First, TA and QA share a common objective: They exist to improve patient well-being and quality of care.

Second, each endeavor has arrived at this objective by a different path. In the 1970s TA acquired a strong identity with cost containment in the minds of many observers. In the 1980s the focus has shifted—certainly in the council's deliberations—to an emphasis on clinical effectiveness. Concerns for adequacy of data, access to information sources, and assessment methods have reinforced this emphasis.

QA, by contrast, evolved from attention to structural and process measures to a substantial interest in outcome variables. This led to increased attention to data and methods issues that parallels the concerns of TA.

Third, the evolution of both fields derives in large measure from the aggressive efforts of both the public and the private sectors to control the growth of health care expenditures. Two questions flow from cost containment and drive the current demand for quality assessment. Does restricted use of resources pose a threat to quality? Does the absence of quality standards imply that resources are being inappropriately used or squandered?

Fourth, these developments have a considerable impact on the medical profession. In the past the guarantee of quality of care was assumed to reside in the attitudes and behaviors of a largely autonomous medical profession. Increasing evidence of poor quality of care, great variations in the use of services across similar populations, and instances of inappropriate care have challenged that grant of autonomy. The future holds out two related opportunities for the medical profession: to participate with both government and private payers in efforts to measure and

assure quality, and to assume active leadership in efforts to develop professional standard-setting programs.

Finally, the shared agenda of TA and QA needs to be recognized and elaborated. Common interests include data and methods. In addition, determining efficacy in the ideal setting and effectiveness in the everyday setting remain characteristics of TA; determining efficacy also anchors the analysis of process and outcomes that lies at the heart of quality-of-care studies. Conversely, quality of care presents patient well-being as the fundamental objective of both enterprises.

Perspectives in Quality Assurance and Technology Assessment

Saul J. Farber

I have long been both an observer and an active participant in the discipline of quality assurance. I try to assure the best provision of quality of care to the individual patient, at a reasonable cost, and with as little risk as possible of causing adverse effects and inconvenience to the patient. These all are also the goals of quality care.

I would like, first, to review the process of problem solving in medicine and, then, to examine the relationship between the concerns of the public and those of the medical profession in assuring that each patient gets the best available care. The outcomes we strive for as physicians are cure when possible, alleviation of pain and suffering, and, in the best of worlds, prevention—all directed to assuring our patients a higher quality of life. In emphasizing these goals and the professionalism of the medical practitioner, we serve both the public's interest and the best traditions of medicine.

THE PROFESSION OF MEDICINE

Medicine is a profession, which is defined as a calling, requiring specialized knowledge, often requiring long and intensive preparation including instruction in the skills and methods of practice as well as in the scientific, historical, or scholarly principles underlying such skills and methods. The profession maintains, by force of organization and concerted opinion, high standards of achievement and conduct. Its members are committed to a lifetime of continued study and to the kind of work that has for its prime purpose the rendering of public service. Each of these factors is important.

This essay was adapted from Dr. Farber's keynote address. The editors thank William N. Hubbard, Jr., M.D., for helpful comments on an earlier draft of the essay.

Historically a learned profession is defined as one of the three professions—theology, law, and medicine—traditionally associated with extensive learning or erudition. Although the definition comes from the dictionary, I would add that medicine, as a learned profession, has as its primary distinction the responsibility to render public service and to preserve life.

Problem Solving in Medicine

Every profession has a logic of its own. This logic is revealed in the problem-solving styles of each profession.

The problem-solving methods of differential diagnosis, which is at the heart of medical practice, were well described by Abraham Flexner in 1910. Although Flexner was not a physician, he studied the medical education system and sought by constructive criticism to ensure the vitality of medicine as a profession. Comparing the logic of clinical diagnosis with the logic of research in the basic sciences, he wrote, "the main intellectual tool of the investigator is the working hypothesis. The scientist is confronted by a definite situation. He observes it for the purposes of taking in all the facts. These suggest to him a line of action. He constructs an hypothesis. Upon this he acts and the practical outcome of this procedure refutes, confirms, or modifies his theory. Between theory and fact, his mind flies like a shuttle and theory is helpful and important just to the degree to which it enables him to understand, relate, and control phenomena."

This is essentially the technique of research. How, you might ask, is it relevant to bedside practice?

The physician, too, is confronted by a definite situation. He or she must seize its myriad details, and only the power of observation, trained in actual experience, will make that possible. The patient's history, conditions, and symptoms form the observational data. From these the working hypothesis, now called the first impression or initial diagnosis, is framed. It requires a line of action, in part to test if it is right or wrong. Have the necessary facts been collected? Does the working hypothesis properly put them together and account for them? The professional competency of the physician is proportional to the ability to observe and heed the signs that nature makes in response to his or her ministrations. The logical processes of research in science and the intelligent practice of medicine use, therefore, exactly the same techniques.

In this post-Flexnerian era, we can extend his articulate descriptions to the processes of interventions or treatment. First, we must develop a hypothesis; therefore, we need knowledge—the ability to

acquire and use a store of information pertinent to particular diseases or illnesses. Whether from textbooks, journals, computer-based information, lectures, peer consultation, or conferences, we must insist that in the first instance we have appropriate knowledge.

Without an understanding of pathophysiology—the body of knowledge comprising the medically related basic sciences—we forfeit our opportunity to understand the cell, organ, and system functions that create the clinical presentation and direct effective therapy. The bedside and the conference room do and should continue to stress pathophysiology.

Second, we need the ability to apply that information in the clinical setting of the problem as it occurs in a specific individual. Clinical organization is the ability to systematize an individual's clinical data—the history, physical examination, and appropriate laboratory data—and related knowledge in a medically logical and purposeful manner. Organization is followed by synthesis, which is a process of combining all the relevant pieces of information and constructing from them an integrated hypothetical concept, in this case, an initial diagnosis. Synthesis includes recall of solutions to similar problems that occurred in past experience. We keep adding both useful knowledge and the results of scientific research to our memory banks and draw on them when we need them to form a synthesis.

Third, we employ clinical judgment, which is the ability to make clinical decisions about interventions for a specific person. Clinical judgment is the decision point for employing procedures and formulating details of treatment. It is done within a frame of reference based on probability and solid understanding of the clinical problem.

The problem-solving approach in medicine is based on a clinically logical process that embodies (1) the essential features of the intellectual challenge in developing a hypothesis and formulating therapy, (2) the diagnostic studies and the technologies we use that help confirm our diagnosis or force us to consider other possibilities, and (3) the treatment that can then be meaningfully instituted for the individuals.

Other Attributes of Medical Practice

Clearly, medicine has additional important attributes central to its practice. These include the acceptance of responsibility, empathy for the patient, interpersonal skills of communication, and motor and technical skills. Exemplary humanistic, moral, and ethical behavior of a physician should be the norm for all physicians providing public service.

The physician's primary concern and responsibility is the welfare of the individual patient. The physician provides comprehensive and con-

tinuing care; he or she is sensitive to the patient as an individual and to the emotional, cultural, social, and physical needs of the patient. A doctor must be willing to consult with other physicians when his or her own skills are limited and to make appropriate use of their advice. It is irresponsible to undertake management of a patient or diagnostic or therapeutic process for which one is not qualified.

Accompanying the scientific approach are elements of uncertainty and unpredictability. Medicine is not yet a complete science with a set of facts and interventions that unfailingly predict an outcome. The unpredictable and unexpected happen all too frequently; we have not conquered nature. The intellectually honest practitioner will always say when appropriate, "I do not know, but I will attempt to help." The bottom line of the relationship between doctor and patient is that the physician is accountable—primarily to the patient and, frequently, to the patient's spouse and family.

The individual patient and the physician have become associated with a third party—the payer for services. This association leads us to perform in a world complicated by economics. Indeed, health care has become so technologically intensive and costly that it accounts for more than 10 percent of our gross national product. It has also become unusually rewarding financially to some individuals and corporations. The advances in technology have improved our capabilities to diagnose, cure, and care, but the growth of technology has also changed the nature of health care and made it into an industry in which we physicians are only a major component.

In applying the industrial model to our professional activities, economists and health care planners designate physicians as elements in a complex organization that includes providers, managers, payers, and investors. The implications of the term "provider" are difficult for us to accept. The payers (be they individuals, the government, insurance companies, or employers) are concerned about expenses, investors are worried about their investments, and managers about both. We "providers" are told that our country cannot afford the care, that resources are scarce, and that people have different wants in addition to health care. That is, the public may have alternative uses for these resources. The question remains, how much does society want to pay for health care?

BALANCING EXPENDITURES AND QUALITY CARE

As a major industrialized nation in the world, the United States is struggling with the problem of allocation of its resources. Among these resources is the capacity to deliver the best quality health care. The ques-

tion of who should receive care is no longer an issue; access to health care is accepted as a right by all who reside in the United States. The question of how and how much to pay for complex and expensive services is a subject of strong debate, with government and employers having a major impact on medical care cost decisions.

The changes taking place in payment for services interfere with the physician's determination to fulfill his or her responsibility as a professional. Can we continue to provide society with high-quality care with these limitations? Finding methods to control the cost of care and to limit the expenditure of resources while assuring quality of care is a major objective of all the groups organized to deal with this difficult and elusive question. More germane and of immediate importance is what kind of quality are we providing now?

TECHNOLOGICAL ADVANCE AND QUALITY OF CARE

We live in an era in which rapid advances in technology play a major role in our daily activities. Very little that we experience is not dependent on some form of technology, including the practice of medicine. New techniques and tools have been and are vital to discoveries made in biomedical science and to development of the sophisticated equipment we use in the care of patients. Most of the progress in biology and medicine is initiated by first defining a need and then stimulating the development of new technologies or modifying and revising existing ones to fill this need. Rapid advances in medical technology have resulted in patterns of diagnosis and treatment that are powerful and effective, yet costly. Nevertheless, safety, efficiency, and clinical indications for appropriate use are not always adequately evaluated before or after a technology's introduction into practice.

For example, the need for a cardiac pacemaker was evident years before it was introduced. This instrument, based on physiology and pathophysiology of the rate and rhythm of the heart, is often lifesaving. Thirty-two years elapsed between the conceptual formulation of the implanted electric pacemaker and the first successful implant in 1960. Dramatic improvements in cardiac surgery played a major role in the success of pacemaker implantation. Much progress has been made since 1960, and we now have pacemakers that provide electrical impulses that come closer and closer to replacing or substituting successfully for the normal pacemaking stimuli in the heart.

Quality-of-Care Issues

No one will question the lifesaving benefits of the pacemaker; its safety and efficiency are established; pacemaker sales approach 100,000

annually. The cost to patients, third parties, and government, however, is great, and that cost brings up questions about overuse. Assessors ask if the pacemaker is being applied in a skillful and balanced manner. How certain are we physicians and surgeons that the patient's needs have been carefully defined and met when we decide to implant the instrument? Who will define quality of care in this instance? How will the standards for quality be formulated, and how will use be monitored? It used to be the responsibility of the profession to establish and define these standards. Now, physicians must be accountable not only to patients but to payers as well.

The pacemaker is a relatively straightforward example of the issues inherent in assessing quality, efficiency, and cost of technology; the entire assessment discipline is rather more complex. Technology assessment and quality-of-care issues have only recently begun to be aired. The term "technology assessment" first appeared in the *Index Medicus* in 1980. Since then, many procedures have been and are being developed to measure the safety, efficiency, cost, and cost-benefit ratio of old and new technologies.

Technology Assessment Issues

Technology has been accurately described as a moving target. Diseases change in their incidence, new technologies are introduced, and new discoveries stimulate further advances in machines and instruments. Timeliness in decision making frequently becomes a problem.

For instance, the introduction of machines and instruments may be delayed by the need to accumulate enough data and by the mandated decision-making procedures of government regulatory agencies. More important, however, the withdrawal of faulty techniques may be delayed by the lack of prompt reporting of adverse data. Gastric freezing for treating duodenal ulcers, radiation treatment to the thymus of infants, diethylstilbestrol treatment for miscarriage, and oxygen treatment of premature infants are examples of errors that could have been corrected more promptly than our open system permitted. Perhaps some cancers and blindness could have been avoided, but this is a statement made by onlookers rather than by those using innovations in the front lines. Nonetheless, I believe that with a more organized approach to a technology assessment system than is being formulated at the present time, we can expect to do a better job.

Primacy of the Quality-of-Care Dimension to Technology Assessment

The quality of medical care delivered by physicians, nurses, and

institutions is an issue that we have been struggling with for a very long time. To me, the moving target here is the changing of disease patterns and, related to this, changing methods of diagnosis and treatment, particularly of the moderately and severely ill. Methods of diagnosis, procedures, drugs, and all manner of intervention change with time. Acquired immune deficiency syndrome (AIDS) is an example of a disease unknown just a few years ago. Although its etiology and molecular biology have been rapidly described and it can be diagnosed with some surety, its treatment and prevention are still under close scrutiny.

The current urgent focus on quality is partly a response to the prospective payment system (PPS), a system designed to contain expenditures on hospital care. To paraphrase the sentiments of a leading hospital administrator: with Medicare's PPS, the hospital's role as a production system is enhanced; all hospitals must learn to balance the new economic realities as they and their medical staffs adapt to a changed future. Hospitals and physicians are indeed adjusting their practices, but whether for good or harm remains to be seen.

The implementation of PPS resulted, for instance, in shorter stays for hospitalized patients. That the quality issue would become supreme was predictable, as patients were discharged earlier in their hospital stay. The accusation that hospitals were discharging patients "sicker and quicker" was heard in Congress as well as in almost every hospital and doctor's office.

The Health Care Financing Administration (HCFA) has declared that quality measurement is the issue of the late 1980s. Although quality has been important to the medical profession for a long time, it is appropriate that the primacy of the issue be emphasized by planners and government officials who are responsible for the changing economic environment.

HISTORICAL PERSPECTIVES ON QUALITY ASSESSMENT

Attempts to deal with quality began soon after the birth of modern surgery. Dr. Ernest Codman, a surgeon at the Massachusetts General Hospital (MGH) in the first decade of this century, was a strong advocate of accountability, efficiency, competition, and of advertising statistics of performance. He proposed an "end result card" to describe a follow-up examination one year after surgery or discharge from the hospital. Dr. Codman wanted to know, for example, what was the patient's problem? Did the patient get entirely well? If not, why not? Was it the fault of the surgeon, the disease, or the patient? What can be done to prevent similar failures in the future? Dr. Codman was using outcome measurements.

Clearly, he was ahead of his time. His strong advocacy for quality

did not meet with favor. He wanted to link the assessment of errors to promotions in hospitals and medical school faculties. He proposed that patients be made aware of the data that would help them to select a doctor and to choose a particular hospital for their medical care. When he could not convince his colleagues and superiors at MGH, he resigned and established his own surgical hospital. Dr. Codman died in 1940, having realized that "the patients and the public do not yet understand the problem of professional accountability."

Mortality Rates as a Quality Measure

Forty-six years after Codman's death, HCFA began publishing mortality rates for hospitals with emphasis on hospitals that were judged to have abnormal mortality rates. The California peer review organization (PRO) also released hospital-specific mortality rates. Although some consumer groups applauded the revelations, hospitals objected strenuously. They insisted that a high mortality rate does not mean a hospital provides poor-quality care: Raw unadjusted data could not and did not take into account the types of patients treated, whether patients had acute or chronic conditions, the severity of illness, the risk of treating different age groups, and the demands of the patients.

Stimulated by release of these data, headlines such as the following appeared in newspapers around the country: "Twenty-three area hospitals hit for high death rates." "Lists of hospitals that are above and below average United States mortality rates." Advertisements appeared from hospitals with low mortality rates. Hospital talk concentrated on the interpretation and critical analysis of the lists and data. Physicians, surgeons, and administrators were anxious about other lists being developed and appearing in the press. Concerns were heard that hospitals might be reluctant to admit severely ill patients if unadjusted mortality statistics were used to judge the quality of a hospital. All agreed that the methodology had to be improved and that data released had to be responsible.

Researchers have begun to define ways of adjusting death rates for severity of illness in hospital intensive care units (ICUs). APACHE II, developed by Dr. William Knaus and colleagues at George Washington University Medical School, provides one severity of illness score for this purpose; it is based on physiologic variables and coexisting chronic illness for each patient. With it, ICUs can be ranked according to how seriously ill their patients are and their mortality rates can be compared. An adverse score could alert hospitals to study the physician and nursing care in their ICU.

Such data are important in identifying suspicions of poor quality. In my opinion, however, we are unable to define just what levels of serv-

ice hospitals, doctors, and nurses must provide to qualify as quality care. I see no consensus on how quality of care should be measured for specific diseases or, even more important, for an entire institution.

The profession itself is trying to do its part. By requiring periodic self-study by hospitals, the Joint Commission on Accreditation of Healthcare Organizations is attempting to assemble data. Their requirements for accreditation now (or will) include, among other items, infection rates, internal transfer to the ICU, surgical complications, discrepancies in pathology reports, and drug interactions. In obstetrics, blood loss greater than 500 milliliters, fourth degree tears, and eclampsia will be documented.

Quality analysis, it seems to me, is in a developing and experimental phase. Hospitals, physicians, and nurses are the subjects of these experiments; the patients provide the data. What is not fair or prudent, during such experimental and developmental stages, is to release results or draw conclusions before all the data are in.

CONCLUSIONS

How do we effectively consolidate our responsibility and accountability to patients and to society? How can we follow the Flexner mandate to serve each patient and, at the same time, adhere to a social contract with society that today includes a major economic element? Our social contract requires that we serve both the individual patient's interest and what is called the common good.

Are physicians properly obliged to keep health care costs under control, as some advocate, even if doing so compromises the welfare of some individual patients? Does limiting medical care, thereby limiting costs, provide the greatest good for the greatest number? Does the public really wish to limit the availability of expensive medical care for the deserving individual? I do not think we can ethically answer these questions in the affirmative. Sick people want their physicians to be dedicated to diagnosing and healing them, not to be diverted or limited by a preset amount of dollars.

The ethical standards of the medical profession dictate that we care for the individual patient in the most efficient, effective, and high-quality manner available. While doing so, we should and do take an active role in containing costs without sacrificing good care. Critical to this is the assurance that our practices and procedures have demonstrated benefits for patients. We join forces with government and industry in adhering to the principle that the demonstration of benefits and the assurance of safety be accomplished by scientifically conducted evaluations. Equally critical is that we physicians possess the knowledge and skills that make

us capable of the practices we are performing. Our profession must continue to comply with high standards of self-credentialing and should continue to cooperate with government and other agencies in more effective and stringent self-evaluation and self-discipline.

Quality assurance for health care by physicians starts with a sophisticated process for admitting highly qualified individuals to medical school. Standards of knowledge and skill are high for graduation, licensure, residency training, and specialty certification. The public is well served by our system of education, training, and certification; their trust is well earned. Society also demands of its professionals high standards of moral responsibility and the avoidance of behavior that leads to the perception of greed, self-interest, and the assumption of power. We must avoid becoming a special interest group and avoid being accused of using ethical directives as a smoke screen for economic advantage.

We are currently experiencing a revolution in biology and medicine. New discoveries allow us to understand the normal and abnormal on a molecular level. The last decade of the 20th century and the early decades of the 21st will experience advances in diagnosis and treatment that will make today's practice of medicine obsolete. We can predict that cancer will be conquered by the knowledge of special genes that regulate growth, that understanding heart muscle function on a molecular level will benefit the sufferers of heart failure, and that ailments such as Alzheimer's disease will be treated and cured by the knowledge of genes and cellular products that deposit abnormal substances in the brain.

The future of medicine has never been brighter. Cure and prevention of many diseases are realistic promises of the future. We value our professionalism, and we are determined to resist all efforts to deprofessionalize medicine and depersonalize care. Medicine, we believe, is a calling requiring scientific and scholarly principles, high standards of achievement, and dedicated public service. Continuing to deliver high-quality health services will go a long way within that framework toward convincing the public that we serve the public interest and, indeed, serve it well.

Quality Assessment and Technology Assessment: Critical Linkages

Robert H. Brook

Why would anybody want to write a paper on the relationship between quality assessment and technology assessment? Being interested is not a good enough reason; neither is simply adding to the profusion of articles that seems to characterize these two fields today. If there is any justification for such a paper, it lies in the need to underscore two issues: both quality and technology assessments are vital to a successful health system, and each type of assessment depends on the other's success.

The scientific community behaves as if each of these activities can prosper on its own merits. They cannot. To make this point, to clarify the significant interrelations between quality and technology assessment, and to draw some policy-relevant implications for the future are sufficient justifications for undertaking this essay.

To facilitate the discussion of the relationship between these two types of assessments, this paper is organized in the following way. First, I briefly define quality and technology assessment and follow that with an overview of the evidence relating to the lack of integration of the fields. The final section presents some thoughts about why the current situation should change. The research studies I have cited were chosen to emphasize specific points, not to be comprehensive.

This paper takes an optimistic, not a pessimistic, view. Ample evidence suggests that quality of care can be assessed, that technologies can be assessed, and that both types of assessments can dramatically change the practice of medicine for the better. We need not be fearful of these assessments. The only things to fear are, first, to fail to use the assessment tools currently available to us and, second, because of a lack of information about quality and efficacy, to continue to practice medicine in a partial knowledge vacuum.

Without vigorously pursuing these assessments, however, we will

continue to make major errors in the use of technology and to produce a level of quality that is, at best, average and, at worst, far below what could be possible. The American people deserve better from a system that consumes 12 percent of the gross national product. American physicians might even feel better if they knew that their care was improving the health of their patients.

DEFINITIONS

Definitions of technology assessment and quality assessment are hard to come by. The Office of Technology Assessment defines technology assessment as a "comprehensive form of research that permits evaluation of technical, economic, and social consequences of the use of technology including short- and long-term, intended and unintended, and direct and indirect consequences"(1).

We might define quality assessment as the measurement of those aspects of personal health care that cause a patient's or community's health to deviate from its expected natural course. Quality of care consists of technical and interpersonal elements. If measured by structure (e.g., characteristics such as whether a physician is board certified) or process (e.g., procedures such as performing an electrocardiogram on patients who suffer chest pain), only those elements known to relate to outcome can be used. If measured by outcome, the study design must be sufficient to link outcome differences to process or structure (that is, aspects of the personal health care system) rather than to patient characteristics or disease severity.

To understand quality and technology assessment better, two other words must be defined: efficacy and effectiveness (2). Efficacy is the benefit achieved from a medical technology when it is applied to a given medical condition under ideal conditions of use (by the best practitioners in the best hospitals). Effectiveness, while having all the attributes of efficacy, reflects performance under ordinary conditions by the average practitioner for the typical patient. Differences between efficacy and effectiveness could be produced by random error (chance), bias (the inventor of a technology either consciously or subconsciously wants to make the technology appear better than it actually is), quality of care, or factors that may or may not be under the control of the personal health care system, such as patient or system noncompliance.

CURRENT SYSTEM

The above definitions suggest that people who perform quality and technology assessment should be in constant contact; in fact, two sepa-

rate fields have emerged with separate languages, conceptual constructs, methods, practitioners, and meetings. For example, the Institute of Medicine recently published a 573-page book on *Assessing Medical Technologies* (3). The word "quality" did not appear in the index; it may well not have been in the text. A possible synonym for quality was medical practice, but even so, the authors evidently made no direct attempt to link medical technology assessment with the quality assessment movement.

What are the links that should be forged? One major question is whether and how the outputs of technology assessment can be made "user-friendly" so that people who assess quality of care can benefit from them. Much of the technology assessment literature tends to be published in fragmented form or sometimes in "obscure" journals or "fugitive" literature; or, it is presented at meetings that consist mostly of experts in technology assessment. As long as this is so, making the products of technology assessment of use to the quality-of-care community will be difficult.

Those questions assume the issue is primarily one of communication. But we should ask: Is the problem even deeper? Are people who primarily do technology assessment reluctant to be associated with those who perform mainly in the quality-of-care area? Since quality-of-care assessment has an evaluative, judgmental tone about it, perhaps being associated with or forming an alliance with the people who study quality is too threatening to happen naturally.

The willingness of people assessing quality of care to pay attention to technology assessment results is also less than optimal. People in quality assessment have their own jargon, their own journals, and their own meetings. They also must deal with one overriding concern: Should we develop valid and reliable sets of national criteria and practice standards?

By intent or by default, the country appears to be slowly moving toward "national" standards, but because this activity is occurring in a passive manner, it is happening without full use of technology assessment information. For that reason, we may find ourselves with invalid criteria. For instance, let us consider the process of setting criteria for optimal quality of care for adults who seek care for a sore throat. If the criteria are to be valid, the following aspects of care must be addressed:

- Which patients should be asked for a medical history? Given a physical examination?
- What do we mean by an adequate history or physical examination in these circumstances?
- Who requires a throat culture? What do we do about a positive culture?

- Who gets an antibiotic? On what evidence? Which antibiotic?
- What is adequate follow-up care?

To provide at least partial answers to such questions, the technology assessment literature could be analyzed and synthesized.

But who has the resources to do that, especially if that literature is fragmented and difficult to find? If a regulatory agency wants criteria in a very short time, who can produce reliable sets? If the agency wants comprehensive, valid criteria, who has the time? When quality-of-care criteria for treatment of a sore throat are produced by many physician groups or hospitals working independently (especially "over lunch when the quality committee meets"), we can be sure that the outputs of the technology assessment process will not be adequately used.

What I suspect will occur is what happened 20 years ago when I approached some infectious disease experts regarding setting quality-of-care criteria for patients with a sore throat. The experts suggested the following: All adults with a sore throat should, at their first visit, have a complete history and physical and a throat culture. If the culture is positive for a strep infection, a second visit is necessary with the patient receiving a short-acting penicillin injection. The patient should be observed in the office for 30 minutes to make sure an allergic reaction does not occur. The patient returns for a third visit the next day, at which time a history of penicillin allergy is taken and if one is not detected, an injection of a long-acting penicillin is given. If the culture is negative, the patient returns for a second visit only if he or she is symptomatic.

Given the high prevalence of sore throat, the economic consequences of applying such a criteria set for that complaint would be exorbitant. Yet, 20 years later, and after many technology assessments in this area, we still do not have a sophisticated, valid, and reliable set of criteria to judge the quality of care given for such a mundane problem as a sore throat in adults—certainly not one that can be fully supported by conclusions from those technology assessment studies. If the quality and technology assessment people had found ways over the years to communicate better with each other, perhaps the outcome would have been different.

INTERDEPENDENCE OF QUALITY AND TECHNOLOGY ASSESSMENT

The definition of technology assessment quoted previously does not address whether technologies should be assessed under conditions of efficacy, effectiveness, or both. Are we interested in the value of the technology when it is performed under ideal conditions (the best physician, in

the best hospital) or under average conditions? Sensitivity analysis may be useless if the analyst does not or cannot know the conditions under which the procedure will be applied in the real world.

Thus, one link that must be forged is that quality-of-care-related information should be made part of the context within which technology assessments are conducted. For example, the two options by which early cancer of the cervix can be treated are radiation and surgery. Both have equal cure rates, but under ideal conditions the complication rate from radiation therapy is lower than that from surgery. Hence, the radiation approach would appear to produce better overall outcomes; based on technology assessment studies, radiation therapy has been the recommended course of action. However, a national quality-of-care study of the treatment practices of cancer of the cervix found an 11 percent major complication rate from the use of radiation therapy (4). It also showed that the rate of complications correlated with how well the radiation therapy was applied. The results of this study greatly weakened the argument against surgery for stage 1 cancer of the cervix. Had there been a feedback loop from quality assessment to technology assessment, this finding might have emerged more routinely; now-questionable assumptions about the preferable therapy would not have been produced by an accident of faith.

Technology assessments, especially those relating to the use of drugs for cancer patients, often are based on controlled clinical trials at multiple centers. The quality with which the protocols are applied is rarely known, but it can play a key role in the trial's outcomes. One such study on physician compliance with a study protocol for Hodgkin's disease was done by the Southeastern Cancer Study Group (5). The study, based on 938 treatment visits, concluded

- Physicians complied with protocol treatment decisions 64 percent of the time.
- Of the errors, 29 percent related to the timing of the use of toxic anticancer medications, 28 percent to the level of dosage, and 43 percent to the way the dose was calculated.
- Compliance with the protocol resulted in substantially better outcomes, measured by whether blood counts remained acceptable.
- Compliance could be increased to 94 percent if a systematic information system was used.

If these results can be generalized to other efforts in technology assessment, we may ask: How many technologies (in this case, drugs) may have been discarded because of failure to recognize that they were effica-

cious, that failure arising because the quality of care with which they were applied was so low that even the best sensitivity analysis missed the mark?

The situation becomes even more complex if patient values and choices are considered (6). Two therapies are used for lung cancer—radiation therapy and surgery. The five-year survival rate for surgery is better than that for radiation therapy; unlike radiation therapy, however, surgery is associated with the possibility of immediate death. Some people might choose a shorter life, on average, while being certain that they would not die perioperatively. This tradeoff might vary according to the quality of surgical care—i.e., exactly how likely is it that the outcome will be perioperative death? In fact, McNeil's research found that, when 60-year-old patients with lung cancer were asked which therapy they preferred, 7 percent replied that they would choose radiation therapy if the operative death rate were 5 percent (high quality) but 64 percent said that they would choose radiation therapy if the operative death rate were 20 percent (low quality of care).

The quality assessment field is similarly confused. Consider, again, a patient with a sore throat who seeks care from a physician. Let us assume several things: First, the doctor can order one or all of three tests, each of which costs $10; second, all tests will result in more rapid improvement in the patient's health; third, a patient who receives tests A and B recovers more quickly than one who gets A and C; and, fourth, the person who receives A, B, and C recovers quickest of all. How do we compare the quality of care received by the patient who got A and B versus A and C, or versus A, B, and C? The health outcome is better for the first patient than for the second, and the cost of care is the same. The health of the third patient is better than that of the first, but the cost is higher.

Unfortunately, because cost and health outcomes are rarely measured in the same study, the word quality is used by the field to compare such illustrative cases with each other. Thus, in some comparisons quality means obtaining a better outcome with similar resources. In other cases it means obtaining a better outcome but with additional resource use. The consequences of these different uses of the word quality are not trivial, as they confuse the uninitiated or make it possible to defend the status quo when the real need is for honest evaluation and change.

In summary, we need in the future to study the epidemiology of quality of care and to analyze in detail the output of technology assessments, and then we must make the findings of each enterprise available to the other. To underscore these recommendations, I have tried to demonstrate, through some examples, the following:

- The quality and technology assessment fields perceive themselves to be different; they are, however, similar and integrally linked.

- Both quality and technology assessments require, and depend on, a broad definition of health.
- Both fields make evaluative judgments.
- For constructive action to occur, both fields must establish causal links between the process of care and patient outcomes.
- Measurement of cost is not sufficient to separate the fields.

For the outputs of both these fields to have maximum impact on the health of the American people, I would make the following additional points:

First, cross-fertilization must occur. Technology assessments should explicitly include a statement about the level of quality, and quality assessments should use the outputs of technology assessments.

Second, better ways to communicate and share results must be found. Descriptions of studies, reporting of findings, and technical details must be made "user-friendly" to both the quality and technology assessment communities.

Third, both fields must coordinate the selection of assessment topics and decide whether criteria should include obtaining the biggest health payoff for a given level of expenditures, the most savings consistent with a minimum level of outcome, or equity.

Finally, a centralized facility or mechanism should be established to analyze, synthesize, and disseminate the results of technology assessment; to develop, update, test, and store quality-of-care criteria and standards; and to integrate information from those two fields. Absent a single central repository, we can hope that the Council on Health Care Technology will fulfill the first goal. An equivalent "Council on Quality Assessment" might be established to achieve the second goal, and both can cooperate to reach the third.

The final system might operate in the following way: A technology assessment is performed. Its conclusions are then integrated into data banks containing information from previous technology assessments. This step most likely requires additional quantitative analyses. Once these are completed, the new, augmented information replaces the old in models that are used to generate quality-of-care criteria or standards. The results of this effort are presented to a series of experts for their review, revision, and approval. Finally, the ensuing quality-of-care standards and criteria, along with the analysis supporting their use, are made available to organizations or agencies charged with the business of quality assessment.

As the process matures, the outputs of such models might serve as guides to good medical practice. They could be maintained on-line and be accessible by physicians and patients by a home or business computer.

Physicians, no longer dependent on assessing new scientific information themselves (which they do poorly and have little desire to do), would use their on-line links to computer data banks in making clinical decisions—i.e., in charting courses of action to follow in the case of individual patients in an increasingly complex medical world.

A new form of practice might emerge—one that reflects a more active patient, a more satisfied physician, and, ultimately, better health for all at a price we would be willing to pay.

REFERENCES

1. Office of Technology Assessment. Congress of the United States. Strategies for Medical Technology Assessment. September 1982. Washington, D.C.
2. Brook, R.H., and Lohr, K.N. Efficacy, effectiveness, variations, and quality: boundary-crossing research. Medical Care 23:710–723, 1985.
3. Institute of Medicine. Committee for Evaluating Medical Technologies in Clinical Use. Assessing Medical Technologies. Washington, D.C., 1985.
4. Hanks, G.E., Herring, D.F., and Kramer, S. Care outcome studies: Results of the national practice in cancer of the cervix. Cancer 51:959–967, 1983.
5. Wirtschafter, D.D., Scalise, M., Henke, C., et al. Do information systems improve the quality of clinical research? Results of a randomized trial in a cooperative multi-institutional cancer group. Comput-Biomed Research 14:78–90, 1981.
6. McNeil, B.J., Pauker, S.G., Sox, H.C., Jr., and Tversky, A. On the elicitation of preferences for alternative therapies. Preferences for Alternative Therapies 306(21):1259–1262, 1982.

Measuring and Maintaining Quality in a Health Maintenance Organization

Donald M. Berwick

This paper includes a brief review of the current techniques for quality measurement, an analysis of future needs, and a glimpse at a specific program for quality measurement that my colleagues and I have developed at the Harvard Community Health Plan (HCHP), a health maintenance organization (HMO) of some 300,000 members in eastern Massachusetts. I will begin, however, with a few cautionary remarks.

First, although my experience at HCHP is extensive, I am not entirely sure that the lessons learned there apply widely outside the so-called "managed care sector" of the health care industry. Indeed, my local experience may not be generalizable to HMOs other than the staff model that predominates within HCHP.

Second, the topic of this conference is the intersection of quality assurance and technology assessment. Although technology assessment is a research interest of mine, I prefer to discuss quality assessment and leave it to others to make the necessary connections between the two fields.

THE DEMAND FOR INFORMATION ON QUALITY HEALTH CARE

Hardly a single health care conference passes lately without a session on quality. All parties to the health care contract, patients, purchasers, regulators, and care-givers, seem concerned rather suddenly that we are somehow losing our collective grip on the quality of our care systems (1). Undeniably, the major impetus for that concern is the pressure of cost containment and its first cousin, competition for patients. Prospective payment, caps on expenditure, and price competition are forcing health

care producers to trim programs and question expenditures. When we make efforts to tighten the system, it is natural to seek assurances that we are not thereby imperiling the patient or violating our values. But cost pressure is not the only reason to measure, monitor, and improve quality in health care today. At least six other forces make quality measurement an urgently felt need.

First, we appear to be in a general transition in health care from a professional-dominant to a purchaser-dominant industry. The transition is by no means complete, but it is further along than the average physician in this country may realize, and it shows no signs of slowing. The change in structure is accompanied by a change in the need for information. Professionals do not need explicit mechanisms to demonstrate the quality of their work to others. Indeed, one defining characteristic of a profession is that, as a work group, it reserves to itself the authority to judge the quality of what it does. In a purchaser-dominant system, on the other hand, the purchaser has (or shares) that authority. To make choices, the purchaser must have some way to assess the quality of the product being purchased. A purchaser-dominant system requires more explicit measurement of quality than does a professional-dominant system.

The increasing authority of the purchaser also creates a climate for the market to segment itself more fully. Different consumers of health care may prefer health care with different attributes, and the notion of quality itself acquires elements of taste; its definition may no longer be uniform. A segmented health care market requires information on various dimensions of quality, dimensions that may be weighted differently by various consumers.

A second source of momentum toward quality measurement is the accumulating evidence that physicians vary greatly in the way they use health care resources to care for patients with similar clinical conditions (2). Variation in the use of tests, drugs, procedures, and hospitals is so wide that it strains credibility to argue that all physicians are practicing medicine of equally high quality. Can doctors who differ by fivefold in the rates of use of a procedure all be giving correct care? The evidence on variation creates, in the minds of both the producer and the purchaser of care, a need for more comprehensive information on the underlying quality of practice.

Third, the increase in malpractice litigation has given new urgency to the goal of controlling quality. Some hold hope that better quality control may reduce the burden of liability, although that hope is by no means grounded in empirical evidence yet.

Fourth, we face oversupply of doctors and hospital beds in many health care markets. This excess creates both a threat of unnecessary practices to keep the providers busy and an opportunity to select the best

of those available, if only information were available to guide that selection.

Fifth, health care is entering a new age of data management, benefiting from sophisticated computer systems that support the collection and analysis of large bodies of data. My own HMO has a computerized medical records system, and data bases of insurance claims show increasing promise as monitoring tools. Sometimes, quality measurement involves searching for patterns or infrequent events requiring large aggregates of information. The new data systems create technologic opportunities, not previously available, for collecting and manipulating information.

Finally, quality measurement receives new life today because of the very existence of managed care systems. In a cottage industry configured by happenstance and few central controls, information on quality lacks linkage to corrective action. Managed care systems, such as HMOs, create the opportunity for self-conscious designs and broad response to information on quality. Better managed health care systems use information with a directness not possible in a more disaggregated industry.

CURRENT TOOLS

The need is great for tools to measure quality. What is available today to fill that need? The answer is not reassuring. To be sure, quality assurance has had a distinguished intellectual tradition in health care for at least three decades. The research literature contains important themes of investigation on both the targets of quality measurement—what is to be measured—and the methods for measuring—how to measure.

I will not review that literature but will instead put before you one key summative conclusion: Despite a deep intellectual tradition of investigating quality in health care for the past thirty years, we almost completely lack a useful applied technology for the measurement of quality in health care. By an applied technology I mean a set of measurement tools and supporting theory that can be used by the decision makers who guide the expenditures, programs, or organizational structures in health care institutions. An applied technology for health care quality measurement should be connected to action through the choices of those who make decisions.

What are the characteristics of such an applied technology, and where do current tools fall short? At least five characteristics would seem desirable.

First, useful measurement tools would produce timely information. An appropriate applied technology for measuring quality must produce information at a rate that matches the needs of real decision makers over a span of weeks or months, not the years that it generally takes for research-oriented quality assurance to generate conclusions.

Second, useful quality measurement must be able to aggregate information at a level appropriate to the decision maker. The chief of a department needs information about the department. The research literature generally offers views at levels of aggregation far too broad for real-world action.

Third, measurement tools must be sufficiently low in cost. It is not clear how much money will be available to measure health care quality, but it may be revealing that HCHP, which at present is probably investing more dollars in this field than any other health care organization of any size, is spending barely one-third of one percent of its operating budget on a quality measurement program. Useful quality measurement in today's climate must be far more efficient than research models have been to date.

Fourth, the applied technology for quality measurement of the future must attend carefully to methods for displaying its results. Using statistical control charts was a major breakthrough in industrial quality control; standard forms of display made measurement results accessible to others who had too little time or skill to seek patterns in less convenient tables of data. It may be symptomatic of our inattention to the issue of display that few health care managers appear at all familiar with even the seminal investigations in academic health care quality assurance.

The fifth feature of an applied technology for measurement is that it be multidimensional. Many persons believe that quality in health care is, at bottom, connected unidimensionally to health status outcome. According to this orthodoxy, a high-quality health care system is, by definition, one that yields the best health outcome for a given investment of resources. This view seems flawed to me. A more robust view of quality would incorporate more than outcomes (narrowly defined) in the definition. We seek to produce and to purchase more than longevity and function when we see the doctor. Imagine, for example, two hospitals identical in the initial functional status of their patients, and in which results in terms of longevity, pain, and return to work are identical so far as we could possibly measure. Imagine, however, that in one of the hospitals the patients feel cared about, but that in the other the patients feel the staff to be cold and uncaring. Are these hospitals of different quality? They are, and a useful set of tools for assessing quality must take dimensions other than outcome into account. It is possible to create a health care system

with good health status outcomes but with poor quality on other significant dimensions.

LESSONS FROM OTHER INDUSTRIES

In June 1985 I was asked by the board of directors of the HCHP to assume a newly created position of vice president for quality of care measurement (QCM) and to set up a department that would report, one year later and regularly thereafter, on the quality of the HMO. Concerned about cost pressures, the board sought a system whereby it and the HMO managers could adjust their decisions so as to preserve and enhance the quality of care.

As I have already suggested, I found serious limitations in the health services research literature to guide the discharge of this new responsibility, certainly in the required time frame. In industries outside health care I found quite a different story. In visits with airlines, hotel chains, consumer goods manufacturers, AT&T Bell Laboratories, the National Aeronautics and Space Administration, and elsewhere, I became acquainted with theories and technologies that I call, collectively, industrial quality control.

In contrast to health care quality assurance, industrial quality control is very much an applied science (3,4). With roots in statistical theory and with standard tools dating as far back as the 1930s, quality control engineers in fields other than health bring to their task a discipline equipped to deal with the pressures of time and practicality that managers and workers face. It is undeniable that health care has its own special needs in quality assessment, but, with equal certainty, we have much to learn from other industries.

A full exploration of such importable lessons is beyond the scope of this report, but a few specific examples are useful.

First, successful industries have, in many cases, learned that high quality depends on continual improvement. Quality measurement is best guided by a firm conviction that one can be better tomorrow than today. Specifically, an institution should not attempt to prove that it is good enough now and need not improve. Quality control is grounded in aspiration, not in defense.

Second, the search for quality relies on a clear sense of values and constancy of purpose. Perhaps values are not clear enough today in health care. In particular, new structures of health care should look to establishing goals that are negotiated with the clients so that the definition of quality will be a shared one, not dictated by a profession only.

Third, industrial quality control experts know that quality

improvement cannot be based on fear. Too much fear exists among today's health care providers who are faced with a demand that quality be studied, displayed, and improved. Whereas some industries (or at least some companies) have embraced quality control as a set of tools through which to do better, much of health care has reacted almost purely in defense, figuring out how to avoid quality measurement instead of how to do it better. In Japan it is said "Every defect is a treasure." The discovery of a problem is recognized and valued as an opportunity to improve, not as a threat to one's security.

Fourth, in industry, quality control is grounded in statistical theory. Measurement is done to acquire knowledge, and knowledge is acquired so that action may occur (4). Medical training alone does not equip health care professionals for the sorts of inferential tasks that allow knowledge to guide the improvement of quality; the statistical principles used in industrial settings are not arcane, but neither are they intuitively obvious.

Fifth, improvements in productivity, some industries know, follow from improvements in quality (5). We are doing it backward in U.S. health care today. The budget mechanisms and the cost containment apparatus being put into place in hospitals and managed care systems emphasize control of cost and productivity as the route to an improved system. By contrast, what some U.S. industries have learned, and what I think U.S. health care may discover, is that if quality is improved first, productivity improvement will follow; but it will follow in a somewhat longer time frame than we are currently using to make policy about cost containment in this country.

Sixth, quality problems in industrial settings tend to be systems problems. The opportunities for improvement of quality lie in improved design of systems, and I believe that this is also true of health care systems. If we try to improve U.S. health care quality simply by eliminating the lowest stratum of producers, we will be repeating a mistake that failed most industries during the past two decades.

Finally, industrial quality control engineers know that inspection of the final product is an extremely inefficient way to improve quality. No one ever inspected quality into anything. Quality is improved by improving the production process.

THE HCHP EXPERIENCE

Guided by these and other lessons from outside health care, and informed by the research traditions of quality assurance, my staff and I began to develop some managerially useful tools in our HMO. Our strat-

egy was guided by the need to embed measurement in relation to values on the one hand and in relation to action (in the form of control systems) on the other. We seek measurement for action.

Although our budget was generous, it seemed prudent to focus measurement in areas where hazard was most likely to lie. The accumulated experience in our HMO suggested that if our values were to be offended in caring for a patient, it would likely be due to one of four kinds of flaw: denying access, failures in communication, blurring lines of responsibility, or deficiency in supervising or training of nonmedical staff.

These may not be hazardous areas in other forms of health care, but in a prepaid group practice they are not at all surprising. Access is an important tool for rationing care under prepayment; indeed, downward adjustment of staffing ratios is becoming as important a mechanism for cost control in ambulatory care as diagnosis-related groups (DRGs) have been for hospitals. Control of access is a blunt tool and can be hazardous to health.

Failure of communication can occur, especially in large practices, when patients are transferred between clinicians or institutions. For all its ills, fee-for-service medicine probably reinforces communication, at least among physicians. In complex health care systems, communication is essential for the safety and comfort of the patient, yet the technology of communication, such as ways to report laboratory results or to connect specialists with each other, has lagged.

As communication suffers in large systems, so can responsibility for the patient become blurred. Prepaid environments can lose some of the informal mechanisms that the incentives in fee-for-service care preserve. In our HMO we often find unclear responsibility close to the root of a quality problem.

Supervision of support staff may be an underestimated source of quality in health care systems. Operations research would reveal that a great proportion of the "care" given to patients comes from health care workers other than doctors or nurses. Receptionists and other staff can help determine whether a patient is safe and content or at risk for avoidable morbidity and discomfort.

To these four hazards in our managed care system, a fifth must be added, as it compounds the threat of the others; namely, the demoralized doctor. The current atmosphere of conflict, suspicion, and demand in health care cannot help but take its toll in the potential decathexis of work by the doctor. Although not yet directly relevant to the measurement of quality, I believe that quality in U.S. health care requires that the physician, no longer in control of the system, nonetheless be helped and encouraged to invest as much emotionally in the care of the patient as the

traditions of the profession have always claimed was part of the duty of doctoring.

The measurement strategy at HCHP is organized according to a framework that specifies a collection of dimensions, or attributes, of quality and investigates them as aspects of different encounters within the HMO. Types of encounters include, for example, well-care visits, symptomatic visits, emergency visits, hospitalizations, and visits to support departments such as radiology, the pharmacy, or the laboratory. It would be equally plausible to assess the quality of care for a segment of the population (for instance, the elderly) or for victims of a specific illness (for example, diabetes), but we chose the encounter as the unit of analysis because data are most conveniently available at that level. Ultimately, we will try to increase the amount of population-based measurement, since it seems that the HMO enrollee, as opposed to the user of care only, is a more suitable target of measurement.

Care at HCHP may be assessed according to any or all of eight attributes of quality. This list of eight is by no means exhaustive of plausible components in a multidimensional view of quality, but they provide, as a beginning, a rounded view of performance sufficiently detailed for some management action. An attribute of care is a property that a conscientious physician or reasonable patient would wish to have in health care, if performance on other attributes was held constant.

1. *Outcomes.* The orthodox view of quality requires outcome measurement. As many have found, outcome measurement is difficult and expensive, and we have begun to invest in prospective data bases as one appropriate system of support. At present, an obstetrical data base system and a tumor registry are fully operational, and a cardiology data base is being designed. Functional status assessment of some categories of patients will begin soon.

2. *Technical process.* Part of quality is the performance of the system in delivering the processes of care it intends to deliver. Thus, we established at HCHP specific criteria or algorithms for the care of well adults, diabetics, new elderly enrollees, people with major depressive disorders, and others, and we assess adherence to these specified care processes.

3. *Access.* We invest heavily in measuring the ease of access to scheduled appointments, telephone care, and other sites of entry, as well as facility waiting times. Because our experience has grown, we are able to compare performance to historical benchmarks, and we can spot problems in trajectory before they become actual hazards. For some access measures, we have drawn on industrial techniques of simulation and probes.

4. *Interpersonal sensitivity.* Through surveys using modern psycho-

metric principles, we assess and track patients' attitudes toward the care they receive. The assessments are dimensional themselves, involving not just global measures of satisfaction but also local ones of particular use to managers. In this and other realms, data are collected and analyzed at the level of the individual doctor or nurse and fed back to them for their personal use.

5. *Integration.* Mainly around issues of communication, we seek surrogate measures of integration of care. Measures include both direct and indirect assessment of communication among specialties and of the integrity of the medical record system.

6. *Physical environment.* Using tools borrowed from the hotel industry, we monitor and score more than 60 attributes of the physical space and environment of the HMO in eight dimensions, with subscale scores sent to facility managers and training department staff.

7. *Staff attitudes.* Through questionnaires, now in multiple generations, we assess the attitudes of professional and nonprofessional staff toward the support departments they rely on, toward the affiliated hospitals, and toward the HMO as a work environment. Information is shared with hospitals and has led to direct and effective action in several cases.

8. *Reliability.* Not truly a dimension of its own, the notion of "reliability" involves the simultaneous assessment of variability in performance along with average performance. In any dimension a reliable care system not only delivers high quality on the average but also rarely deviates from the average. Health care traditionally incorporates among its goals a strong aversion to risk, and a good health care provider sometimes will consciously incur a decrease in average expected performance to reduce the risk of deviation from that average. Our measurement systems seek to make such trade-offs explicit by reporting not just mean performance but also suitable measures of variance.

The combination of attributes and types of encounters yields a matrix of quality for which HCHP holds itself accountable. The display of measurement results follows this matrix, and it tries to focus attention on areas of especially high priority for repair or preservation. At the moment our displays are not very efficient; we mainly use tables and figures entirely too cluttered for rapid consumption. In the near future we hope to make greater use of conventional formats for display from industrial quality control, such as statistical control charts and other graphics, to enhance the connections between measurement and action.

CONCLUSIONS

Progress in measurement of quality is long overdue in health care. An applied technology is needed and achievable. But it also holds danger

if we believe that measurement, alone, can improve or preserve quality.

To improve quality, measurement must be linked directly to values, goals, and aspirations on the one hand, and to action, control systems, and change on the other. If I could lay out an agenda for the Institute of Medicine in health care quality, it would include these three cornerstones: clarifying values, enhancing design and process, and measurement. These three would be unified by the goal of continual improvement; they would derive their energy not from fear, defense, or reluctant remedies to a demanding environment, but rather from the firm intention to do the best we can, and always better, with and for our patients. In that latter pursuit good measurement serves us well.

ACKNOWLEDGMENT

The author is grateful to Ms. Judith Kalman for her assistance in preparation of this document.

REFERENCES

1. Ginsberg, E. The destabilization of health care. New England Journal of Medicine 315:757–760, 1986.
2. Wennberg, J.E., and Gittelsohn, A. Variations in medical care among small areas. Scientific American 246:120–134, 1982.
3. Wadsworth, H.M., Stephens, K.S., and Godfrey, A.B. Modern Methods for Quality Control and Improvement. New York, John Wiley and Sons, 1986.
4. Shewhart, W.A., and Deming, W.E. Statistical Methods from the Viewpoint of Quality Control. Washington, D.C., The Graduate School, U.S. Department of Agriculture, 1939.
5. Crosby, P.B. Quality Is Free: The Art of Making Quality Certain. New York, McGraw-Hill, 1979.

Quality Assurance in Hospitals in New York State

Thomas Q. Morris

The New York State Department of Health (NYSDOH) has assumed a major leadership role in addressing the issue of quality health care. Through a series of legislative and regulatory guidelines, including specific measures directed at identifying the reasons for the spiraling cost of medical malpractice, hospitals must gather and retain detailed information concerning both physician behavior and hospital practices. Specific reporting responsibilities for a broad array of clinical incidents have also been imposed on hospitals. This information, in combination with outcome or mortality data, is used by the NYSDOH to identify indicators of possible poor quality health care.

Hospitals, historically, have expected their own staffs to determine both appropriate levels and indicators of quality health care. At the present time, however, hospitals in New York State find themselves imposing considerable regulatory pressure on their professional staffs. This causes a significant amount of discomfort, both within the institutions and the medical community at large.

During the past 18 months, the NYSDOH has aggressively sought to assess quality of care through a variety of surveys, reporting requirements, and data analyses. The results of these efforts to date are mixed. Although a number of important observations and findings concerning presumed quality of care have been made, a distinctly adversarial atmosphere has developed between the providers of patient care—the physicians and the hospitals—and the NYSDOH. In addition, public confidence in several major academic and local institutions throughout the state has been significantly shaken.

To discuss these matters succinctly, however, it is important, first, to identify the origins of the recent emphasis on quality assurance in New York State; second, to describe the activities of the Hospital Association of New York State (HANYS) both in response to and independent of

these pressures; and, last, to point out the impact of the regulatory quality assurance system on hospitals in New York State.

ORIGINS OF EMPHASIS ON QUALITY ASSURANCE

Emphasis on quality assurance and quality of care in New York cannot be discussed without initially emphasizing the very distinct, direct, and personal involvement and commitment of Dr. David Axelrod, Commissioner of Health of New York State. He has given this the highest priority both within the department and within the state. He is directly involved in the central review process. In doing so, however, he is quick to point out, as he has publicly done, that hospital care in New York State is "better than ever." But he does see significant room for improvement. His zeal for improvement is highlighted, I believe, by comments he made before the 1987 annual meeting of the Greater New York Hospital Association. From his perspective, he noted that there were no complications in patient care; there were only incidents.

Legislative and Regulatory Initiatives

The renewed emphasis on quality assurance in New York State also has legislative and regulatory origins. In July 1985 a medical malpractice law was signed by Governor Mario Cuomo. The obvious intent was to prevent a patient care crisis created by physicians' refusal to see patients in many areas throughout the state because of the unavailability or inadequacy of malpractice insurance coverage. The law also obligated hospitals to initiate new monitoring and reporting mechanisms, which have had a major impact on hospitals.

Responding to the mandate that "every hospital" must maintain a coordinated program for identifying and preventing medical and dental malpractice, a series of regulations were promulgated in January 1986. These were sweeping in nature and necessitated the revision or replacement of existing quality assurance processes. Every hospital was directed to identify and implement a quality assurance risk management program to enhance the quality of patient care and reduce or prevent medical and/or dental malpractice. The new program integrated the review activities of all hospital services and mandated the following procedures:

1. Identify actual or potential problems in the care of patients through the following activities: review of all staff work; review of all mortalities and unimproved cases; review of morbidity and circum-

stances other than those related to the natural course of disease or illness; review of infections, complications, errors in diagnoses, transfusions, and results of treatments; and review of medical records, medical care evaluation studies, complaints, incidents, staff suggestions regarding patient care or safety, utilization review findings, profile analyses, and other pertinent data sources.

2. Assess the cause and scope of the problems or concerns identified.

3. Develop and recommend proposed courses of action to address and identify problems.

4. Implement, through established mechanisms, the action necessary to correct the identified problems.

5. Evaluate the effectiveness of the corrective actions taken in addressing the identified problems.

That is the first section of the regulations. This list continues for seven single-spaced pages. The sweeping nature of these guidelines is evident in many areas.

The procedures for the initial granting of privileges to physicians as well as for the continuation of privileges is based on a system with definite standards for the evaluation of each new or continuing applicant by the credentials committee of the hospitals. Medical staff appointments and reappointments and the delineation of privileges must be based on recommendations by the active medical staff to the governing body.

It is also mandated that the quality assurance committee for the hospital include trustees of the hospital who do not hold any other position in the hospital. The regulations also deal with specific physician activities in considerable detail, resulting in the development of what has been called the "physician profile." The hospital must maintain 16 items concerning the activity of each physician including the outcomes of morbidity and mortality review, blood utilization review, infection control review, utilization review, safety committee review, peer review organization (PRO) data, surgical case review, any medical care evaluations performed, tissue review, medical record review, incident report review, complaints, liability claims data, prescription review, and medical case review, plus evidence of a continuing education program and other training. All allegations of malpractice occurring in any setting, not just proven cases, must be chronicled for the past ten years.

Many of these requirements are not new; many have been in place for quite some time in various ways, but now the hospital administration is obligated to maintain these files. Physicians' concerns about access to these files, the confidentiality with which they are maintained, and the potential uses to which they might be put are, understandably, enormous.

Mortality Data

The second origin of the current emphasis on quality assurance resulted from a review of mortality or outcome data by the NYSDOH. For several years, New York State has required a variety of data to be submitted to the statewide planning and research cooperative system, the so-called SPARCS. These data come from all general hospitals throughout the state and include, among other information, diagnoses, outcomes, procedures, and responsible physicians for all patients. Consequently, statewide outcome data are available for specific review and analysis.

Partly in response to the 1986 Health Care Financing Administration (HCFA) mortality studies, the NYSDOH undertook its own statistical analysis of mortality data. The Department of Health acknowledged that quality-of-care determinations cannot be made from mortality alone; it concluded that targeted case reviews were both appropriate and useful. At least three such case reviews are in progress.

First, "rare deaths" were studied—diagnostic categories in which five or fewer deaths occurred during the period 1983 to 1985. A preliminary review by medical consultants concluded that approximately 320 cases required further medical record review. Second, a high mortality rate study for selected diagnosis-related groups (DRGs) was initiated. Hospitals with substantially higher than average mortality rates were reviewed by Medicaid review agents. These studies have been completed, and although the results have not been released yet, it is my understanding that statements of deficiencies have been sent to 27 hospitals.

Third, a quality indicators mortality study was conducted. Medical consultants retained by the Department of Health have identified 12 specific factors that may indicate quality problems. These have been applied to a study of 5,000 mortalities in downstate New York. The case reviews are being conducted by Medicaid review agents and have yet to be reported. So the outcome emphasis, the mortality issue, remains very strong in the efforts of the NYSDOH.

Cosurvey Process

The third origin of the emphasis on quality arose from the cosurvey process developed in New York State. Historically, the Department of Health always participated in the surveys of the Joint Commission on Accreditation of Hospitals (now, of Healthcare Organizations). In 1985, however, after an intensive survey of multiple hospitals in Buffalo, the Department of Health decided to embark on its own evaluative studies. It concluded that the survey process could be enhanced signifi-

cantly by assigning a large team of surveyors either concurrently or independently of the Joint Commission. This has resulted in a NYSDOH presurvey process, which Presbyterian Hospital experienced prior to the arrival of the Joint Commission team. The presurvey identifies what is believed to be a series of significant problems in the institution. These data are given directly to the Joint Commission surveyors when they arrive so their efforts are more focused. The future of the cosurvey is uncertain, but the state will certainly continue to put a great deal of effort into this process.

The origins of state regulation of and interest in quality assurance programs are multiple—first, malpractice crisis and the ensuing legislation; second, the review of collective mortality data from different perspectives; and, third, a more aggressive survey process. The interest of the public in these data obviously is great. The public release of this information by the Department of Health is a regular process under the Freedom of Information Law. Both politicians and the press have chosen to sensationalize some of this information.

HOSPITAL ASSOCIATION INITIATIVES

The second major area of New York State quality assurance activity involves the HANYS and the local and regional hospital associations. This activity also was undertaken for a variety of reasons: first, to highlight the commitment of institutions to high-quality care; second, to provide guidelines to hospitals throughout the state, particularly in dealing with the multiple regulatory pressures that were then developing; and, last, to respond to the increasing public concern about quality assurance.

A year ago the HANYS formed a Quality Assurance Task Force. I had the pleasure of chairing that group, which consisted of about 30 individuals including administrators, quality assurance directors, physicians, and other interested parties. Our charge (Table 1) was global. We were to

TABLE 1 Charge of the Quality Assurance Task Force of the Hospital Association of New York State

- Identify and evaluate issues relating to the delivery of quality patient care
- Define quality indicators and standards of quality
- Integrate regulatory and surveillance requirements of the Department of Health and the Joint Commission
- Develop a model quality assurance plan

identify and evaluate issues relating to the delivery of quality patient care; to define quality indicators and standards of quality; to integrate regulatory and surveillance requirements of different agencies with the goal of developing a model quality assurance plan. The latter effort proved to be the most pressing at the time because the institutions throughout the state were not prepared to address these issues and were in significant disarray.

The objectives of the model plan (Table 2) were to achieve compliance with regulations and to be comprehensive and yet flexible enough to deal with both large and small hospitals. We wanted to delineate responsibility and accountability within each institution; to coordinate and integrate quality assurance functions; and to create change through information, education, and collegial commitment to quality. Finally, we wanted to build in some redundancy and the expectation that each person who could identify a quality issue would do so and would report it.

We formed three work groups:

1. Quality Indicators and Standards
2. Regulation/Surveillance Integration
3. Quality Assurance Plan

Work group 3 has succeeded. Group 2 has had some success, and group 1 remains an area that we have not addressed fully. We are reconfiguring the group to see if we can begin to think about indices of quality care in a more productive way now that we have developed an overall process.

Our efforts have produced a 35-page book entitled *Quality Assurance Plan Guidelines.* It contains guidelines for individual hospitals, an outline of how to cope with the current regulatory pressures concerning quality assurance, and preliminary definitions of some indices of quality of care.

One important aspect of this effort was to define the authority and

TABLE 2 Objectives of the Model Quality Assurance Plan

- Achieve compliance with regulations and Joint Commission standards
- Be comprehensive yet flexible and adaptable
- Clearly delineate responsibility and accountability
- Coordinate and integrate quality assurance functions
- Create change through information, education, and collegial commitment to quality
- Build in redundancy—the expectation that each person in a position to identify an issue will do so and will report it

responsibilities within the institution. The law in New York State permits the quality assurance committee to report directly to the governing body. The task force concluded, however, that the chief executive officer of the institution is responsible and should be held accountable for the design, implementation, and effectiveness of the quality assurance plan.

RESULTS

What has happened? As I said before, quality of care has become headline material in New York State for the newspapers as well as the nightly television news. Some of it has been highlighted appropriately, and some inappropriately so. The release of mortality outcomes has certainly had its effect, but within New York State the most important element generating publicity has been the incident reporting system that was mentioned earlier. All hospitals are obligated to report the following items to the Department of Health within 24 hours: (1) patient deaths or impairments of bodily functions in circumstances other than those related to the natural course of illness, disease, or proper treatment in accordance with generally accepted medical standards; (2) fires in the facility that disrupt patient care or cause harm to patients or staff; (3) equipment malfunctions during treatment or diagnosis of a patient that had or could have had an adverse effect on a patient or health facility personnel; and (4) poisoning occurring within the facility. Strikes, disasters, and termination of any general services such as electricity, telephone, gas, or water are also reportable, but the first four types of incidents are the ones most directly related to patient care.

During 1986 approximately 4,400 incidents were reported. These concerned treatments, procedures, patient falls, equipment failures, and poisonings. Falls are by far the number one item reported. However, significant incidents concerning treatment, medications, and equipment also occur. Review of more than three-quarters of these was completed rather quickly, although hospital investigation was required in almost all of them. The Department of Health has finished the investigation of more than half of the remaining 1,000 incidents. So far, the investigation has revealed approximately 200 significant deficiencies. Between 15 and 20 percent have resulted in enforcement procedures—but enforcement procedures are proving to be very difficult for institutions in New York State. Hospitals are participating effectively in this process, reporting a large amount of data to New York State, on which the media frequently capitalize.

The most serious consequences for hospitals occur when the Department of Health interprets an incident as negligence in the care of a

patient and cites the hospital and, perhaps, physicians for a number of deficiencies. Negotiation then ensues between the institution and the Department of Health to develop an effective and reasonable plan of correction to minimize possible future recurrence of a similar event. Throughout this dialogue, all processing of certificate-of-need applications of the hospital in question is suspended.

In evaluating an incident the Department of Health frequently contends that the governing body and administration have failed "to effectively discharge duties and responsibilities for the conduct and operation of the hospital." At times, this allegation is appropriate for the administration of the hospital; in general, however, trustees throughout the state think that this charge is totally inappropriate. In fact, it has led to some trustee resignation. In addition, institutions that submit new establishment requests to New York State and have previously agreed to stipulations arising from enforcement proceedings may have their trustee character and competence questioned. Consequently, a significantly adversarial situation has arisen as a result of the incident reporting system, its publicity, and the linkage to the certificate-of-need application process.

In addition, the cost of this monitoring is burdensome for the hospitals. No additional reimbursement is available to support the necessary staff.

Finally, investigation of incidents can be most disruptive. When single cases are analyzed, teaching hospitals have a particular problem. The normal interaction between residents at various levels and the attending physician—the "give and take" that normally occurs in a teaching setting—may be overinterpreted by the Department of Health and the media. The comments of first- or second-year residents are scrutinized with great detail; they may be used, not to the advantage of patient care, but to discredit both the institution and the individual involved. In fact, some residents-in-training have been charged with professional misconduct, which, I think, is a serious aberration. Fortunately, the Commissioner of Health agrees with this position and supports statutory protection for supervised resident physicians.

CONCLUSIONS

This brief overview of quality assurance activity in hospitals in New York State highlights both positive and negative aspects of current efforts by hospitals and the Department of Health. The impact on patient care appears to be positive. Institutions are certainly more observant about their patient care practices, and physicians are clearly more diligent

and attentive to various details. These changes have occurred in an unnecessarily adversarial setting, however. The real challenge for us is to move away from using gross indicators of quality of care—defining areas in which deficiencies occur and correcting them—to a more sophisticated level such as Dr. Donald Berwick discussed. We must begin to analyze quality of care with indices sensitive to intensity of illness and individual population demography. I anticipate that when that can be done we will bring great benefit to all.

Quality Assurance in Hospitals: From Process to Outcomes

James S. Roberts

I shall discuss the major steps that I believe are necessary to improve the effectiveness of quality assurance activities in hospitals. The discipline of technology assessment plays an important role in advancing our knowledge of effective mechanisms to monitor and improve the quality of care.

First, let me offer some definitions that provide the context in which I view quality and quality assurance. Although high-quality care can be defined in many ways, I believe it is most appropriate to focus the definition on the basic purpose of health care: meeting the health needs of individuals or groups. Thus, I define quality health care as:

> The degree to which the health care needs (educational, preventive, restrictive, and maintenance) of an individual or a group are identified in a complete and timely manner and the resources (human and other) necessary to meet these needs are provided as effectively as the practical state of the art allows.

This needs-driven concept of quality emphasizes timely and accurate assessment and treatment; it stresses that expectations of the effectiveness of health care should go no further than that which is achievable under the conditions of good day-to-day practice.

I hope that, by focusing on the degree to which individual or group needs are identified and met, we might also face more squarely the implications that barriers to access to care, inadequate resources, and inappropriate limitations of insurance coverage constrain the ability of the health care field to provide high-quality care.

FACTORS INVOLVED IN ACHIEVING HIGH-QUALITY CARE

However one defines high-quality health care, several factors contribute to its achievement (1). These structural and process variables,

singly and together, shape the results or outcomes of care and constitute the measures used in quality assurance to gauge the quality of care. They include

- Variables related to the patient, the practitioner, and teams of practitioners caring for the patient.
- The organization under whose auspices the care is rendered.
- The diagnostic and therapeutic process itself.
- The nature of the local, state, and federal health care policies that frame the structure, insurance benefits, and reimbursement approaches for the patient and practitioner.

QUALITY ASSURANCE

As the final stage-setting definition, it will be helpful if I present my concept of quality assurance. One can view the achievement of high-quality care as the objective of everything done in health care and, thus, properly encompassed under quality assurance review. Indeed, as our notions of quality assurance evolve, this will be understood to be true. For purposes of this volume, let me discuss quality assurance in the more contemporary fashion—quality assurance is the process used to define and measure clinical performance. Its results, in turn, guide the improvement of the quality of care. Thus, an effective quality assurance program in a hospital has three main results: (a) it maintains high-quality care where it exists, (b) it improves the quality of care where it is wanting, and (c) it incorporates advances in state-of-the-art medical care into day-to-day practice.

The quality assurance process has the following major steps, each of which is familiar and needs little elaboration

- Identification of indicators related to quality.
- Data collection relative to the indicators.
- Data analysis and problem identification.
- Peer review/problem analysis.
- Feedback of findings.
- Pursuit of improvement through effective and appropriate action.

I will return to the central issues of data collection and peer review later in this paper. For complex and understandable reasons, it constitutes one of the weakest links in this chain.

The range of subject matter for quality assurance is broad. Using indicators related to the structure, process, or outcome variables noted

above, quality assurance efforts have generally concentrated on individual health care processes, health problems or diagnoses, or individual health care practitioners. In the future, more attention will be placed on health care teams and organizations and on the quality of care in a geographic community.

This shift in emphasis is being created by numerous environmental forces. They include increased practitioner subspecialization and the resulting need for creation of formal and informal health care teams; greater competition between health care organizations and the resulting demand for measures that differentiate quality among competitors; the growing confusion over ultimate responsibility for quality as local health care markets are characterized by sometimes interconnected and sometimes competing health care organizations; and the growing number and influence of vertically integrated regional health care systems.

UNIVERSAL NEEDS OF QUALITY ASSURANCE PROGRAMS

Whatever the subject being evaluated in a quality assurance program, certain needs must be met for the program to be successful. They are

- Methods to identify the health care problems or needs of a patient or group of patients and/or descriptive data concerning the types of services provided by a practitioner, an organization, or a patient or group of patients.
- Indicators of high or substandard quality care for these types of needs or services.
- Thresholds for evaluation of data relative to these indicators.
- Data on performance relative to these thresholds.
- Methods to analyze and compare performance related to thresholds, past performance, trends, and comparable groups.
- Effective peer review and problem identification.
- Understandable and relevant feedback of performance to those being evaluated and to change agents.
- Effective methods and actions that improve performance.

Of these familiar components, let me highlight two: data collection and peer review. The first concerns the need to compare the subject being analyzed with another relevant measure of performance. At present, the United States lacks data bases that can be used to describe normative performance related to anything but the most crude measures of quality. The Health Care Financing Administration (HCFA) has released

crude and "adjusted" mortality rates for Medicare beneficiaries and will continue to do so (2). These data are typical of the limitations of current health care data bases. The indicator of quality (mortality rate) is not very useful in discriminating good from poor quality care; the data are limited in scope to one population (Medicare beneficiaries) and, for the moment, to one setting (hospital); moreover, the measures are not well adjusted for the severity of illness of each hospital's population.

The landmark decision by HCFA to release hospital-specific mortality data directs attention to the limitations of all health care data bases in terms of their usefulness for monitoring the quality of care. Much work is needed here. I urge you to focus attention on how we can improve the content, collection, analysis, and feedback of data to national and local quality assurance efforts.

In addition, I wish to point out the importance that peer review plays in any first-rate quality assurance program. Those involved in evaluating clinical performance understand that clinical decisions—the ultimate issue in quality assurance—are made by practitioners armed with a less than complete scientific underpinning (3). Medicine is still very much an art. Patient care decisions are made by practitioners who, most often, are doing their best to apply existing knowledge to the needs of their patients. But research is often inadequate to help decide for or against surgery, the relative probabilities of the benefits outweighing the dangers in using a potent medication, or the relative risks of early discharge of a patient from the hospital.

Further, the indicators and criteria used in quality assurance programs are not definitive enough to identify poor-quality care. Mortality rates, hospital-acquired infection rates, the frequency of unplanned returns to the operating room, the incidence of readmission to the hospital, and all other indicators lack the power to allow final decisions on quality.

The understandable imprecision of even the best clinical decision making and the weakness of current quality indicators couple to highlight the importance of peer review. Indicators are flags for subsequent peer review. In most instances it is only after effective peer review that final judgments can be made about the quality of care. Later in this paper, I will address the importance of societal protection of peer review and the need for further research into the process of peer review.

JOINT COMMISSION INITIATIVES AND THE "AGENDA FOR CHANGE"

Throughout its history, the Joint Commission on Accreditation of Hospitals (now, Healthcare Organizations) has continually modified its

standards and survey processes to stay current with and, often, to lead the health care field (4). This history of measured change has entered a new phase with the commission's developmental project called the "Agenda for Change."

Designed to improve substantially the commission's standards and surveys, the set of initiatives encompassed in the Agenda for Change is drawn from a detailed evaluation of the current and future environment of health care conducted by the Board of Commissioners. This analysis made three points clear. First, it is necessary and feasible to use data related to key clinical and organizational indicators as a tool to better evaluate health care organizations. Second, it is essential that Joint Commission standards reflect the most current research that identifies the characteristics of effective health care organizations. Finally, it is possible to combine the monitoring of clinical performance with better analysis of organizational structure and function to create an improved on-site survey process focused more on organizational effectiveness.

These are lofty goals, and time does not permit me to describe the details of this multiyear research and development effort. I will, however, outline its five major components and highlight certain key concepts.

Selection of Clinical Indicators

The substantial investments made over the last two decades in clinical research, health services research, quality assurance, and risk management have resulted, we believe, in a body of literature and experience that will be central to the success of the major new element of the Agenda for Change, namely, the capability to monitor important indicators of an organization's clinical performance. This past work represents the substrate from which the Joint Commission will, with the aid of expert panels, select clinical performance indicators to use in the monitoring process, which I will describe below.

The performance indicators will be limited in number and will identify important clinical processes and, more often, clinical outcomes worthy of monitoring. They will be related to one area of care or will cross specialties, and they will be drawn from research and experience that identifies them as potentially useful screening devices. The first task forces are in obstetric care, anesthesia care, and hospitalwide indicators. They will be followed by similar work in the other major hospital services and for nonhospital care as well.

To those involved in clinical care, quality assurance, or risk management, the indicators that are selected will come as no surprise. The use

of mortality data, major patient complications, or risk-driven clinical events is (or should be) commonplace in institutional clinical monitoring activities. The major contribution of the commission's new effort will be to identify the most valid of these indicators and to create a national data base of normative performance to which an organization can compare its own performance.

As noted earlier for these indicators—whether used individually or in aggregate—the research that supports their use is insufficient, and the factors that influence clinical performance relative to them are too numerous. We believe, however, that they will serve as very useful screening devices to aid health care organizations in identifying possible deficiencies in the quality of care and the Joint Commission in analyzing an organization's problem-solving capabilities.

Such information does not change the commission's historic interest. We will continue to want to be sure that the organization is rigorously identifying and evaluating (through problem analysis and peer review) possible clinical problems and improving care where necessary. Accreditation decisions will flow from answers to these questions, not from sterile review of clinical outcome information.

Case-Mix Adjustment

Closely related to the clinical indicators initiative is the effort concerning case-mix adjustment. Clearly, one reason clinical performance varies within and among organizations is the differences in the severity of illness of their patient populations (5).

As the Joint Commission analyzes data on the selected clinical indicators, it will need to make adjustments for differences in case mix. The technology of case-mix adjustment, however, is new, not well refined, and rapidly evolving. What are the clinical, social, genetic, and other factors that help place a specific medical complication (heavy postoperative bleeding, third-degree laceration during delivery), morbidity indicator (eclampsia, major arrhythmia during anesthesia), or mortality (maternal, infant, cancer patient, postoperative patient) into proper perspective? Surely, the answer goes beyond age, race, sex, and diagnostic category. Yet the solution must not, as a practical matter, involve abstracting a large volume of clinical data on each patient.

The truth lies in modest but relevant clinically based adjustment, which can narrow the full spectrum of variation in clinical performance to a confined range in which there is a reasonable expectation (but not a guarantee) that a problem exists. This approach meshes nicely with our

belief that clinical performance data are not adequate by themselves to evaluate quality. Case-mix-adjusted clinical data are useful only in raising questions—they give no final answers relative to quality. The judgments about quality must be made in a professionally responsible manner after effective peer review.

Organizational Indicators

In recent years all segments of the U.S. economy have been under heavy pressure to become more efficient and in many cases to improve the quality of their product or service. Health care has not been immune from this pressure.

This need prompted best-selling books and academic papers on organizational excellence, effective leadership, and new managerial concepts. The writings tend to focus attention on such concepts as clarity of organizational purpose, employee involvement in planning and decision making, effective cross-organizational coordination, strategically oriented planning, top-to-bottom commitment to quality, and responsive customer service.

Using academic experts from the "organizational effectiveness" field and a spectrum of experienced health care managers, we are testing our belief that it is possible to identify the major characteristics of an effective health care organization. If it is possible to gain consensus on such characteristics, we plan to take two important steps. The first is to modify and streamline our current standards to bring them in line with these major concepts of organizational design and management.

The second is to identify indicators of possible organizational dysfunction that are worthy of monitoring. The need for monitoring such events has been apparent to the Joint Commission for several years. Our survey experience demonstrates how quickly a health care organization can deteriorate or improve as a result of a major organizational occurrence. A hospital can change from good to bad with a disruptive change in ownership, sudden loss of key management personnel, or inability to hire or retain sufficient numbers of qualified nurses. The quality of care in a nursing home can improve dramatically with the hiring of a new director of nursing.

With the rapid change occurring in health care, we believe it is necessary to monitor the occurrence of these events between full surveys and to assess quickly the impact they are having on the organization's ability to provide high-quality care.

Improvement in Monitoring Between Surveys

Using both clinical and organizational indicators, the Joint Commission will monitor clinical and organizational performance between surveys. This monitoring system will aid the commission and the accredited organization in assessing its performance continually relative to its past performance and to other organizations providing similar services. A system as dynamic and—some would contend—as fragile as our health care system demands such monitoring to catalyze timely and targeted improvement.

In setting this goal the Joint Commission is well aware of the data collection, reporting, analysis, and feedback implications of a continuous monitoring system. Our plan is to limit data to those that are essential and to be as parsimonious as current technology allows in our use of existing data streams.

Improved Survey Process

The end product of Joint Commission work is an on-site evaluation of a health care organization and a resulting judgment on compliance with commission standards. The new monitoring approaches described above do not negate the need for an on-site evaluation. They will, however, target such review to those elements of an organization's structure and function that have the most effect on the quality of care. The revised standards will be more relevant, and data from the monitoring system will allow better targeting to areas of possible organizational or clinical dysfunction.

This design is, of course, considerably different from the Joint Commission's current approach. It will require surveyors who can credibly and accurately assess organizational performance using more relevant data. Further, the survey process itself will have to change from one focused on capability to an assessment of actual effectiveness. For those concerned that the Joint Commission has, in the past, prompted irrelevant paper shuffling, this change will be welcome indeed.

MAJOR OBSTACLES TO EFFECTIVE QUALITY ASSURANCE

I would like to turn to an explicit discussion of the major factors that I believe impede the effectiveness of quality assurance programs in health care organizations and may limit the impact of the Joint Commission program I just described. I will address two major categories—socie-

tal obstacles and technologic/methodologic obstacles. Both must be addressed if quality assurance is to have its desired effect.

Societal Obstacles

As with most endeavors designed to produce behavioral change, quality assurance is affected mightily by the manner in which our society deals with health care quality and the health care field. Listed below are several problems that must be addressed by health care policy leaders—both health care professionals and those who formulate government, insurance, industry, purchaser, and consumer policy directions. A brief description follows each.

Lack of Full Commitment by Our Society to Providing High-Quality Care to All

To be fully effective, health care practitioners and organizations must have the freedom to provide high-quality care whenever and wherever it is needed. Constraints on access to or financing of care for segments of our population limit those involved in providing, evaluating, and improving care. This is particularly relevant to quality assurance programs in institutions serving a high proportion of uninsured or underinsured patients.

Lack of General Consensus About What Our Society Means by "High-Quality Health Care"

This is an important and complex issue. As noted earlier, views on what high-quality care means differ widely. For example, some believe access to care is a quality issue and some do not. Most consider patient satisfaction to be an important element of quality but differ widely on its definition; some experts limit satisfaction to measures of patient perceptions of system responsiveness and practitioner kindness, whereas others include measures of the congruence between patient expectations of symptom relief and that actually achieved.

Coupled with the growing expectation that quality assurance programs will improve care must, in fairness, come consensus on those elements to be included in the definition of quality health care.

The Gap in the Perception Versus the Reality of the Science of Medicine

Our society continues to cling to the notion that health care is no longer a "practice" but is a fully mature science. This view—partly cre-

ated by the health care field itself—prompts unrealistic expectations about the operation of quality assurance programs. In particular, there is a perception that measures of quality are easy to identify and that performance relative to such measures can be aggregated to produce valid rankings for health care organizations. For example, some view it as feasible to rank-order hospitals based on the gap between their actual and "expected" mortality rates.

This perspective fails to take into account the dearth of literature defining the relative probabilities of success, complications, or death from various diagnostic or therapeutic procedures and the lack of well-tested methods to adjust for case-mix differences among hospitals. Lacking a full scientific base, quality assurance must depend on individual case review done by peers, and we must all learn to live with the consequences of a process that must depend on the informed, but subjective, judgments of peers.

Inadequate Legal Protection for Effective Peer Review

The importance of peer review raises an issue essential to the success of any quality assurance activity. If meeting the professional obligation to engage in peer review threatens the personal assets of health care practitioners, quality assurance will not be done well. Our society must protect those who participate in good-faith peer review. On this note it is encouraging to contemplate the protection available under the Health Care Quality Improvement Act of 1986 (6).

Diffusion of Responsibility for Quality at the Community Level

Until recently the physician and the hospital were understood to have sole responsibility for the quality of care received by patients in most communities. That reality has changed dramatically with the revolutionary reconfiguration occurring in our health care system. In examining the structure of community-level health care, one finds a confusing mixture of organizations involved in defining services and in financing, insuring, and providing care. Health maintenance organizations (HMOs), preferred provider organizations (PPOs), independent practice associations (IPAs)—the alphabet soup of new financing and delivery organizations abounds in local "markets." Further, efficiency demands have prompted many health care organizations to contract with others for provision of some care.

All of these changes have created great diffusion in responsibility for the quality of care. Is the HMO or the hospital, or both, responsible for

the quality of inpatient care provided to an HMO member? What is the relative responsibility of the hospital and the home care agency for the quality of the care provided under a joint venture between the two organizations?

Quality assurance programs depend for their success on the support and action of those who are ultimately responsible for the quality of care being reviewed. These individuals provide the policy context, encouragement, and resources necessary for the programs and assure that corrections are undertaken to improve quality where it is found deficient. From the confused local health care scene must emerge those who are responsible for quality so that they can, in turn, support and participate in quality assurance activities.

Technologic/Methodologic Problems

Along with the societal problems, quality assurance is limited by a series of technologic or methodologic problems. These range from scarcity of a solid research underpinning to shortcomings in the clinical data available in hospitals.

Scarcity of Sound Research Identifying the Organizational Variables that Have a Proven Relationship to High-Quality Care

Each of us knows that high-quality health care depends, in part, on how the health care organization responsible for our care is structured and operates. We could agree, I believe, that a hospital that (a) has effective coordination between clinical and support services, (b) is clear about its basic mission and role in its community, and (c) has a cooperative, quality-centered relationship among its governing body, management, and medical staff is more likely to provide high-quality care than a hospital lacking these characteristics.

Yet a look at the literature to confirm or refute these perceptions is disappointing. The science of health care management is early in its development. Those interested in organizational effectiveness are left to depend heavily on expert opinion and experience to identify characteristics of organizational structure and function more likely to produce high-quality health care. Although such reliance is not without its benefits—real-world orientation is essential—research into the organizational determinants of quality should be expanded to solidify the scientific base of this field. This need is surely heightened by the remarkable restructuring occurring in our health care system.

The Limitation in Our Understanding of the Clinical Process and Outcome Indicators Related to High-Quality Care

The health care literature abounds with studies that explore the efficacy of given diagnostic or therapeutic procedures. Yet, many of these studies lack scientific rigor, are conducted in a controlled environment not achievable in day-to-day practice, or relate to narrowly defined populations of patients. Each of these characteristics of the clinical research literature decreases its usefulness to quality assurance.

For example, put yourself in the shoes of a department director who is interested in national mortality data for a given procedure—a potentially useful benchmark for the department's quality assurance program. It is possible to find studies on this subject, but on careful review, it is also likely that most articles report data for a limited number of patients and show considerable variation in mortality rates. Should this literature be used to set a mortality standard? If so, which rate should be used? How should the differences in case mix be accounted for between a particular department's patients and those in published studies?

Inadequate Attempts to Synthesize Existing Knowledge and Disseminate It in Ways that Will Help to Change Clinical and Organizational Performance

Despite the negative picture I painted concerning the quality and usefulness of existing health services literature, it is also true that much that is good in the literature is not fully known or used. This country needs more coordinated attempts to draw together the knowledge and experience concerning given diagnostic or therapeutic processes and assure the broad and effective distribution of such information. In this regard the National Institutes of Health Consensus Conferences and the clinical efficacy work of the American College of Physicians should be commended.

At the local level, quality assurance programs should be a vehicle to help assure the integration of existing knowledge into day-to-day practice. The selection of clinical care areas to be monitored, the identification of indicators and criteria to be studied, the peer review prompted by poor compliance with criteria, and the actions designed to improve care are all steps in the quality assurance process that could—and should—be used to integrate the best clinical information into routine practice.

Technology assessment could be helpful in identifying those techniques that are useful at the national and local levels in integrating existing knowledge into quality assurance programs.

Lack of a Full Understanding of How to Design and Support Valid Peer Review and Measure Its Effectiveness

I have already noted the importance of societal protection for good-faith peer review. In addition, more research is needed on the techniques of peer review. What personal and professional characteristics make an individual an effective peer reviewer? What are the approaches to constructing a peer review process to assure its accuracy, validity, consistency, and effectiveness? How does one measure the effectiveness of an admittedly subjective process such as peer review? These issues are critical to the success of peer review, yet they remain largely unanswered.

Limited Availability of Clinical Data Useful for Monitoring and Comparing Clinical Performance Across Practitioners and Organizations

The data problem confronting quality assurance programs is substantial. It is our strong perception that data systems in hospitals have been constructed largely to answer financial and resource tracking needs. Hospitals have made little investment in the routine collection, storage, and analysis of clinical performance information so vital to effective quality assurance. In part, this is the result of the problems noted above, but it is also a predictable consequence of the inattention of hospitals to day-to-day monitoring of quality—inattention unheard of in other sectors of our economy.

Likewise, little clinical performance information is available at the national level. Where does one go to compare performance on clinical processes or outcomes across hospitals, insurance plans, or patient populations? There is no such repository. One of the driving forces behind the Joint Commission's Agenda for Change is the creation of such a data base.

The Need for Conceptual Clarity and Methodologic Improvement in Measuring Severity of Illness and Case Mix for Use in the Evaluation of Quality

Related to the need for better comparative data is the importance of solidifying methods to characterize a patient's severity of illness and to aggregate such measures into a case-mix classification for a procedure, a department of a hospital, or the hospital as a whole. Most such systems have been developed to analyze resource use, not for quality assurance. Likewise, they vary in their conceptual base and are rapidly changing.

As we have begun to identify the clinical indicators we will be testing in 1988, we are finding that, in many areas of clinical practice

(trauma, coma, myocardial infarction, anesthesia), clinically based classification systems have been developed and are in widespread use. These sharply focused systems might prove most useful in adjusting for the differences in patient populations among hospitals. Efforts to evaluate carefully the various severity classification systems and to assess how they should be used in a quality assurance program would be of great help to the field.

Better Methods to Synthesize Organizational and Clinical Performance Information into Valid Judgments About an Organization

Finally, we all recognize that health care organizations are being asked to be more publicly accountable for their clinical performance. Organization-specific outcome information is, and will continue to be, available. Having such data, consumers, governmental agencies, insurers, and businesses are naturally interested in knowing how good a hospital, HMO, nursing home, home care agency, or other organization may be.

Yet, we also know that clinical outcome data do not tell the whole story and are not sufficient to differentiate the good from the bad. More important, poor quality care can and does occur at even the best hospitals. There can be no absolute guarantees when one enters a hospital. The important issue, once basic levels of clinical performance are reached, is an organization's persistence in continually improving the quality of its care. As a patient, I am interested less in point-in-time clinical and organizational performance than in the fact that the organization is demonstrably dedicated to constant improvement in performance.

Thus, we need both organizational and clinical information and approaches to synthesizing this material in a manner useful to a variety of audiences. Exploring how this is accomplished in other industries might prove beneficial.

REFERENCES

1. Roberts, J.S., and Prevost, J.A. Using outcome indicators to evaluate the quality of care. The Internist. September 1987;10–15.
2. Health Care Financing Administration. DHHS. Medicare Hospital Mortality Information 1986. Washington, D.C., December 1987.

3. Eddy, D.M., and Billings, J. The Quality of Medical Evidence and Medical Practice. Paper prepared for the National Leadership Commission on Health Care. Washington, D.C., 1987.
4. Roberts, J.S., Coale, J.G., and Redman, R.R. The history of the Joint Commission on Accreditation of Hospitals. Journal of the American Medical Association 258:936–940, 1987.
5. Pollack, M., Ruttimann, U., Getson, P., et al. Accurate prediction of the outcome of pediatric intensive care. A new quantitative method. New England Journal of Medicine 316:134–139, 1987.
6. Iglehart, J.K. Congress moves to bolster peer review: The Health Care Quality Improvement Act of 1986. New England Journal of Medicine 316:960–964, 1987.

Assessing the Quality of Care in Family Physicians' Practices by the College of Family Physicians of Canada

Alexander E. M. Borgiel

Everyone involved in the delivery of health and medical care has a commitment to quality. It is one thing, however, to be committed to quality care and another to implement programs that are effective in assuring quality.

In North America setting standards for quality medical care in hospital and office practice has traditionally been the responsibility of the medical profession. Quality assurance has been an explicit part of the rationale in establishing certification procedures, initiating peer review programs, promoting continuing medical education, and discussing relicensure or maintenance of certification.

A growing demand for accountability of the medical profession and hospitals has arisen out of public concern about the quality-of-care provided in the United States and Canada. The Canadian Council on Hospital Accreditation now requires quality assurance programs for accreditation by member hospitals. Other professional associations, such as the Canadian Physiotherapy Association and the Canadian Nursing Association, are developing program standards and practice criteria. In Ontario the College of Physicians and Surgeons, a licensing body, has been involved in peer assessments of practices since 1978, using quality-of-care methods and criteria (1).

Incentives for change have been rapidly mounting during the past 20 years (2,3). Among those incentives are (a) the escalating costs and increasing government funding of health care (4); (b) an increase in the number and severity of malpractice claims and suits; and (c) ethical considerations in the face of a growing awareness of the magnitude of preventable patient injuries occurring in our hospitals and doctors' offices.

The College of Family Physicians of Canada (CFPC) contributes to the education and performance of its members to ensure an optimum

standard of patient care. During the past 25 years the CFPC has encouraged the establishment of undergraduate and residency training programs for family medicine in all medical schools in Canada. Graduates of residency programs and previously practice-eligible physicians take examinations comprising written questions, simulated patient interviews, and orals for certification in family medicine. Members and certificants must complete 50 hours of continuing medical education every year to maintain certification. In addition, the CFPC has developed a maintenance of certification program. CFPC membership represents more than half of the 16,000 family physicians in Canada (5,6).

PRACTICE ASSESSMENT PLAN OF CFPC

The CFPC, as the accrediting body in family medicine, realized that, although significant advances have been made in the education and training of family physicians, less progress has been made in assessing the function of these same family physicians in the practice setting (7). In 1980 the CFPC established the Committee on Practice Assessment, which is composed of epidemiologists, research-oriented family physicians, professors of family medicine, sociologists, and a practicing family physician. This committee, acting with the full endorsement of the executive committee, prepared a feasibility study, funded in June 1982 and completed and published in the *Canadian Family Physician* in April 1985 (8). Dr. Robert H. Brook was the consultant for our project.

The words "assessment," "quality of care," "practice audit," and "quality control" have produced very few positive sentiments among family physicians in the past. Why then would a practicing family physician such as myself, unattached to a university, feel so strongly and positively about this work? It is my contention, and that of the CFPC's Committee on Practice Assessment, that unless we do this and do it quickly, quality assessment will be done by others with motivations much less positive than our own. As a practicing family physician for 20 years, I know that the only important improvements in the care and quality that patients have received in my practice have been from innovations and procedures that I learned from other physicians during my trips to their offices and in the hospital setting. Family physicians can learn from each other, not in isolation, but as a community pooling our resources and ideas to improve our satisfaction and patient care. It was this philosophy that generated the research program of the CFPC, which during the past 15 years has evolved to the present stage.

The mandate given to our committee in 1980 includes the following objectives:

- Develop a set of quality-of-care assessment procedures that is practical, economical, and acceptable.
- Use these same assessments to identify strengths and deficiencies in individual practices and variations across practices.
- Develop strategies and tools directed toward specific patterns of practice to help physicians improve the quality and efficiency of medical care.

The bottom line was that the assessment procedures were to be voluntary and that the benefits and costs must be acceptable to family physicians since they ultimately would pay for this process.

METHODOLOGY

The methodology for the assessment of family physicians' offices is composed of three instruments used to obtain data for analysis: physician interview questionnaire, chart abstracts, and patient questionnaire.

Physician Interview Questionnaire

This instrument is sent out two weeks prior to an on-site visit and is completed by the physician and his or her staff (7,9,10). In this part of the assessment we attempt to obtain information in seven areas: the practice profile; office facilities; after-hours coverage; referrals and consultations; hospital practice; medical records; and education and research.

These facts are reviewed and clarified at the time of the on-site visit. The questionnaire usually takes between 45 minutes and one hour to complete.

Chart Abstracts

Formulating how to use the family physician's charts as a measure of quality of care required the most intensive work. In reviewing the literature on quality-of-care assessments and appraisals the project committee was struck by the limitations of attempting to use structural characteristics, utilization reviews, observational studies, indicator conditions, or criteria mapping in assessing primary comprehensive and continuing care as defined by the CFPC. More specifically, we were aware of the uncertainty in family medicine arising from the indeterminate nature of most presenting problems, the subjectivity of assessments and response, the inability to resolve clearly most primary care problems through management,

and the wide range of problems that are presented (11).

After several meetings of practicing family physicians (in groups of five or six) in London, Toronto, and Ottawa, Ontario, we reached a consensus for criteria required in the family physician's chart. In rating medical care for episodic visits we sought explicit criteria for the range and scope of problems seen by family physicians. We began with the criteria statements from the project for the evaluation of health maintenance organizations (HMOs) (12), which were modified in New Mexico and California. Added to them were criteria statements based on the indicator conditions of Sibley and associates (13) and the recommendations of the Canadian Task Force on the Periodic Health Examination (14). Our aim was to develop explicit criteria statements based on professional consensus. We made the assumption that each visit in a two-year time frame could be evaluated for both positive and negative factors (15–19).

Each chart abstract required an assessment, which we divided into five parts: criteria statements, chart format, prevention, use of drugs, and resource utilization.

Criteria Statements for Chart Abstract

We defined 182 diagnoses (e.g., otitis media; acute bronchitis) under the following dimensions of care: history; physical; diagnostic procedures; radiography; verification of diagnosis; management (including drugs and physical therapy; education and prevention); follow-up; psychosocial orientation; dangerous; and essential.

For each diagnosis, criteria were worked into appropriate dimensions. Each diagnosis did not necessarily have all the dimensions; indeed for areas such as patient education, psychosocial, hazardous, and essential, the criteria were sparse. The practicing nonuniversity-associated family physicians who formulated these criteria did not expect a book to be written for each diagnostic visit, but they were able to agree that for specific diagnoses a good family physician should be expected to do certain things to provide good care.

Ultimately, our criteria scores are applied to medical care provided over one or more visits. The goal is to identify strengths and weaknesses of the physicians on each dimension of care across problems rather than to provide information on the management of specific diagnoses. Hence, our final units for analysis are the dimensions of care. This is the fundamental difference between our approach and the tracer/indicator conditions approach in which indicator conditions are purposely selected and scores are determined by summing across factors for each diagnosis.

Chart Format

Abstracting the chart format enables us to assess the basic registration data, the drugs used, the legibility of the chart, specific aspects of the general assessment, and notation of allergies.

Prevention

The chart abstract for prevention assesses the recording of blood pressure, Pap smears, and breast examinations according to specific time frames and age categories. It also assesses behavioral history (smoking, alcohol, and use of street or proprietary drugs) and immunizations during the previous two years.

Use of Drugs

Eleven types of drugs commonly used in family medicine were assessed for appropriate use. They include antibiotics, nonsteroidal anti-inflammatory drugs, antidepressants, tranquilizers, anticoagulants, barbiturates, and steroids.

Resource Utilization

Our abstract of resource utilization assessed the use of imaging studies, number of visits to the physician in a two-year time frame, consultations initiated, and the number and amount of prescriptions for narcotics.

Patient Questionnaire

The third instrument used in our evaluation was a patient questionnaire developed and improved on on many occasions. We incorporated some material developed by other authors in the United States (20,21) and Australia (22) and modified the questionnaire for the CFPC. This, in many ways, is the most interesting and best-selling feature for family physicians. To increase their interest in this project, we gave physicians an option of including two of their own questions, which may not have been covered by our questionnaire.

The questionnaire included the following broad areas:

- Patient satisfaction with the doctor-patient relationship, access to physicians, and the timing of receipt of care

- Patient's assessment of the physical aspects of the doctor's office as well as the hours of operation
- Unmet medical needs resulting from health problems (e.g., chronic pain, headaches, insomnia) and problems in living (e.g., depression, problems at work, marriage problems)
- After-hours coverage
- Preventive medicine counseling (e.g., smoking, alcohol, obesity, exercise)

MECHANISMS OF ASSESSMENT

On the appointed day the assessment team arrives. The team is composed of a liaison physician and a trained nurse-abstractor. They come with a computer. Our assessment instruments have been incorporated into a computer software program developed for us by Dr. Michael McCoy of the University of California at Los Angeles. Figures 1 and 2 provide examples of the partial content of information collected during the chart abstract (here for acute bronchitis).

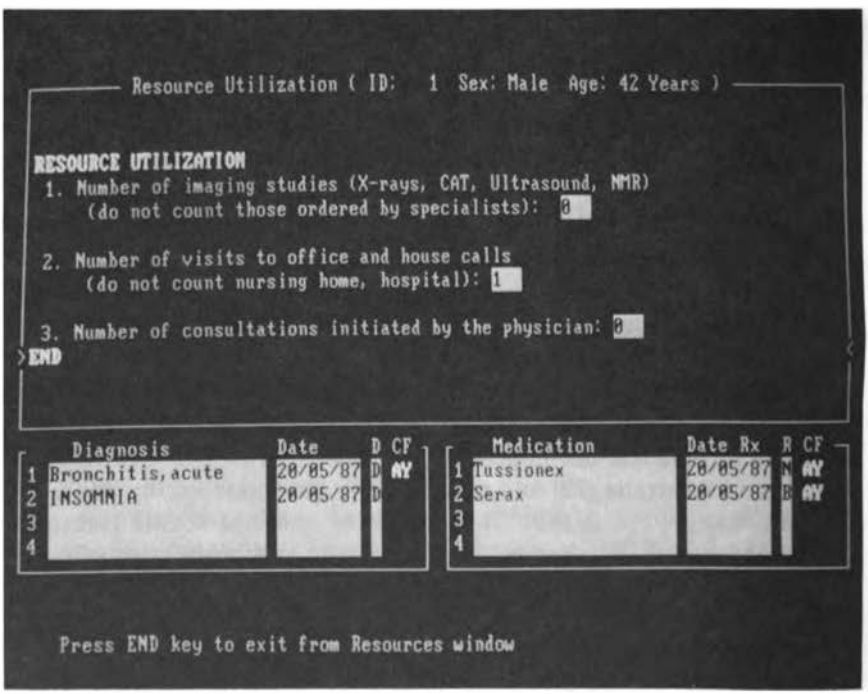

FIGURE 1 Process criteria set for acute bronchitis. Resource use.

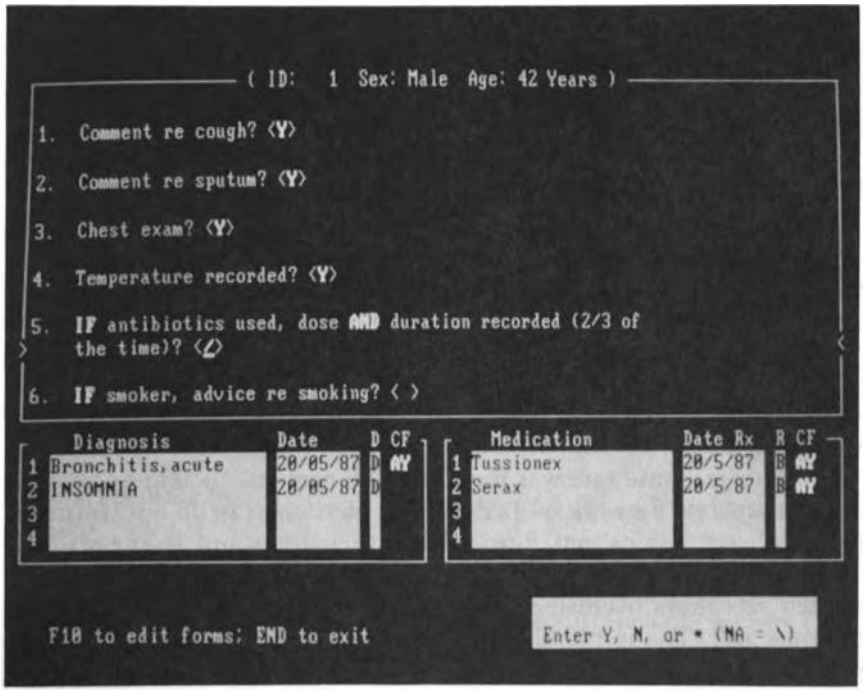

FIGURE 2 Process criteria set for acute bronchitis: history.

Twenty-five charts of patients who had visited the physician within the past two years are randomly picked by the liaison physician. In group practices only patients for whom at least two-thirds of the medical care had been provided by the candidate physician were eligible for selection.

The charts are assessed by the abstractor. This process takes between five and one-half and nine hours, depending on the variable characteristics of the charts, such as legibility, length, and individual mannerisms. The liaison physician also discusses the physician questionnaire that had been filled out by that physician and his or her staff.

The practice nurse or receptionist obtains names of 60 current, randomly selected patients equally distributed throughout the year. Once this information is obtained, the patient questionnaire is sent out. Subsequently, the results are keypunched and analyzed at the Kellogg Center in Montreal.

Following analysis, the results are returned to the physician in a narrative format. The computerized format, which we are developing with the help of Dr. McCoy, will enable us to send the physician not only

charts, figures, and comparisons but also an explanation that will address particular interests and problems.

Finally, I should note that our three instruments dovetail into each other and corroborate many areas of the practice profile in medical care.

RESULTS OF OUR FEASIBILITY AND PILOT STUDIES

Our research team has completed a feasibility study of 10 family physicians as well as a 10-doctor pilot project (8). In February 1987 we completed our assessment of 120 volunteer family physicians in southern Ontario. The candidate physicians were randomly selected from six districts in Ontario. They included nonmembers of CFPC, members of CFPC, practice-eligible certificants of CFPC, and residency certificants of the CFPC.

Our response rate was more than 80 percent. To achieve this, we asked a separate network of 32 dedicated physicians to do our recruiting (personal communication). Respected by their peers and, in the majority of cases, knowing the candidate physicians personally, this group of trained recruiters overcame several obstacles to produce the favorable response rate. The response rate for our patient questionnaires was also greater than 80 percent. Patients were anxious to reply and some even wrote letters explaining some points raised by the questionnaire.

At this time the results of our three instruments are being compiled and analyzed at the Kellogg Center in Montreal.

As a result of the activities outlined here, we can get the following global perception of our procedures. First, we have a method of analyzing a practice through examining the results of physician questionnaires, chart abstracts, and patient questionnaires, using computer technology. Second, we provide physicians with computerized narrative results that enable them to compare their patterns of practice with those of peers. Third, we can indicate to physicians weaknesses in the method and structure of their practice and in the care provided to patients. Fourth, we have a voluntary program, which ultimately will cost physicians only between $600 and $800.

FUTURE GOALS

What future purpose does the CFPC see for this procedure and mechanism? The following broad outlines were defined by the CFPC executive committee. First, this procedure can be used as a form of in-practice assessment of the practice-eligible candidates for the certifica-

tion examination. Second, it can serve as a stimulus to continuing medical education and professional satisfaction. Third, it can serve as a method of practice accreditation by the CFPC. Finally, it can serve as a stimulus to participation in the maintenance of the certification examination.

SUMMARY

In summary, the quality-of-care practice assessment program as presented and perceived by the CFPC is a positive, constructive method of practice analysis in ambulatory care. The analysis has a very positive effect and provides the reassurance that all family physicians need from time to time.

REFERENCES

1. McAuley, R.G., and Henderson, H.W. Results of the peer assessment program of the College of Physicians and Surgeons of Ontario. Canadian Medical Association Journal 131:557–561, 1984.
2. Canadian Medical Association Task Force on Education for the Provision of Primary Services. Ottawa: Canadian Medical Association, 1984.
3. Brief of the Executive Committee of the College of Family Physicians of Canada to the Canadian Medical Association Task Force on Education for Primary Care, 1983. (Personal communication.)
4. McWhinney, I.R., Johnson, C.A., Moore, C.A., Palmer, W.H., and Rosser, W. Postgraduate training for family practice and the specialties: There are enough positions. Canadian Medical Association Journal 133:967–968, 1985.
5. Grava-Gubins, I. Manpower study, 1986. Canadian Family Physician 32:2313, 1986.
6. Statistics from the College of Family Physicians of Canada. (Personal communication.)
7. Borgiel, A.E.M. Quality care: Assessment of family practice. Canadian Family Physician 23:823–825, 1977.
8. Borgiel, A.E.M., Williams, J.I., Anderson, G.M., Bass, M.J., Dunn, E.V., Lamond, C.T., Spasoff, R.A., and Rice, D.I. Assessing the quality of care in family physicians' practices. Canadian Family Physician 31:853–862, 1985.
9. Williams, J.I., and Steele, K.O. Constructing an Index of Accessibility for Primary Care Services. Toronto: University of Toronto, Health Care Research Unit, 1981.

10. Davis, D.A., Delmore, T., Bryans, A.M., Williams, J.I., Dunn, E.V., Krauser, J., Scott, D.J., and Heron, A. Continuing medical education in Ontario. Annals Royal College of Physicians and Surgeons, Canada 65:442–447, 1983.
11. Kessner, D.M., Kalk, C.E., and Singer, J. Assessing health quality: The cases for tracers. New England Journal of Medicine 288:189–193, 1973.
12. Leighton, R.E. Criteria for Evaluating Quality Assurance in Health Maintenance Organizations: Prepaid Health Research Evaluation and Demonstration Project. United States Department of Health, Education, and Welfare. Health Care Financing Administration. Washington, D.C., 1979.
13. Sibley, J.C., Spitzer, W.O., Rudnick, K.V., et al. Quality of care appraisal in primary care: A quantitative method. Annals of Internal Medicine 288:189–193, 1975.
14. Canadian Task Force on the Periodic Health Examination. The periodic health examination. Canadian Medical Association Journal 121:1193–1250, 1979.
15. Lohr, K.N., Brook, R.H., Kamberg, C.J., Goldberg, G.A., Leibowitz, A., Kessey, J., Reboussin, M.S., and Newhouse, J.P. Use of medical care in the RAND health insurance experiment: Diagnosis- and service-specific analyses in a randomized controlled trial. Medical Care 24:S1–S87 (September Supplement), 1986.
16. Moscovice, I. A method for analyzing resource use in ambulatory care settings. Medical Care 15:51–67, 1977.
17. Pineault, R. The physician episode of care: A framework for analyzing physician behavior regarding use of clinical and technical resources. Clinical Investigations in Medicine 8:48–55, 1985.
18. Schneeweiss, R., Rosenblatt, R.A., Cherkin, D.C., et al. Diagnostic clusters: A new tool for analyzing the content of ambulatory medical care. Medical Care 21:105–122, 1983.
19. Solon, J.A., Feeney, J.J., Jones, S.H., Rigg, R.D., and Sheps, C.G. Delineating episodes of medical care. American Journal of Public Health 57:401–408, 1967.
20. Aday, L.A., Anderson, R., and Fleming, G.V. Health Care in the U.S.: Equitable for Whom? Beverly Hills, Sage Publications, 1980.
21. Ware, J.E., Jr., Davies, A.A., and Stewart, A.L. The measurement and meaning of patient satisfaction. Health Medical Care Service Review 1:1–15, 1978.
22. Douglas, R.M., and Stevens, I.D. The Uses of "Kits" for Self-evaluation in General Practice. Department of Community Medicine, University of Adelaide, Adelaide, Australia, 1982.

PSRO/PRO— The Acronyms of Peer Review and Quality Assurance

Kenneth A. Platt†

Historically, structured peer review had its origin in the old Foundations for Medical Care, which were largely concentrated and developed in the 1950s and 1960s in California. The prototype foundation, from my perspective, was the San Joaquin Foundation started by Dr. Don Harrington in that era. Originally, the foundations concentrated on monitoring use of outpatient services with attention paid to the frequency of visits and the use of ancillaries. They were largely a response by the private sector to the competitive pressures of the Kaiser system. For several decades, this movement remained mainly on the West Coast and did not begin to spread nationally until the late 1960s and early 1970s. The increased interest was due to the economic and quality concerns of the federal payer systems of Medicare and Medicaid as well as to the increasing attention being paid to competitive medical plans.

In the years 1969 through 1972 the foundation movement spread eastward, especially to the Rocky Mountain States and Midwest, as more and more attention was paid to the problems of utilization and quality of care provided to the Medicare population. Those of us who recognized the inevitability of public accountability by the medical care system for the vast amounts of public and private expenditures used in the provision of health care turned to the foundation movement as a vehicle for providing the formal peer review needed to achieve that accountability. Recognizing the crude and rudimentary system then in existence for obtaining data that would allow us to do adequate peer review for utilization and quality, we turned to the federal government for seed money to develop new systems of data retrieval for peer review.

† Dr. Kenneth Platt, who was widely recognized as a major figure in the medical peer review movement, died in June 1988. He served the medical community of Colorado, and indeed the entire country, with distinction and dedication to the highest principles of the medical profession; his leadership will be greatly missed.

ORIGIN OF PROFESSIONAL STANDARDS REVIEW ORGANIZATIONS (PSROs)

Using monies obtained under the old Experimental Medical Care Review Organizations (EMCRO) grants, states such as Utah and Colorado put together the nucleus of statewide peer review organizations. They assembled new data retrieval instruments not so closely linked to the payer system and started a new profession of nurse reviewers to be the frontline workers in this new concept.

Coincidental with the new professional awareness of the need for public accountability through the peer review process was the activity of the Senate Finance Committee in setting up a formal monitoring system to assure proper care for Medicare beneficiaries. Under the guidance of Mr. Jay Constantine and Dr. James Monghan, the PSRO program was proposed as an amendment to the Social Security Act by Senator Wallace Bennett of Utah. When President Richard Nixon signed the Social Security Act in October 1972, the PSROs were born.

Their original mission, which remains in effect today, was to assure that the care provided to Medicare beneficiaries was provided in the appropriate setting, met acceptable standards of quality, and was provided over a reasonable amount of time with appropriate utilization of resources. In an attempt to keep the PSROs independent of organized medical societies on a state and national level they were deliberately created with multiple numbers in states with more than 2,500 physicians. The resultant group of more than 100 PSROs made implementing the program difficult, with heavy administrative and data costs slowing the process significantly. Organized medicine largely opposed this program because of its antipathy to the intrusive nature of peer review in general and to the linkage with government in particular.

The medical community generally viewed this whole process as a thinly disguised cost containment process, and it fought the program as an unnecessary and unwarranted interference in the traditional doctor-patient relationship. Nonetheless, despite the unpopularity of the PSRO program with the profession and segments of the federal government, it was put in place gradually in most regions of the country. Although the original fears—that PSROs would largely be a cost containment effort through utilization controls—were confirmed, some efforts toward quality measurements were begun under the label of medical care evaluation studies.

The decade or more of PSRO was a learning experience for the government and the profession. Proficiency in peer review and data retrieval did occur and some results were achieved in utilization controls. In general, the program was a toothless tiger in enforcing more appropriate utili-

zation of resources and in correcting the inadequate provision of quality services. Given the political climate in which the program operated and the inadequate funding from its inception, no one should have been surprised at its modest accomplishments. In retrospect, its most significant contributions to our current efforts may well have been the recruitment and training of a cadre of concerned professionals who became proficient in the peer review process and the stimulation of entrepreneurial entities to develop data support systems so essential to the process.

ADVENT OF PEER REVIEW ORGANIZATIONS (PROs)

With the advent of the PRO as the replacement for the old PSRO system, concurrently with the prospective payment system, the review focus shifted from utilization to quality. The new organizations had the same basic goals as the PSROs, i.e., appropriate care in the proper setting that met professionally acceptable standards of care. They differed greatly, however, in their numbers (54 as compared to more than 130), their focus on quality, and their potential for enforcement through retrospective denial of payment and the sanctioning process. With increased authority came increased responsibility and the need to recruit true peer reviewers of judicial temperament. Indeed, in the past three years the use of denial of payment and the sanctioning process has produced swirling clouds of controversy that, at times, obscure the honest efforts of the profession to self-regulate as a method of patient service and protection.

The Colorado Foundation for Medical Care has been in existence since 1970 and currently holds the PRO contract for Colorado. In addition, it has substantial contracts with the state of Colorado to provide review services for the Medicaid population in both acute and long-term care institutions, as well as some 700,000 privately paying patients through numerous private and public contracts. These contracts cover the entire spectrum of review efforts including preadmission and preprocedure approval, concurrent and retrospective review, second surgical opinion programs, quality assurance review and monitoring, and medical consultants to the OCHAMPUS program. With 14 years experience as a PSRO/PRO, one would think that making the shift from focusing on utilization and cost containment to quality monitoring and assurance would be a simple shift of gears. Nothing could be further from the truth.

QUALITY REVIEW

Quality is a poorly defined entity, subject to varying definitions

from differing perspectives. The technique of quality assurance is in its infancy. On a case-by-case basis it is more expensive than utilization control and subject to more professional variation in criteria setting. Medicine still is an art as well as a science, and identical outcomes can result from different processes of care. Equally competent peer reviewers will differ in their opinions on quality issues, especially where those opinions are based on retrospective chart review rather than on concurrent, hands-on evaluation.

We are in the process of implementing quality review for the PRO program from several viewpoints. One of the most obvious is the use of "generic quality screens." These examine a patient's hospital stay from several perspectives. These are

- Adequacy of discharge planning.
- Medical stability on discharge.
- Unexpected deaths.
- Nosocomial infections.
- Unscheduled return to surgery.
- Trauma.

If a case fails one or more of these screens, it is subjected to further medical scrutiny. This evaluation can result in denial of payment, imposition of sanctions with adverse publicity, and difficult hospital-staff relationships. Although these screens work, they need a severity of illness index to screen out unnecessary quality reviews.

Quality issues are also often addressed in objectives agreed to at the time of PRO contract negotiations, such as reducing the number of urinary tract infections associated with indwelling catheters and improving the appropriate use of aminoglycosides.

Other quality reviews occur as a result of patient complaints, concerns referred by the fiscal intermediary, or cases that have raised quality issues during routine reviews by the nurse coordinator with which the physician advisor concurs.

TECHNOLOGY ASSESSMENT

Although utilization review and quality assurance have not been widely applied to technological concerns, there is no reason that they could not serve as very useful tools in judging efficiency versus hazard, appropriateness of technological application, and the outcomes associated with the application of existing or new technologies. A case in point is Dr. William Knaus' APACHE II study, which is directed at the outcomes

resulting from the use of intensive care units. Using a system of point grading on admission to the units, a prediction of outcome, i.e., life versus death, can be made with what appears to be amazing accuracy. This allows comparative outcome records between similar institutions to call attention to possible improvements in those units that do not achieve normal predicted outcomes. As further technologies are developed with their often high initial costs and somewhat limited applications, technological assessment both before and after implementation will be a necessary task. In the past, basic research was often translated into commercial application on a broad scale without adequate clinical testing on a more limited basis.

Given our current era of limited resources, an aging population, and continued development of new and expensive technologies, it would appear prudent to assess those technologies on a limited, trial basis prior to widespread implementation. Should such a process be adopted, physician organizations such as PROs could serve as a useful, professional forum in criteria development, data collection, utilization and quality monitoring, and outcome evaluation.

NEEDS FOR THE FUTURE

In conclusion, let me briefly dwell on our perception of what is needed to improve the medical profession's ability to self-regulate. First and foremost is the need for the profession to accept and actively participate in the concept of public accountability. Health is a vital commodity of paramount importance to each of us individually and the nation collectively. The responsibility for maintaining the public's health and using new and sophisticated technology lies with the licensed professional. That privilege cannot be looked on as a coronation of professional elites with keys to the royal treasuries; it must be seen as a sacred and honored trust to be used humbly and for the public good. With that trust comes the responsibility to self-regulate to prevent abuses, assure quality, and conserve resources. Once that concept is widely accepted, true peer review will make our task easier.

In addition to wider professional acceptance of the responsibility of peer review must come broader participation in the process by the public, third-party payers, and governmental agencies. Data properly displayed and interpreted can and should be a useful tool in increasing the public's knowledge of both the potential benefits and the shortcomings of current and future technologies. Improved data collection instruments and software packages such as MedisGroups and Computerized Severity Index will make the tasks easier.

Let me conclude with a word of caution. In the struggle to conserve limited resources and assure appropriate utilization and quality, let us not strangle innovation and experimentation. We must not settle for mediocrity monitored by computers and measured against national criteria. Such an outcome would, in my opinion, stifle that creative 10 percent on whom we all depend for progress. It could in the worst case scenario lead to technology rationing on the basis of cost rather than need.

Technology Assessment: A Physician's Perspective

Earl P. Steinberg

Clinical practice is in a constant state of change. Basic and clinical research continually provide new insights into the etiology, pathogenesis, detection, and treatment of disease. In addition, hundreds of new diagnostic devices, drugs, and therapeutic procedures become available each year for use in clinical practice. The challenge to physicians individually, and to the medical profession as a whole, is not just to keep abreast of these changes but to define the optimal strategies for applying our constantly increasing diagnostic and therapeutic options to individual patient problems.

This paper defines, from a physician's perspective, the role of technology assessment in this process. In so doing, the relationships among technology assessment, quality assessment, and quality assurance can be clarified. The paper begins by defining these terms and providing a conceptual framework for understanding their interrelationships. It then describes in detail the technology assessment activities conducted by the American College of Physicians (ACP) and the implications of those activities for quality assessment and quality assurance. It concludes with a call for increased technology assessment activity involving more primary data collection efforts in addition to literature syntheses and secondary data analyses.

DEFINITIONS AND CONCEPTUAL FRAMEWORK

From the physician's perspective, technology assessment consists of three distinct but interrelated activities:

- Evaluation of the efficacy, effectiveness, and safety of emerging and established technologies in a variety of clinical circumstances.

- Determination of the appropriate (and, when possible, optimal) use of alternative technologies in specific clinical circumstances, based on performance characteristics and safety profiles of those technologies.
- Consideration of which types of physicians ought to be performing various technical procedures, based on evaluation of the knowledge and skill required to perform those procedures properly.

Quality assessment, in contrast, involves the comparison of actual clinical practice to either optimal clinical practice or some agreed on standard of clinical practice. Within this construct, quality assessment depends on technology assessment to define the optimal clinical practices or standards of practice to which actual clinical practices should be compared.

Quality assurance, in turn, incorporates both technology and quality assessment activities in a system of evaluation and interventions whose purpose is to increase the likelihood that actual clinical practice will mirror standards of practice or optimal practices once defined.

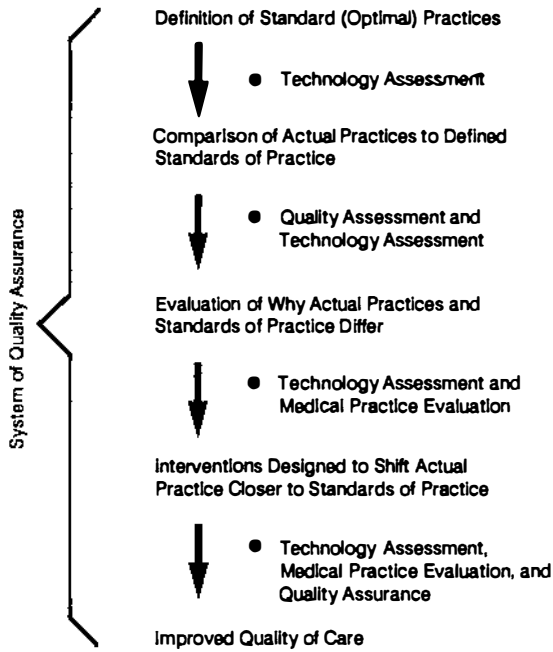

FIGURE 1 The relationship between technology assessment, quality assessment, and quality assurance.

Figure 1 clarifies these interrelationships. Through critical evaluation of individual technologies, technology assessors help define appropriate, and occasionally optimal, patterns of clinical practice. Both quality assessors and technology assessors engage in evaluations that compare actual patterns of clinical practice to the appropriate or optimal patterns of practice defined through assessment of individual technologies. When differences between actual clinical practice and standards of practice are identified, researchers engaged in technology assessment and medical practice evaluation try to identify the clinical, psychological, social, and economic reasons for the observed differences. Having considered the multiple factors responsible for observed discrepancies between actual and desired patterns of practice, researchers engaged in technology assessment and medical practice evaluation, as well as individuals engaged in quality assurance activities, try to develop and implement interventions that will move actual clinical practice closer to defined standards or optimal patterns of practice.

THE ACP's TECHNOLOGY ASSESSMENT ACTIVITIES

The ACP is a national medical society with more than 63,000 internist members; it works to promote the highest standard of health care through its activities in continuing education, technology assessment, quality assurance, and health policy analysis. Over the past decade, ACP has made a substantial commitment to technology assessment.

The Clinical Efficacy Assessment Project

The centerpiece of ACP's technology assessment activities is its Clinical Efficacy Assessment Project (CEAP), which is directed by a six-member Clinical Efficacy Assessment Subcommittee (CEAS) of ACP's Health and Public Policy Committee (HPPC). The purpose of ACP's CEAP is to improve the quality of medical practice by providing practitioners with guidelines and recommendations regarding the appropriate use of tests, procedures, and therapies and the rationale for those recommendations. CEAP focuses on diagnostic, therapeutic, prognostic, and preventive technologies and practices.

History of CEAP

CEAP had its origin in a collaborative relationship with Blue Cross/Blue Shield's BC/BS Medical Necessity Project, which began in 1976.

Together with the American College of Surgeons, the American College of Radiology, and the Council of Medical Specialty Societies, the ACP worked with BC/BS to identify tests and procedures that were no longer justified in routine medical practice. Between 1976 and 1978, 76 such tests and procedures were identified.

The process during this period was straightforward: BC/BS presented lists of procedures to ACP for review, and through a process of consensus of expert opinion, ACP identified those that were considered outmoded or useless. The practical result of this process was that reimbursement for these outmoded or useless procedures was stopped. BC/BS believes that this process resulted in a modest reduction in health care costs, although supporting data have not been published.

Although this arrangement proved beneficial to Blue Cross, ACP believed it provided little value to college members. Instead, the major challenge confronting ACP members was determining which new or established procedures to apply in specific clinical circumstances.

To broaden the scope of its technology assessment activities, ACP secured a three year grant in 1981 from the John A. Hartford Foundation and, with it, established the current CEAP. Between 1981 and 1983, this project was largely an in-house operation, with college staff performing approximately 50 technology assessments. Advice and comment on those in-house reviews were provided by subspecialty societies.

As an indication of the importance attached to this technology assessment activity, ACP began funding CEAP internally in 1984 when the Hartford Foundation grant expired. At that time, the college began contracting with outside consultants to perform its technology assessments.

In 1986 ACP resumed a collaborative relationship with Blue Cross, through which ACP undertakes a small number of technology assessments that are of particular interest to Blue Cross and which satisfy ACP selection criteria. Financial support is provided by Blue Cross to defray the costs of performing these assessments. In 1986 ACP also began a study of the feasibility and advisability of adding a prospective data collection component to its technology assessment activities. (See description of Physician Network, below.)

In total, between 1978 and 1986, the ACP performed 114 technology assessments; 78 were of diagnostic technologies, and 36 were of therapeutic technologies.

The CEAP Process

The strength and weakness of CEAP is the process it uses to assess technologies. This process has seven components, summarized in Figure 2.

1. Identification of Candidate Technologies
2. Selection of Technologies for Assessment
3. Selection of Technology Assessor(s)
4. Literature Synthesis and Secondary Data Analysis
5. Development of Background Paper and Summary Statement
6. Internal and External Review
7. Dissemination of the Technology Assessment

FIGURE 2 The Clinical Efficacy Assessment Project (CEAP) process.

The first component is identification of candidate technologies. Suggestions for technologies to be assessed arise internally from the members of the CEAS, from other ACP members and committees, and from outside organizations, such as the Office of Health Technology Assessment (OHTA), BC/BS, and manufacturers of devices, particularly of new ones.

An important dimension of the selection process is the framing of the question, or the specification of the technology, to be assessed. In the past many of CEAP's assessments have focused on all the uses of a particular technology. More recently, CEAP has focused on the application of an individual technology to a particular clinical problem. Thus, in addition to undertaking assessment of individual technologies, such as intravenous pyelography, in all of their uses, CEAP is evaluating alternative technological approaches to studying an anatomic area, such as the gall bladder, brain, spine, or carotid arteries, as well as appropriate diagnostic testing in a specific clinical circumstance, such as after an acute myocardial infarction.

Once a group of candidate technologies is proposed, final selections are based on a set of general selection criteria. These include the committee's perception of the degree of interest in the technology by practicing internists, the existing or potential volume of use, the cost of the technology, the perceived impact of the technology in terms of its benefits and its risks, and the potential impact of an ACP assessment. Relevant in this regard is whether other researchers or organizations have recently performed or are planning similar evaluations and whether new information has recently become available that creates a need for either an initial assessment or a reassessment of a previously assessed technology.

For example, ACP recently undertook a reassessment of techniques for evaluating bone mineral content, largely because of the newly available dual photon absorptiometry. Other relevant considerations in the selection process are the committee's perception of whether a technology is being inappropriately used and the feasibility of performing an assessment. Factors considered in the latter include the degree of difficulty of performing an assessment, ACP resources (which are sufficient to fund approximately 10 to 12 assessments per year), and the extent to which relevant information is available. The ACP believes that an assessment is valuable even if it leads to a statement that insufficient data exist on which to base a recommendation.

Once technologies are chosen for an assessment, the next step is to choose consultants to perform the assessments. The ACP has debated the pros and cons of choosing generalist versus specialist physicians as assessors. The advantage of the former is that they tend to be more skilled in clinical epidemiology, statistics, and decision analysis and have less of a vested interest in any particular technology. The danger in choosing only generalists, however, is that they may lack the necessary content expertise to perform an assessment. To avoid this problem, ACP has frequently teamed methodologic experts with content experts in its assessment process.

CEAP assessments consist of a detailed review and synthesis of the literature. Various techniques such as meta-analysis, decision analysis, and Bayesian probability assessments may be applied to data that are published in the literature to develop practice guidelines. By relying only on literature syntheses and secondary data analyses, however, CEAP is limited in the scope of the technology assessments it can perform.

Two documents emerge from a CEAP assessment: a detailed review of the literature in the form of a background paper, and a statement—a short summary of the background paper and the clinical practice recommendations that emerge from it. These documents include a description of the technology, its safety, data regarding efficacy and effectiveness, indications and contraindications for its performance, data on its cost, and, finally, conclusions and recommendations. Typically, conclusions and recommendations consist of suggested guidelines for appropriate use. Draft documents are subjected to detailed reviews by the CEAS itself, relevant medical societies, content and methods experts, the *Annals of Internal Medicine*, and the *Annals'* external review process.

After approval of a statement by CEAS, the statement is referred to HPPC and ACP's Board of Regents for their approval. Papers are simultaneously submitted to the *Annals of Internal Medicine* for consideration for publication. The *Annals* has an annual circulation of about 100,000. CEAP statements are also published in the *ACP Observer* (with a circula-

tion of about 60,000). Abstracts of CEAP statements appear in the *International Journal of Technology Assessment in Health Care.* CEAP statements are also mailed to a regular list of subscribers, including state and county medical societies, government organizations, third-party payers, and other policymakers. Press releases are sent to more than 700 professional and lay publications. The ACP has made available a collection of CEAP reports performed over the past several years.

A CEAP assessment takes approximately 9 to 15 months from inception to publication. Approximately 114 such assessments have been performed to date.

Impact of CEAP

The ACP has made an effort to determine the value and impact of its CEAP. One indication of its value is the interest a diverse audience has expressed in CEAP reports. Between July and December 1986, ACP received 156 orders for 1,771 CEAP reports from practicing physicians, directors of housestaff training programs, chairmen of departments of medicine, hospital quality assurance committees, third-party payers, and government agencies. A study undertaken by Blue Cross in 1983 documented that CEAP reports were being used as part of housestaff education programs, as standards of practice against which actual practice could be compared, and as the basis for reimbursement policies used by BC/BS, Medicaid, commercial insurers, and in some cases Medicare.

Relevance of CEAP to Quality of Care

CEAP has several important implications for quality assurance. First, CEAP reports are one form of physician education. ACP hopes that these reports lead to changes in practice, which in turn lead to a higher quality of care. Second, CEAP reports themselves provide physicians with training in the critical appraisal of literature. This will make physicians more sophisticated consumers of information in their own medical reading. Third, CEAP reports and recommendations set standards for practice that can be used in quality assurance and quality assessment activities. Finally, the high quality of CEAP's literature reviews sets a higher standard for the design of research and for the presentation of research results in scientific publications.

The response to a recent CEAP report on the safety and efficacy of performing cardiac catheterization on outpatients in hospital or free-standing catheterization lab facilities demonstrates how the ACP's technology assessment activities can stimulate and facilitate quality assurance activities (1). The CEAP report on this topic concluded that, for low-

risk patients, cardiac catheterization could be performed safely on an outpatient basis when done in a laboratory that is in or adjacent to a hospital. The report recommended, however, that outpatient cardiac catheterization not be performed in laboratories that are not adjacent to hospitals (i.e., in freestanding centers), since it was felt that such facilities do not provide patients with ready access to emergency care and hospitalization and that regulatory mechanisms to assure high-quality care in such facilities do not currently exist. In direct response to this CEAP report, owners and operators of freestanding cardiac catheterization facilities formed the Association of Ambulatory Cardiac Catheterization to develop quality assurance standards and a quality assurance program for freestanding ambulatory cardiac catheterization laboratories and to collect data to document the safety of outpatient catheterization.

Other ACP Technology Assessment Activities

Guidelines for Use of Common Diagnostic Tests

ACP recently extended its technology assessment activities beyond CEAP. The *Annals of Internal Medicine*, for example, published a series of articles that contained guidelines for the appropriate use of 15 commonly used diagnostic tests. This series of articles has been reprinted as a separate book (2). The series grew out of a collaborative arrangement between BC/BS's Medical Necessity Project, ACP's CEAS, and the *Annals*. The rationale for this project was the presumed overuse of commonly ordered tests. BC/BS estimated that more than $6 billion are spent annually on unnecessary tests. BC/BS's and ACP's interest in this project grew out of a belief that much inappropriate use emanates from a lack of understanding regarding appropriate use. The project's goal was to produce guidelines that would both increase the quality and decrease the cost of care. The desire, in essence, was to change physicians' behavior so that instead of ordering a test unless it is contraindicated, physicians would order a test only if it is indicated.

Fifteen commonly used tests were chosen for analysis, including routine chest radiography; routine admission preoperative electrocardiography; laboratory tests, complete blood count and differential; serum electrolytes; osmolality, blood urea nitrogen and creatinine; biochemistry profiles in ambulatory screening and preadmission testing; urinalysis and culture for diagnosis of infections in women; prothrombin test and partial thromboplastin time; throat cultures and rapid tests for diagnosis of streptoccocal pharyngitis; syphilis testing; arterial blood gasses; erythrocyte sedimentation rates; carcinoembryonic antigen; cardiac enzymes and

blood cultures. Outside consultants were hired to perform detailed literature reviews on each of these topics. These reviews resulted in the publication of background papers and a set of guidelines that can serve as standards for determination of quality assurance and quality of care.

Reaction to these guidelines has been both positive and negative. ACP views the guidelines as important educational tools to guide appropriate clinical practice. Others, however, have criticized the guidelines as not being "in the interest of patients" and as "invading medical judgment." Concerns were also expressed that the guidelines would lead to increased malpractice litigation. Reports about the guidelines were published in the *New York Times*, *Los Angeles Times*, and *Wall Street Journal*.

National Physician Network

A second new venture being pursued by ACP involves exploring the establishment of a national network of physicians who would systematically collect information regarding the use and impact of selected tests and procedures in routine ambulatory practice. Like that of CEAP, the purpose of this effort is to develop guidelines to enhance the quality of care while containing its costs. The ACP initiated a feasibility study of a physician's network in September 1986 with the help of a grant from the Henry J. Kaiser Family Foundation.

The rationale for the project is that the prevalence of disease, as well as the sensitivity and specificity of various tests, likely varies in community and university settings; yet most research is performed in university centers. In addition, known variations in physicians' practices provide an opportunity to observe the results of natural experiments, which will help ACP determine the relationship between use of particular tests and patient outcomes. If ACP decides that the project is feasible and worthwhile, it will proceed with a strategy focused on selected tracer conditions or problems. An effort will be made to determine the frequency and variation of use of various tests, procedures, and types of referrals and to relate this variation in use to outcomes of care. The ACP's ultimate goal is to develop education and feedback interventions to modify physician practice.

Clinical Privileges Project

Another activity of ACP that has implications for both technology assessment and quality of care is ACP's clinical privileges project. In this project ACP is writing guidelines for 70 procedures in internal medicine. Guidelines will consist of statements regarding the required knowledge, technical skills, and training necessary to establish initial competence and to maintain it.

CONCLUSION

Over the past several years ACP has tried to improve the quality of medical practice by commissioning and overseeing technology assessments that consist of critical appraisals of the published literature. Although this activity can and does guide physicians in their clinical practice, and serves as the basis for some quality assessment and quality assurance activities, its impact is limited for two reasons. First, it is a small-scale activity, consisting of only 10 to 12 assessments per year. Second, it consists exclusively of secondary data analysis with much, if not most, of the data collected in university medical centers.

To address the latter problem, ACP is exploring the feasibility of extending its technology assessment activities by establishing a network of physicians who would collect data in their offices. These data could be used to evaluate the relationship between diagnostic and therapeutic practices and the outcomes of patient care. In addition, ACP is working to establish guidelines on the clinical competence required to perform a variety of technological procedures. Ultimately, to meet the goal of providing high-quality, cost-effective care to patients, the medical profession and society as a whole must commit increased resources to technology assessment activities that involve collection of new data as well as analysis and synthesis of existing data.

REFERENCES

1. Health and Public Policy Committee, American College of Physicians. The safety and efficacy of ambulatory cardiac catheterization in the hospital and freestanding setting. Annals of Internal Medicine 103:294–298, 1985.
2. Sox, H.C., Jr., ed. Common Diagnostic Tests, Use and Interpretation. Philadelphia, American College of Physicians, 1987.

Technology Assessment: View of a Multihospital System

John H. Moxley III

I shall discuss technology assessment from the view of a multihospital system: How might such systems benefit from the type of technology assessment under consideration by the Institute of Medicine's (IOM) Council on Health Care Technology? What have such systems to contribute to the assessment programs that are being considered? Finally, how are hospitals' technology assessment efforts related to quality assurance programs, if, indeed, such a relationship exists. I shall also pay particular attention to how quality assurance programs affect multihospital systems.

In general, I will use American Medical International (AMI), the company that I am associated with, as an example. I think it is safe to assume, however, that similar technology assessment activities are going on in other multihospital systems to one degree or another. In fact, it is increasingly difficult to define a multihospital system; several individual hospitals have branched out and become mini systems. Most of what I say applies not only to the investor-owned hospital companies but also to not-for-profit hospital systems and corporately reorganized individual hospitals.

CHANGING ENVIRONMENT OF HEALTH CARE

Certainly, a great deal has changed within the investor-owned health care companies since I first met with the IOM's Council on Health Care Technology some 13 months ago. The combined effects of several years of very rapid growth and diversification plus the introduction of competition and prospective pricing has brought relatively hard times to the investor-owned hospital community.

Perhaps the effects of the recent changes in the health care system

are more apparent in the investor-owned companies because they are publicly owned and, therefore, required to make regular reports to the public. But, as I travel around the country, my sense is that not-for-profit hospital systems are experiencing similar difficulties. At AMI the changes have encouraged a significant reexamination of every aspect of the company. The reassessment effort has concluded that the company must reconcentrate its efforts on its core business—the ownership and operation of acute care community hospitals. In my judgment all the systems are restructuring to some extent. Some are moving away from acute care hospitals; others are refocusing on them.

At AMI, at any rate, the significant reorganization has led to a very different view of technology than we had in the past. Just a few years ago, the only questions about new technology were: Did the physicians want it? Was it legal, that is, had it been approved by the Food and Drug Administration or other appropriate regulatory body such that it could be introduced into clinical practice?

In fulfilling those criteria, our system alone purchased, on the average, $500 million a year in new technology; this included intravenous solutions and x-ray film, which others might not consider technology. The purchases also included construction funding necessary to house the new technology. The $500 million, however, was exclusive of the $50 million in drugs consumed by our system annually.

Technology assessments and purchases were, for all practical purposes, carried out at the local facility level. Arrangements were negotiated by the hospital administrators, their staff, and the local sales representatives of the manufacturers. The hospital directors in our system were told that physicians were their customers and the route to success was to keep their customers happy.

FIRST STEPS: CHANGING TECHNOLOGY ASSESSMENT STRATEGIES AT AMI

Beginning in 1982, when the storm warnings were first seen on the horizon, we began to codify our technology assessment and purchasing systems with a primary effort toward improving the purchasing function. We created a materials management group in the corporate office and instructed them to centralize technology assessments and to develop, wherever possible, national purchasing arrangements. Efficient purchasing was the emphasis; technology assessment was swept into it. We were fortunate enough to recruit some very capable and experienced materials managers, and they, largely on their own initiative, began to look at technology assessment issues.

The new group began to receive and appraise all requests for new technology. If a hospital requested a widely used technology, they carried out some relatively simple tasks. They surveyed the literature and gathered the advice of several consultants. On the basis of the gathered information, they made a decision whether to buy X or Y technology.

MORE FORMAL EXPERT PANELS

Soon it became clear that the process needed to be further formalized. Panels of experts with considerable experience were organized from among physicians and the other health care professionals within our system. The panels were asked specific questions about alternative types and brands of technology. The process has progressed to where we have standing committees representing most of the specialties and specialized areas of care, such as intensive care and emergency medicine.

The review committees can be asked very broad questions, such as what the company should do about magnetic resonance imaging (MRI) or lithotripsy. Or, the panels can be presented with much narrower questions, such as which of four or five brands being requested ought to be bought. Specific to each inquiry is the aspect of quality assessment. This is one of the few ways in which technology assessment directly intertwines with quality assurance.

The expert panels are specifically not given cost information, and they are instructed that cost is not to be one of their considerations. If they reach a clean decision about which brand to buy, the materials management group negotiates a national contract for the purchase.

If, however, the committee divides or cannot reach consensus, we conduct a field trial. As an example, we had recently been reviewing IV controllers and pumps. There are about five products that physicians in various parts of our system were anxious to have. The review committee could not reach a clear choice. Consequently, we tested each of the products in several of our hospitals and through that process were able to come up with a product that seemed to be more acceptable than the others. The decision was made, therefore, to purchase the product judged to be superior.

Once the technology assessment is completed, we move forward to negotiate with the manufacturers. I will only point out that the negotiations are complex. We are a large purchaser; therefore, we have to worry about whether the manufacturer can keep up with our needs, whether the manufacturer offers the proper training for our people, and so forth.

Despite the elaborate process that has evolved over the past four or five years, at times a user group is unhappy about the decision reached. When that happens, and it has happened two or three times, we go to a blinded trial, often using the group that is most unhappy. Such a situation came up when we decided to purchase a certain brand of x-ray film. At that time, a group of radiologists insisted we were going to buy an inferior product to save money. They were, however, willing to engage in a blind trial. The group selected the film from the new company. That put the issue to rest, and we went forward and negotiated a contract.

Drugs

I am not going to comment extensively on how we handle drugs. I realize the Council on Health Care Technology considers drugs as technology. We do also. I will make just three brief comments because they reflect what we are doing in the general area of technology assessment.

We created a corporate pharmacy group, which carries out many of the same functions as the corporate materials management group. We are now developing a systemwide formulary to limit duplication. At first we thought a systemwide formulary would be viewed as contentious by the physicians who use our hospitals. In reality, however, we have received excellent cooperation, and development has moved smoothly. We are also in the process of introducing clinical pharmacists into the AMI system. We have found that clinical pharmacists improve quality of care and concurrently reduce costs.

PRESENT AND FUTURE RESPONSES TO CHANGES IN HEALTH CARE

Since our technology assessment and purchasing systems have been introduced, a number of changes have occurred in the health care world. As one consequence, our overall expenditures for technology, both replacement and new, must be not only contained but also markedly reduced. Hospital Corporation of America has announced that they have reduced their capital expenditures by 50 percent, from $1.4 billion in 1985 to $700 million in 1986. We have reduced our capital expenditures by at least that percentage. I am certain that similar reductions are taking place throughout the hospital world. Although I have some serious concerns about what might happen if hospitals are forced to continue to reduce capital expenditures, I will put them aside. Suffice to say that all the pressures on hospitals at the present time are going to force them to be increasingly conservative purchasers of new and replacement technology.

COLLABORATING WITH NATIONAL TECHNOLOGY ASSESSMENT PROGRAMS

How might AMI and other hospital systems benefit from collaborating with the national efforts being discussed in this volume? The answer is as obvious as it seems. Assuming that the assessments were performed in a manner equal to or better than we do in our system and the cost to our system was of approximately the same magnitude, it would be easier and "cleaner" for us to collaborate with national efforts than to continue assessing independently. I say "cleaner" because no matter to what lengths we go to assure objectivity, someone is going to charge that our primary interest is cost, not quality and, therefore, call our process flawed. Hence, if we could collaborate with regional or national efforts, we would protect ourselves from that criticism. It is also likely that having a wider breadth of input into the assessment process, even better decisions would be reached.

CONTRIBUTING TO NATIONAL TECHNOLOGY ASSESSMENT PROGRAM EFFORTS

What can we contribute to a national effort? We have a large hospital system of more than 90 hospitals, populated by health professionals. Further, we have a very good materials management group that has considerable experience in evaluating both established and new technologies.

Our existing technology assessment system might be used by a national effort. Alternatively, we might have to change some of our processes. But it seems it would be easier for us to adapt than for new systems to be cut from whole cloth. Trying to put together either a system or a group of hospitals that have not collaborated and, therefore, cannot move in a systematic way is difficult. We had to do it. Therefore, using existing systems may have mutual benefits—to technology assessment and to us. In addition, we have accumulated a significant amount of technology assessment data over the past few years which we could share with a national effort.

We can also share with national efforts some opinions we have developed about technology assessment methods. For example, our assessment people believe that it is very important not to identify members of technology assessment panels. They found that when panel membership is known, members are lobbied and pressured by their physician and nurse colleagues across the country to select certain types or certain brands of technology. No matter how well intentioned, the lobbying is both time consuming and disruptive. Therefore, the specific identity of

review committee members is not revealed within our system. Their backgrounds and credentials are disclosed but not their names.

In appointing a review committee we try to choose the most skilled individuals to whom we have access. We think that we have made some significant strides in recent years. We hope that we have progressed enough to withstand some of the increasing pressures on the health care system.

Some further specific suggestions from our materials management group include the need to become much more proactive in general in technology assessment. They are concerned that it is increasingly difficult to separate the "bells and whistles" from the technology per se. With greater frequency, particularly in our smaller community hospitals, we are forced to purchase machines with capacities far in excess of those needed to do the job. It is increasingly difficult to find a scaled-down model that can do 500 procedures in an eight-hour shift rather than a machine that does 5,000 tests per eight hours, whether needed or not. The manufacturers are increasingly in control. The provider has to buy the technology and often that technology is more than is needed to accomplish the task.

TECHNOLOGY ASSESSMENT ISSUES FOR THE COUNCIL ON HEALTH CARE TECHNOLOGY

Recently we brought together a group of our hospital managers and asked them some questions about technology assessment. They were asked to assume that they had an opportunity to tell the IOM's Council on Health Care Technology about the issues important to them. They produced the following list of issues.

First, hospital managers would like to have more information on how new technology will affect existing technology and how new technology will influence the management of disease. For example, is the technology going to introduce an entirely new therapeutic approach? If so, will the new technology force a change in the treatment site? How can cost-effectiveness be more precisely measured? Hospital managers think that they constantly have to balance cost, technology assessment, and quality of care and come up with what is being termed cost-effectiveness. They would like as much information about each of the variables in that complex equation as possible.

Second, hospital managers would like specific information on where the technology should be sited. They are under a lot of pressure to introduce new technology but have little information about the proper treatment site. How are they to come to grips with the situation intelligently?

Third, guidelines for dissemination of technology are also needed. AMI is primarily in the community hospital business. When is it appropriate for a community hospital to introduce a technology as compared with a tertiary care center or a freestanding ambulatory center?

Fourth, better estimates of generation turnover time are needed. We mean, by this, turnover time based more on true technology innovation and less on the "bells and whistles."

Finally, hospital managers are not happy with the Health Care Financing Administration's (HCFA) payment approval process. They would like to see the HCFA's approval process streamlined and based on a more realistic payment level for a proven new technology. Their consensus, shared by many, is that the current process is based largely on cost containment foot-dragging, not expeditious technology assessment.

In my judgment the issues raised by our hospital managers reflect the fact that traditional capital budgeting has finally come to the hospital industry. In the past, technology purchases were largely based on manipulating the reimbursement formula. Reimbursement formulas were works of art, not analytical conclusions. Hospital managers are under increasing pressure to justify capital budgeting by deciding in a more traditional way whether a piece of equipment ought to be purchased and how it will be paid for over time. The changes in the capital budgeting process are coming to the not-for-profit sector as well. Prospective pricing places similar demands on all providers. If a piece of technology cannot pay its own way, it is going to be more difficult for a hospital, under any form of ownership, to purchase it.

QUALITY ASSURANCE AT AMI

How does AMI do quality assurance? Beginning in the 1970s, we began the following process. The chief operating officer began to make one-day, biennial site visits to each of our hospitals. He was accompanied by a medical records person and a nurse. They inspected the facility, and they interviewed the administration, the hospital department managers, the physicians, and hospital board members. They issued a 20- to 30-page report highlighting the strengths and weaknesses of the physical facility and the staff. Specific suggestions for improvement were made and followed up.

The site visits look at the delivery process but not at clinical outcomes. They do, nevertheless, contribute to quality assurance. The correlation with technology assessment is feeble but does exist. During the visit, the medical staff often made a case for new technology. A more stringent justification is required at the present time, however, than was a few years ago.

During the last few years, AMI has taken two additional quality assurance steps. First, about two years ago, Dr. Robert Dubois, a Robert Wood Johnson Clinical Scholar, came to see me. Dr. Dubois is an internist and was working with Dr. Robert Brook at The RAND Corporation as well as taking some courses at UCLA. He was looking for a system in which to test some quality assessment ideas. AMI welcomed him and helped to fund his study. Dr. Dubois has devised a process for carefully adjusting hospital mortality rates. The process has been applied to the hospitals in our system. The study then looked at the outliers—hospitals that appear to have lower mortality rates than predicted and those with higher than predicted mortality rates. Dr. Dubois then paid a site visit to each of the hospitals and did a very precise chart review for a set of conditions. The chart review results were tested against the judgment of a panel of consultants. We believe that the results will be a significant contribution to quality assessment literature (1,2).

Clearly, AMI's goal was to come up with a system whereby individual hospital managers and system managers could, quickly and frequently, have a printout of quality assessment data across the system. This information must be dependable enough to draw conclusions and to initiate intervention in a particular hospital. Such a process would allow managers to avoid those unfortunate events that occur when hospital quality for one reason or another deteriorates. We are not there yet. But I believe that Dr. Dubois and the study group have taken a significant step in the right direction.

Second, AMI organized an Office of Medical Affairs last year. The office is the focal point for input by our medical staffs. The office has begun organizing groups of our physicians into task forces and asking the task forces to examine issues such as accreditation and physician discipline.

AMI has not done this before, but we believe that we must involve our medical staff much more than in the past as we address issues such as physician discipline in small community hospitals. When only five or six physicians are in a community, they are not only professional colleagues, they are socially intertwined as well. Physicians in smaller communities have often known each other for several years. Introducing new review and discipline practices into such a community is, at best, extremely difficult.

One of the things we must begin to do is relate our technology assessment and our quality assurance programs. In doing so we can expect to get broader and more precise information and improve the quality of care throughout our system. It is my belief that most systems are moving in the same directions—each with its own twists but in the same general direction.

CONCLUSIONS

In conclusion, let me quote—as many of us in medicine frequently do—from Lewis Thomas. Dr. Thomas noted in one of his essays that "it is a gamble to bet on science for moving ahead, but it is, in my view, the only game in town"(3).

I confess that I subscribe to Thomas's view. I believe that if we are to experience, in the future, the single most important characteristic of twentieth century American medicine—its capacity for scientific improvement and technological adaption—we are going to have to stop creating adversarial relationships between the various constituencies involved in the provision of health care. We are going to have to stop increasing the fear level that I sense among physicians and hospital managers and many others. We must realize that it is essential that both the public and private sectors and the professional and lay communities appreciate the importance of the health care mission. Only then can we work together to tackle some of the critical technological assessment and quality-of-care issues and, thereby, preserve what has made our health care system great and a system that we should all be proud to perpetuate.

REFERENCES

1. Dubois, R.W., Brook, R.H., and Rogers, W.H. Adjusted hospital death rates: A potential screen for quality of medical care. American Journal of Public Health 77:1162–1167, 1987.
2. Dubois, R.W., Brook, R.H., Moxley, J.H., et al. Hospital inpatient mortality: A predictor of quality. New England Journal of Medicine 317:1674–1680, 1987.
3. Lewis, T. Making science work. Late Night Thoughts on Listening to Mahler's Ninth Symphony. 1981.

Technology Assessment: Perspective of a Third-Party Payer

Ralph W. Schaffarzick

FOUNDATIONS OF TECHNOLOGY ASSESSMENT IN A STATEWIDE PHYSICIANS' SERVICE PLAN

On February 2, 1939, almost a half century ago, the Articles of Incorporation of California Physicians' Service (CPS) were filed with the Secretary of State. An excerpt from the first paragraph of that document follows:

> We, the undersigned . . . active members of the California Medical Association . . . do hereby declare that the duties and obligations of the profession are not only leadership in the maintenance of high standards of medical service but also in the means of distribution of that service so that all who need it may receive it; that the very advances made by modern science have greatly increased the cost of good medical service and hospital care and will continue to increase that cost as new methods and equipment for diagnosis and treatment are discovered and perfected.(1)

Thus, the founders of the nation's first nonprofit statewide physicians' service plan recognized the importance of making quality health care available "to all who need it." Further, they anticipated the increasing impact of technology on the cost and quality of health care. To monitor and evaluate the "new methods and equipment for diagnosis and treatment," a major permanent committee of the Board of Trustees was established: the Medical Policy Committee (MPC).

In those "innocent" early days the proceedings of the MPC were relatively simple and dealt primarily with subjects concerning standards of medical practice such as the role of consultants, frequency of hospital visits, and "quack medicine." As the surge of medical technology began to gather momentum during the late 1960s, it became apparent that the MPC would need to develop the ability to decide which new procedures,

devices, drugs, and biological products should be recognized for payment. In its role as "prudent purchaser" of services for Blue Shield of California (BSC)/CPS subscribers, the MPC gradually realized that it needed to identify truly valuable technology and recommend that it be considered for coverage.

At that time, BSC/CPS was still closely related to the California Medical Association (CMA). Quite naturally, the MPC sought the advice of the CMA's Scientific Board. Panels of physicians on the Scientific Board were asked to provide opinions concerning the validity of new technologies. Such opinions were based on personal experience or that of their colleagues. Opinions were expressed in terms such as "good," "bad idea," "promising," or "experimental."

Gradually, it became apparent to the MPC that a more rigorous approach must be employed. This recognition was given considerable impetus by the publication of *Costs, Risks, and Benefits of Surgery*, edited by Barnes, Bunker, and Mosteller (2). Increasingly, the MPC required evidence of safety and efficacy, such as published reports of randomized controlled trials or double-blind studies. When Dr. David Eddy introduced us to the discipline of decision analysis for special topics, the process of technology assessment (TA) became more structured and critical.

Even so, the discipline was far from mature when in 1984 Finkelstein, Isaacson, and Frishkopf (3) studied the evaluation processes used by the BSC/MPC and by the Health Care Financing Administration (HCFA). They found HCFA's Medicare evaluations to be:

> highly structured, synthesizing thorough literature reviews, recommendations from the National Institutes of Health (NIH) and other governmental agencies.... In contrast, BSC's evaluations were based largely on presentations and discussions at advisory committee meetings, after receiving informational inputs that were more limited than those used in Medicare evaluations. The fraction of technologies recommended for coverage was slightly more than 50 percent for each carrier.... Still the carriers differed significantly in the stage of development of the practices evaluated and in their willingness to make a coverage decision in the face of both safety and effectiveness data that were regarded as tentative.... While there were meaningful differences between the processes of the two carriers that were studied, these carriers are probably two of the most competent groups conducting this type of evaluation.

In 1982 Bunker, Fowles, and Schaffarzick (4) reported their evaluation of medical technology strategies. They found a lack of consistent and explicit policy on payment for new technologies; an absence of a single, organized, adequately funded program or agency responsible for generating data with which to evaluate technologies; difficulty in establishing priorities for study; and a number of conflicts among the goals of insurers, practitioners, developers, evaluators, and government. To deal

with this situation, the authors proposed the creation of a private, nonprofit institute for health care evaluation, to be composed of representatives from both the private and the public sectors.

The creation of the Institute of Medicine's Council on Health Care Technology represents a major accomplishment, which is welcomed by all of us interested in this important endeavor.

THE PROCESS OF TA AND COVERAGE DETERMINATION OF BCS/MPC

General Procedures

Currently, the MPC, a major committee of the Board of Directors of BSC, comprises 21 members. Fifteen are members of the board, 11 are medical doctors, and 10 are "public" members. The latter include hospital administrators, a health policy economist, former state and congressional legislators, a professor of philosophy, and employee benefits managers for large corporations.

The MPC convenes three to four times a year, meeting in various cities in the state. Invited guests include authorities on the topics being considered, representatives of specialty societies, other third-party payers, including the California Department of Health Services, and HCFA. A unique feature of the meetings is that they are held in open forum. Attendance numbers from 70 to 100 interested parties, including the press and even television.

In preparation for the meetings, new technologies are identified by the medical director and the medical policy administrator. Literature searches are conducted. Opinions of specialty societies and recognized authorities are solicited. Analyses performed by other TA entities, such as the Clinical Efficacy Assessment Project of the American College of Physicians or the Office of Health Technology Assessment of the National Center for Health Services Research, are included. When all available evidence has been assembled, it is analyzed according to criteria developed by the Blue Cross and Blue Shield Association (BCBSA) (5):

- The technology must have final approval from the appropriate government regulatory bodies.
- The scientific evidence must permit conclusions concerning the effect of the technology on health outcomes.
- The technology must improve the net health outcome.

- The technology must be as beneficial as any established alternatives.
- The improvement must be attainable outside investigational settings.

Agenda books, containing the pertinent literature, letters, claims samples, and analyses, are mailed to the attendees well in advance of the meeting. After full and open discussion, including oral testimony of authorities and other guests, the MPC votes on the recommendation of the medical director.

A technology can fall into three categories: experimental, investigational, or acceptable medical practice. Only those technologies achieving the latter category are eligible to be considered for coverage.

In determining coverage, other factors are considered: cost, cost-effectiveness, accessibility, legal and contractual implications, and moral/ethical problems, especially that of distributive justice.

The decisions of the MPC are presented to the Board of Directors, which can accept them or return them to the MPC for reconsideration. After acceptance, the decisions are announced to physicians, hospitals, and other interested parties and are implemented in the claims-processing system.

Both new and established technologies are subjected to periodic review. As further evidence is acquired, an investigational procedure may be promoted to acceptable practice or an established procedure may become obsolete.

Selective Coverage

As technologies became more complex, a new concept in TA/coverage determination emerged. In 1984, for example, the MPC determined that, although human heart transplantation was generally still investigational, in the experience of the Shumway group at Stanford it had become acceptable practice. The MPC decided, therefore, that heart transplants would be covered when performed at Stanford or at any other center that could document the fact that it possessed the skills, resources, commitment, and record of favorable outcomes comparable with those of Stanford. Specific criteria were established for the measurement of these qualities. Subsequently, this concept of selective coverage has been applied to human liver transplant, bone marrow transplant, and other procedures.

Selective coverage promotes high quality of care, accessibility of valuable technology to BSC subscribers, and regionalization of resources

for complex procedures. In the case of human organ transplantation it also inhibits insalubrious competition for donor organs.

Modified Selective Coverage

Not infrequently, BSC receives claims or petitions for coverage of procedures and devices that are still in the investigational stage but have the potential for improving clinical outcomes at lower risk and lower cost. The scientific rationale may be valid, and preliminary clinical trials may be very encouraging. On the other hand, when a device is involved, it may not yet have received marketing approval from the Food and Drug Administration (FDA), and reports of sufficient clinical trials may not have been published in peer-reviewed journals. As a result, the technology must be classified as investigational and, therefore, not be covered.

Even so, some new procedures and devices are not as dramatic as heart or liver transplantation. A specific example of such an emerging technology is the ablation of abnormal conduction foci in the interior of the heart by means of electroshock delivered to the lesion through a heart catheter. The standard, established treatment for such lesions is open heart surgery with surgical destruction of the offending tissue. If the treatment can be accomplished through a cardiac catheter, obviously the patient is not subjected to the risks, trauma, prolonged hospitalization, and cost of open heart surgery.

Clinical trials of the catheter technique are in progress in several centers. The results so far are encouraging, but they are not yet sufficient to assess fully the risks and benefits of this alternative approach. Another problem is that although the catheter has FDA approval for the purpose of electrophysiological mapping study (EPMS), it has not been approved for the ablation procedure. Put another way, the catheter has been approved for purposes of locating the trouble but not for correcting the trouble. As a consequence, BSC can pay for EPMS, for open heart surgery to destroy the lesion(s) identified by EPMS, and for implantation of an automatic cardioverter-defibrillator, but not for catheter ablation.

Although the catheter ablation technique is successful in many cases, it does not always work, and the patient then must undergo open heart surgery. Thus, although catheter ablation may sometimes replace the more formidable and expensive surgery, in other cases it may prove to be an additional source of risk and cost.

To promote the perfection of a medical technology that has demonstrated genuine potential for producing clinical benefit, reduced risk, and lower cost, but that has not yet emerged fully from the investigational category, the concept of a modified form of selective cov-

erage was considered by the MPC. (Probably the true origin of the concept is found in the Code of Hammurabi.[1])

When approached by an investigator of a procedure in this category, BSC might respond by suggesting an agreement to cover the professional and institutional costs of the procedure on eligible BSC subscribers when the clinical outcome is successful. If, on the other hand, the outcome is not successful, the investigator and hospital would absorb the cost and charge neither BSC nor the subscriber for any of the services related to the procedure.

In the case of cardiac catheter ablation, for example, if the technique corrected the cardiac arrhythmia, BSC would pay. If, however, the heart lesion failed to respond and open heart surgery was required, the doctor(s) and hospital would charge no one for the attempted catheter ablation.

The MPC has given permission to the medical director and management to proceed to a trial of the modified selective coverage concept.

CONCLUSIONS

Especially during the past 15 years, the discipline of TA has become more rigorous, structured, and sophisticated in evaluating both new and established procedures. Based on this more critical foundation of TA, providers and recipients of health care are able to make more rational diagnostic and therapeutic decisions. Good TA also provides third-party payers, both private and governmental, with an instrument for coverage (payment) determination. In the case of more complex technologies, such as human organ transplantation, the principle of selective coverage may be used to make investigational procedures available in appropriate circumstances. Selective coverage and modified selective coverage also may assist in the funding of worthy clinical investigation while enhancing the quality of care.

REFERENCES

1. Articles of Incorporation. California Blue Shield, California Physicians' Service, amended April 1944.

[1] "If a surgeon has operated with the bronze lancet on a patrician for a serious injury, and has caused his death, or has removed a cataract for a patrician, with the bronze lancet, and has made him lose his eye, his hands shall be cut off." Code of Hammurabi, 1950–1900 BC (6).

2. Bunker, J.P., Barnes, B.A., and Mosteller, F. Costs, Risks, and Benefits of Surgery. New York, Oxford University Press. 1977.
3. Finkelstein, S.N., Isaacson, S.M., and Frishkopf, J.J. The process of evaluating medical technologies for third-party coverage. Journal of Health Care Technology 1:89–102, Fall 1984.
4. Bunker, J.P., Fowles, J., and Schaffarzick, R.W. Evaluation of medical technology strategies. New England Journal of Medicine 306:620–624, 687–692, 1982.
5. Technology Assessment Criteria. Blue Cross and Blue Shield Association, 1987. (See Appendix.)
6. Knoles, G.H., and Snyder, R.K. Readings in Western Civilization, pp. 3–7. Chicago, Lippincott, 1951.

APPENDIX

National Association Technology Assessment Criteria

The first step in determining eligibility of a medical procedure for coverage is evaluating its health effects, a process known as technology assessment. The Blue Cross and Blue Shield Association conducts such assessments of selected technologies. The Association's role is informational. Each Blue Shield and Blue Cross Plan makes its own coverage decisions. Plans may consider assessment results in their coverage decisions and any additional factors they deem appropriate. The Blue Cross and Blue Shield Association uses the criteria below to determine whether a technology improves health outcomes such as length of life, ability to function, or quality of life. Technologies that meet all five of the following criteria are recommended for coverage consideration.

1. The technology must have final approval from the appropriate government regulatory bodies.

- A device, drug, or biological product must have Food and Drug Administration approval to market for those specific indications and methods of use that Blue Cross and Blue Shield Association is assessing.

- Approval to market refers to permission for commercial distribution. Any other approval that is granted as an interim step in the FDA regulatory process, e.g., an investigational device exemption, is not sufficient.

2. The scientific evidence must permit conclusions concerning the effect of the technology on health outcomes.

- The evidence should consist of well-designed and well-conducted investigations published in peer-reviewed journals. The quality of the body of studies and the consistency of the results are considered in evaluating the evidence.
- The evidence should demonstrate that the technology can measure or alter the physiological changes related to a disease, injury, illness, or condition. In addition, there should be evidence or a convincing argument based on established medical facts that such measurement or alteration affects health outcomes.
- Opinions and assessments by national medical associations, consensus panels, or other technology assessment bodies are evaluated according to the scientific quality of the supporting evidence and rationale.

3. The technology must improve the net health outcome.

- The technology's beneficial effects on health outcomes should outweigh any harmful effects on health outcomes.

4. The technology must be as beneficial as any established alternatives.

- The technology should improve the net health outcome as much as or more than established alternatives.

5. The improvement must be attainable outside the investigational setting.

- When used under the usual conditions of medical practice, the technology should be reasonably expected to satisfy criteria 3 and 4.

Adopted by Blue Shield of California on March 4, 1987.

The Challenge to Technology Assessment: An Industry Viewpoint

Morton L. Paterson

ASSESSMENT GAP

Assessment of drugs, unlike other areas of medical technology, routinely makes use of look-alike control treatments, both active and placebo. This favors the randomized, double-blind trial that scientists prefer. In this sense drug companies have for years been practicing technology assessment, par excellence.

Drug studies have been limited, however, to matters of efficacy and safety. Efficacy trials of drugs are usually done under ideal experimental conditions. Outcomes under "real-world" conditions in the community—the word "effectiveness" has been designated for that—are rarely studied. Moreover, efficacy has been typically defined in as physiological or anatomical terms as possible, such as lesions in peptic ulcer disease, wedge pressure in congestive heart failure, and joint counts in arthritis. Effect of treatment on the patient's overall health status or quality of life has generally not been measured.

This gap in assessment has become important recently because the research-based pharmaceutical companies are increasingly faced with restrictive formulary and reimbursement policies by provider groups—health maintenance organizations (HMOs), Medicaid, hospitals, national health plans abroad—aimed at controlling costs. Cost-effectiveness has become an important part of drug assessment. New kinds of studies, beyond traditional efficacy and safety trials, are needed to assess cost and cost-effectiveness. Unfortunately, other than health economists, few persons readily grasp what cost-effectiveness in medicine means.

COST REDUCTION: THE CASE OF CIMETIDINE

To most provider groups and most industry people, cost-effectiveness means simply cost reduction. A treatment that is priced higher than

others but reduces the total costs of treatment or of a disease is clearly a net good, as long as therapeutic outcome does not suffer. In late 1976 Smith Kline and French introduced cimetidine, a research breakthrough, which improved outcomes in peptic ulcer disease although it cost more than earlier agents. We thought it would reduce the total cost of ulcer disease.

As a first step, we conducted clinical trials. Tables 1 and 2 show results reported by the investigators after a year-long trial of ulcer treatment in Sweden (1). Of the 32 cimetidine patients, one missed work "last week" versus 23 of the 36 patients in the placebo group (Table 1). Far fewer patients maintained on cimetidine lost workdays, 2.8 versus 49

TABLE 1 One-Year Maintenance Trial of Cimetidine: Work Loss Outcome

Work loss outcome	*Cimetidine 400 mg twice daily (32 patients)*	*Placebo twice daily (36 patients)*
Number missing work	1	23
Workdays lost	79	1,405
Average workdays lost per patient	2.8	49

TABLE 2 One-Year Maintenance Trial of Cimetidine: Clinical Outcome

Clinical outcome	*Cimetidine 400 mg twice daily (32 patients)*	*Placebo twice daily (36 patients)*
With recurrences	6 (19%)	30 (83%)
With 2 recurrences	1 (3%)	12 (33%)
With complications	0	4 (11%)
Receiving surgery	1 (3%)	15 (42%)

workdays lost for the group taking placebo. These results suggest a productivity gain for society. The study also found less hospitalization in the cimetidine group, as seen in Table 2. Only 3 percent of patients in the cimetidine group versus 42 percent in the placebo group went to surgery. This has obvious economic implications for health care costs.

This trial was a standard randomized trial of cimetidine versus placebo done for medical, not cost-assessment, purposes. It asked whether cimetidine causes healing of the ulcer, or prolongs the healed state, as determined by endoscopic examination. The patients were carefully

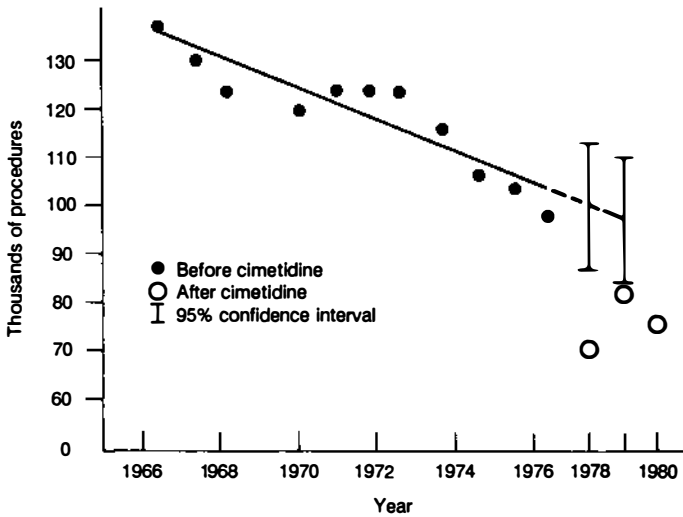

FIGURE 1 Partial gastrectomy and vagotomy surgery in United States, 1966–1980.

selected and monitored. More important, comparison was made against placebo, partly because no anticholinergic or antacid regimen had been proven effective at that time. Thus, the clinical trial was only suggestive of what cimetidine might accomplish in real-world practice. The comparative results obviously could not be translated to specific percentages of reduced work loss or hospitalization in the community. From a cost-benefit viewpoint, the value of the trial came later, when studies of the natural population supported observations made during the clinical trials.

After the market introduction of cimetidine, we looked for associated changes in statistical series. Fortunately, the acceptance of cimetidine was rapid and extensive. In the United States it was used in more than 60 percent of visits to physicians for ulcer within six to nine months after introduction of the drug. The uptake was similar in other countries. This allowed the possible effects of cimetidine to be detected by time-trend analysis.

One observation was a drop in the frequency of ulcer surgery. The United States data, collected by Fineberg and Pearlman (2), reflect the downward trends observed in several other countries where data were available: the United Kingdom, France, the Netherlands, and Sweden. Figure 1 shows the number of partial gastrectomies and vagotomies—the only operations for ulcer of any frequency—done each year in the United States. Note the downward trend well before the introduction of cimetidine. Cimetidine, however, was associated with a sharper-than-

expected drop; the expected trend line after introduction would fall within the brackets, whereas the observed number of operations fell below the brackets. Subsequently there was a rebound, then a leveling off. The 1983 through 1985 figures were found to be at approximately the 1982 level (H.V. Fineberg, personal communication). Similar downward trends were observed in the other countries (3).

Of course, *post hoc non ergo propter hoc* (after this, not therefore because of this). This is the difference between, on the one hand, controlled clinical trials, where cause and effect is elicited through randomization and control of possible confounding variables, and, on the other hand, epidemiology, where phenomena are observed in the community. Fortunately, Dr. Fineberg obtained gastrointestinal operations, appendectomies, cholecystectomies, and herniorrhaphies in particular, as natural controls (2). As Table 3 shows, these operations did not decline significantly in frequency after cimetidine was introduced as did ulcer surgery. Some general downward trend was apparently not at work, leaving cimetidine as the most likely cause.

In sum, the role of cimetidine as the cause of the sharper drop in ulcer surgery is supported by four observations:

1. Cimetidine penetrated the market very rapidly, a necessary condition for a detectable effect on health events.
2. The control operations stayed level, indicating no general change in the health care system was at work.
3. The decrease in ulcer surgery in the United Kingdom occurred in 1977 after cimetidine's introduction in late 1976, and the drop in the United States occurred in 1978 after introduction in late 1977. The drop in elective ulcer surgery in Sweden in 1979 showed a similar relationship to cimetidine's introduction in mid-1978. Such convincing repetitions are rare in time-series data.

TABLE 3 Selected Abdominal Surgical Procedures in the United States, 1970–1978

	Rates per 10,000 Population			
Procedure	*1970*	*1975*	*1976*	*1978*
All abdominal surgery	122	138	133	132
Partial gastrectomy and vagotomy	6	5	5	3
Appendectomy	16	15	14	14
Cholecystectomy	18	21	21	20
Herniorrhaphy	25	26	24	24

SOURCE: National Center for Health Statistics, National Hospital Discharge Survey.

4. No other major factors were apparent, such as other new ulcer drugs, to cloud the cause-and-effect picture.

Beyond all this, cimetidine represented a totally new method of acid suppression that dramatically changed outcomes in ulcer disease. An interesting question for technology assessment is: When will there be another opportunity to see so clearly in statistical series the beneficial effect of a new drug? Probably not soon.

What about cost reductions? In Rhode Island we were able to obtain the costs charged to the health insurers Blue Cross and Blue Shield for an ulcer patient's stay in the hospital. Table 4 illustrates that in 1978, charges were $4,874 for a stay with surgery and $2,075 for a stay without surgery. Thus, a savings of $2,799 was realized if the ulcer patient was hospitalized but avoided surgery; $4,874 was saved if the patient avoided hospitalization altogether (4). If these amounts are multiplied by the fewer number of ulcer operations observed nationally—the difference between the expected number and the observed numbers in the time-series graphs—savings in national charges can be projected at $60 million to $97 million per year in 1978 dollars in the United States alone (4).

Surgery costs represent direct, treatment costs. Indirect costs include the cost of days lost from work, Work-loss data by diagnosis are rare, but they are available for West Germany where, as seen in Figure 2, cimetidine's introduction was associated with a sharp reversal in the number of workdays lost by hospitalized ulcer patients (5). The work-loss results from the Swedish clinical trial are thus confirmed by the West German data. In Sweden, data on absenteeism are not available; however,

TABLE 4 Charges per Hospital Stay for an Ulcer Patient, Rhode Island, 1976 and 1978

Type of hospital stay	*Year*	*Routine plus ancillary charges*	*Surgery physicians' fees*	*Total charges*
With partial gastrectomy or vagotomy (weighted average)	1976	$3,252	$674	$3,926
	1978[a]	$4,081	$793	$4,874
Ulcer and ulcer related without ulcer surgery	1976	$1,654	—	$1,654
	1978[a]	$2,075	—	$2,075
Savings of admission avoiding ulcer surgery	1978	$2,006	$793	$2,799

[a]Inflation from 1976 to 1978: Routine plus ancillary charges = +25.5% (R.I. Blue Cross average daily hospitalization costs); surgery physicians' fee = +17.7% (Physicians Services, Consumer Price Index).

data on early retirements in ulcer disease patients are. Figure 3 shows a sharp drop in the number of early retirements due to disability from ulcer in Sweden (6). These data also represent productivity gains for society.

We also supported decision-analysis studies of the costs of cimetidine versus surgery (7) and retrospective cost studies, mainly in Medicaid populations, of patients taking cimetidine versus other agents (8). Altogether the studies suggest that cimetidine has caused reductions in ulcer surgery and ulcer-related disability, which more than offset the costs of the drug.

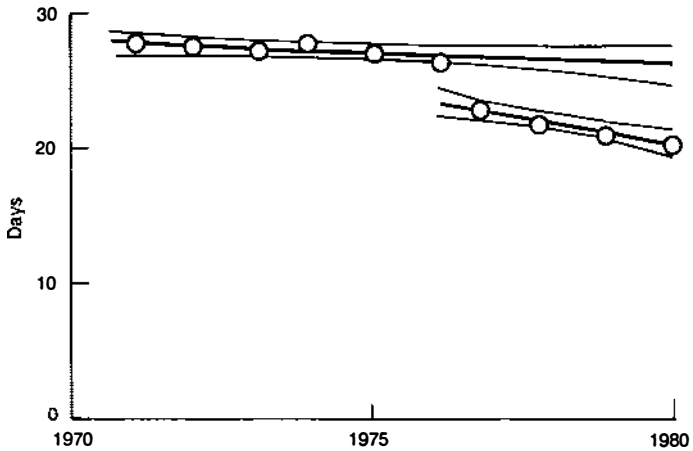

FIGURE 2 Number of workdays lost per hospitalized case of peptic ulcer in West Germany, 1970–1980.

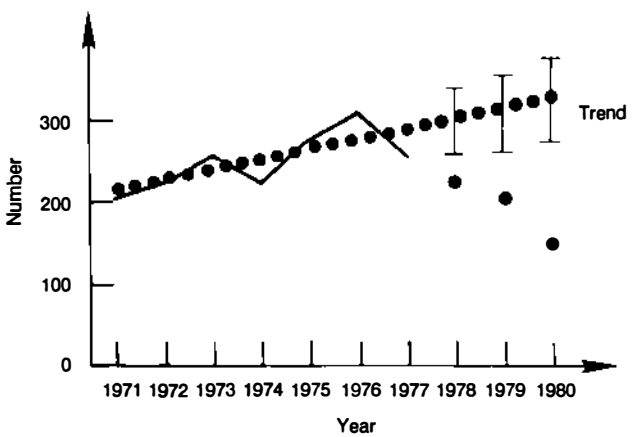

FIGURE 3 Number of newly granted disability pensions for the diagnosis of peptic ulcer in Sweden, 1971–1980.

DIFFICULTY OF MEANINGFUL COST STUDIES

Two points above bear reemphasis. First, had cimetidine not been a dramatically effective new agent widely and rapidly accepted in an otherwise "quiet" environment, it would not have affected national statistics on surgery or work loss in a detectable way. Second, although the clinical trial data suggested that reductions in surgery and work loss were likely in the community, the trial data, based on a placebo comparison, exaggerated the degree of reduction and did not furnish a guideline for the amount of cost-reduction or productivity gains. Thus, problems await those who would like to calculate the cost-reducing effects of a new technology, problems likely to be more difficult if the technology offers a less-than-immediate or a minor advantage.

Consider, for a moment, the cost savings from the nicotine chewing gum that facilitates giving up cigarettes. Stopping smoking reduces heart disease, but the cost savings from that would take a long time to occur and probably could not be isolated in trend statistics. A randomized clinical trial would show the efficacy of the gum in facilitating withdrawal from smoking under controlled conditions of compliance, but it would not predict the results of less motivated use in the real world of the community. Thus, a cost-benefit analysis of the gum would have to be based on estimates of compliance and on an interpolation of the effects of stopping smoking as determined by other long-term studies.

Consider a more difficult example: a nonsteroidal anti-inflammatory drug (NSAID) with the advantage of once-a-day administration. Assume that efficacy and safety are equivalent to other NSAIDs but that price per day is higher. A possible cost saving might conceivably be claimed from the fact that once-a-day convenience improves compliance, thereby improving efficacy, reducing visits to physicians, and possibly lowering time lost from work. I cannot, however, conceive of the clinical trial of one NSAID compared with another NSAID that would demonstrate this with statistical confidence. Published statistical series would show nothing, as any trend changes would be imperceptible.

What is needed is a new kind of trial, one that combines randomized prescribing of approved drugs and hands-off follow-up with recording of medical outcomes and determination of costs from routinely generated, computerized patient records. For some drugs there may be no other way to study real-world effects on cost.

COST-BENEFIT AND COST-EFFECTIVENESS ANALYSIS

An important new drug, like other new technologies, may simply not reduce total treatment costs. New antihypertensives, analgesics, and

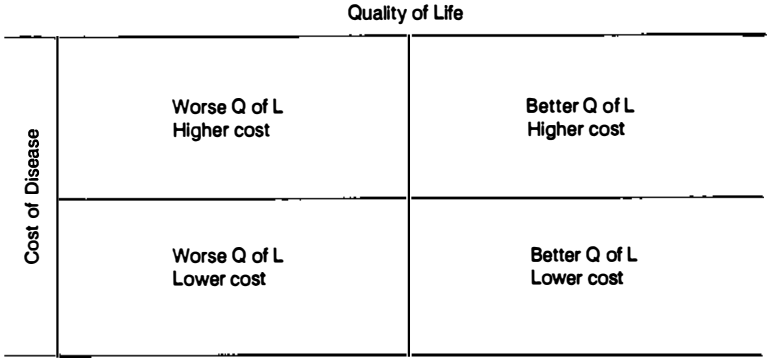

FIGURE 4 Cost-effectiveness outcomes.

antihistamines, among others, may be of this type. What then? We need to turn from cost reduction to cost-effectiveness analysis.

There are four permutations of results that can occur when cost is related to total medical outcome (see Figure 4). Cimetidine would fall into the lower right corner because it lowered cost and produced better outcomes for ulcer patients. In the cimetidine studies, improved medical outcome was ignored based on the conservative assumption that it was no worse than traditional therapy. The lower left-hand box—lower cost but worse outcomes—may represent what actually occurs when cost-containment policies such as shortening patients' stays or discouraging office visits are implemented. The upper left result—worse outcome at higher cost—is obviously undesirable. The upper right box—better outcome at higher cost—is where many new technologies may fall, possibly those most in need of technology assessment.

Table 5 illustrates how cost-benefit analysis might work in this situation.

In ulcer disease we said, in effect, that the value of an agent equals benefits minus costs. If, before cimetidine, benefits of treatment were val-

TABLE 5 Measuring the Benefits of a Technology

Changes in costs or benefits	*benefits*	-	*costs*	=	*value*
Original case	10	-	5	=	5
Cost drops	10	-	3	=	7
Original case	10	-	5	=	5
Cost increases	10	-	7	=	3
Benefits and cost increase	13	-	7	=	6

ued at $10 million and costs of treatment were $5 million, the resulting value of the treatment was $5 million. With cimetidine, let us pretend the benefits stay at $10 million but reduce the costs to $3 million. The result is a net gain in value of $7 million. The gain comes from reducing net costs. Shown below that is what happens when we assume that benefits stay the same but costs increase. If benefits are still valued at $10 but costs increase to $7, value drops to $3. Thus, we cannot stay with the assumption that benefits need not be measured. We would like to be able to find that, if costs increase by $2 (from $5 to $7), benefits increase, say, by $3, to $13, giving a value of $6. Remember, we are on the upper right box of Figure 4, where costs increase but outcome is better. The problem, of course, is that we cannot value the overall benefits in dollars. Thus, we cannot subtract costs from benefits.

The solution is to shift to a different expression of value, cost-effectiveness. In cost-effectiveness, benefits are not measured in money terms. We express them as an added unit of some desirable outcome—years of life, pain-free days, daily function, trips out of the home, workdays. The relationship becomes a ratio—for example, $50 in added cost per pain-free day gained, or $50,000 per year of life. The question immediately arises whether the ratio is high or low, favorable or unfavorable. In cost-benefit analysis, all we had to do, in theory, was to subtract costs from benefits and hope to find a positive result. In cost-effectiveness analysis, we must either judge the results ratio in isolation or compare it with the ratio for another treatment.

A STANDARD UNIT OF EFFECTIVENESS: QUALITY-ADJUSTED LIFE-YEARS

The quality-adjusted life-year, or QALY, is a concept allowing comparison of the effects of different treatments in the same, standardized unit of benefit, a year of healthy life. It takes a bit of thought to understand a QALY, because usually any one person receives only part of a QALY. For example, if a person has arthritis and improves during the next month from 70 percent of perfect health to 90 percent of perfect health, that represents a gain of 20 percent of perfect health. Assuming that gain is experienced for one year, 20 percent of a QALY has been generated. If four others also receive the same benefit, together the five people have received one whole QALY. If our one-QALY gain was associated with $20,000 in added costs, the cost-effectiveness ratio is $20,000 per QALY. This compares favorably with the $85,000 or more per QALY produced by certain uses of prophylactic leukocyte transfusion (9); it compares unfavorably with $4,000 or less per QALY produced by screen-

ing for phenylketonuria (10). Thanks to the QALY, we have a better perspective for judging the cost-effectiveness ratio of the technology in question.

The Case of Auranofin

The foregoing served as a goal as Smith, Kline and French Laboratories undertook the study of the cost-effectiveness of auranofin, its new oral gold agent for rheumatoid arthritis. Auranofin would normally be added to existing regimens of NSAIDs; its labeling required monthly tests for certain adverse effects that can occur with gold therapy. Thus, it was expected to increase drug and related office visit costs and, very possibly, total treatment costs, at least during the measurable near term. The drug, like other so-called disease modifying agents, had a slow onset of action of approximately two to three months before efficacy was observable. We intended to measure changes in quality of life and, if possible, total treatment cost produced by auranofin and to express the result as a cost-effectiveness ratio.

To do all this scientifically required a randomized trial. Our hypothesis was that a gain in quality of life would result from adding auranofin to NSAIDs. The control therapy was existing NSAIDs plus placebo, and the active therapy was existing NSAIDs plus auranofin. In this case the control therapy was close to what would be used in the real world. Preparing for an intent-to-treat analysis, we wanted to keep all patients in the trial randomized, i.e., blinded as to their assigned medication and follow-up. To keep deteriorating or adversely affected patients on a blinded agent for more than six months was considered unethical, so the study was limited to six months, a rather short period for a chronic, variable disease such as rheumatoid arthritis. It is important to note that it takes two to three months for auranofin to work. This could be particularly important in evaluating costs, as the effect of auranofin on hospitalization and disability might take one year or more to be seen.

We selected a wide variety of quality-of-life measures. They included Activities of Daily Living (ADL) questionnaires specific to arthritis, such as the Health Assessment Questionnaire (HAQ) by Fries et al. (11). We also used broader measures of function and well-being, such as the Quality of Well-Being Questionnaire (QWB) (12,13). This questionnaire asks questions such as, "On which of the past six days did you travel or use public transportation?. . . On which were you confined to bed or wheelchair?. . . What was the reason you were in bed?" The answers classify a patient into various categories of mobility, physical function, and social function as well as into symptom/problem categories, each of

which has a weighted score. The scores are totaled and subtracted from 1.000. A total of 1.000 thus represents perfect health. Zero would be death.

Other questionnaires included simple global questions on how the patient would rate his or her overall condition. We even used measures at the cutting edge of quality-of-life measurement: time trade-off and standard gamble questions. In these the patients' overall condition is calculated from odds or risks of harm he or she is willing to undergo to have a chance of cure. Finally, we also used the traditional measures: counts of swollen and tender joints, time to walk 50 feet, duration of morning stiffness, and grip strength.

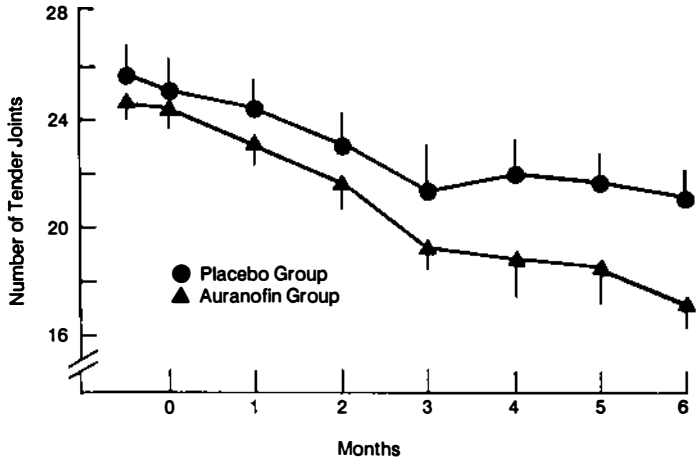

FIGURE 5 Results of auranofin clinical trials: number of tender joints.

We used many nontraditional measures because each approached different aspects of quality of life and because we did not know which, if any, would show a statistically or clinically significant treatment effect for auranofin. There were no useful statistics on the variability of most of the measures in a clinical trial setting. No previous trial existed for guidance.

The questionnaires were administered by trained interviewers living at or near the 14 investigational centers. They did not have medical backgrounds and were not employed by the centers, but they were trained and supervised by an excellent outside organization, Rhode Island Health Services Research (SEARCH). The outcome assessors received home study materials and nearly a week of centralized training in interviewing and

questionnaire administration. They interviewed the patient regularly during the six-month study, with the interviews averaging approximately one hour. All early interviews and many later ones were tape recorded for monitoring by SEARCH. The questionnaires were sent to SEARCH each week and the data computerized there. The assessors were thus constantly monitored for standardized execution of the questionnaires. In sum, the data quality was excellent, better, in fact, than that from some of the traditional measures. There was nothing "soft" about the data obtained, a word often heard when quality-of-life measures were first considered. Finally, local cost data for treatment and indirect-cost events were obtained during the six months.

Selected results are displayed in the following figures (14). Figure 5 shows reduction in the number of tender joints (vertical axis) over time, a traditional measure of benefit, desirable but not clearly indicative of whether the patient is doing better. At six months the auranofin group had fewer tender joints ($p = 0.01$).

Figure 6 shows the reduction in pain, as measured by the McGill Pain Questionnaire. The auranofin group had less pain ($p = 0.02$), scaled on the vertical axis. More indicative of total outcome are the HAQ results, shown in Figure 7. It measures arthritis-specific disability, how well the patient can perform ADLs such as dressing, walking, and grooming, expressed on a 0-to-3 scale. A low score means less disability. The auranofin group attained better scores ($p = 0.01$). Figure 8 shows the QWB results, expressed on a scale from 0 to 1.000. A score on this general health measure represents patient performance in the areas of mobility

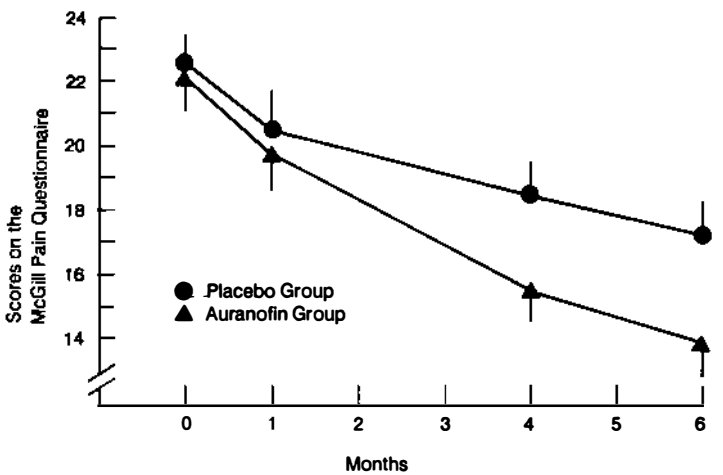

FIGURE 6 Results of auranofin clinical trials: McGill Pain Questionnaire.

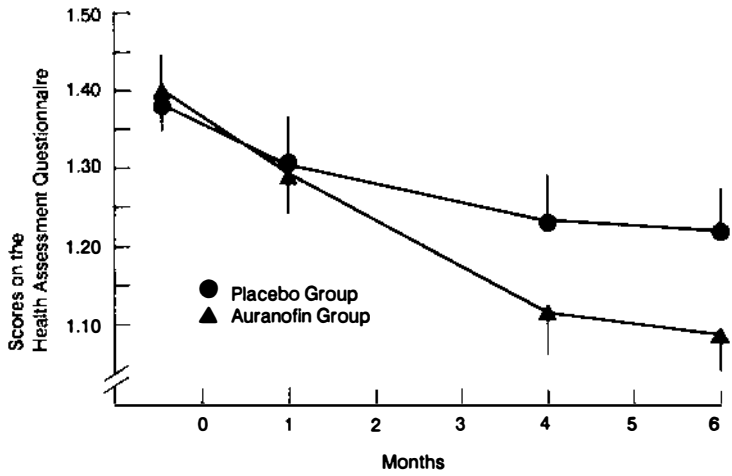

FIGURE 7 Results of auranofin clinical trials: Health Assessment Questionnaire.

and physical and social function adjusted for the negative effect of symptoms or problems, including adverse effects of therapy. The higher the score, the better. The score is also preference weighted, reflecting the importance to the patient of the various levels of performance or symptom/problem. The auranofin group had a higher score at six months, closer to 1.000, which is perfect health ($p = 0.005$).

Finally, to avoid statistical fishing expeditions, the approximately 20 measures were grouped into four composite dimensions. Treatment effect was assessed for each dimension and expressed in terms of percent change from baseline (Table 6). Each shows significance at the 0.0125 level or better, except pain, which decreased as did the other three. Analysis was of the intent-to-treat type. Thus, by rigorous statistical standards,

TABLE 6 Health Status Measures in Two Clinical Trials of Auranofin

	Change in composite scores (% of baseline standard deviation)		
Health status dimension	*Auranofin*	*Placebo*	*p value*
Physical/clinical	34.5	15.7	0.003
Functional	29.2	4.9	0.001
Pain	74.1	49.7	0.026
Global	50.0	26.8	0.006

TABLE 7 Hypothetical Health Status Results of Auranofin Clinical Trials

	Hypothetical results		
Health status measure	*Baseline*	*Month 6*	*Change*
Health Assessment Questionnaire[a]			
Auranofin	0.375	0.625	0.250
Placebo	0.375	0.375	0
Treatment effect			0.250
Quality of Well-Being Questionnaire			
Auranofin	0.500	0.600	0.100
Placebo	0.500	0.500	0
Treatment effect			0.100

[a]Inverted scale.

auranofin improved quality of life as measured by a wide range of outcome measures.

Interpretation of Quality-of-Life Results

The composite scores are, of course, abstract, representing summaries of the total scores of each of the questionnaires. The score of each questionnaire is already abstract, representing the individual questions it contains. For example, within the functional composite is the score from the QWB questionnaire. As we noted, this questionnaire can produce scores from 0 to 1.000, with 1.000 representing perfect health.

Let us pretend that the results of the HAQ and the QWB were as shown in Table 7. (These are hypothetical results only, arranged for arithmetical simplicity to illustrate cost-effectiveness analysis.) We would have a net treatment effect—that is, a benefit—from auranofin of 0.250 on the HAQ score and 0.100 on the QWB. Since few physicians and cost-conscious administrators know what these questionnaires are or have much feel for what their scores mean, we are faced with a new task: to represent the real, concrete improvements in the auranofin patients' daily lives that are reflected in the change in the abstract scores. What we may need is to present the individual items within the questionnaires and the results of each one so that the nonexpert can appreciate how the drug benefits patients.

EXPRESSING COST-EFFECTIVENESS RESULTS

The detailed cost findings from the auranofin trial have been published (15). In order not to duplicate that, let us use the hypothetical treat-

Table 8 Hypothetical Dollar/Score Ratios at Six Months from Clinical Trials of Auranofin

Health status measure	*Change in composite scores (% of baseline standard deviation)*		
	Placebo	*Auranofin*	*Added cost/ score improvement*
Health Assessment Questionnaire[a]	$1,000/0.375	$1,500/0.625	$500/0.250
Quality of Well-Being Questionnaire	$1,000/0.500	$1,500/0.600	$500/0.100

[a] Inverted scale.

ment effect and hypothetical cost results to illustrate points of interpretation. Let us pretend that during the six months, auranofin patients would spend, on average, $500 more in treatment costs than would the placebo patients. This is the cost number for the numerator of the cost-effectiveness ratio. Using the efficacy numbers from Table 7, we can express cost-effectiveness as shown in Table 8.

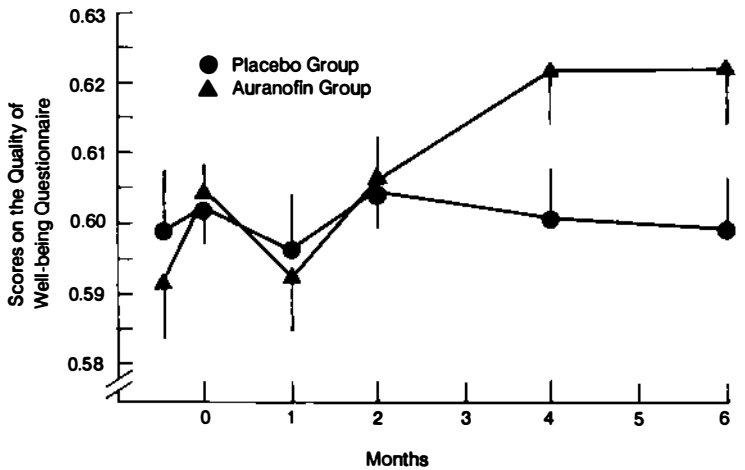

FIGURE 8 Results of auranofin clinical trials: Quality of Well-Being Questionnaire.

TABLE 9 Hypothetical Dollar/Quality-Adjusted Life-Year (QALY) Ratios at 12 Months from Clinical Trials of Auranofin

	$/QALY Ratios Hypothetical 12-month results			
Comparison therapies	*Placebo*	*Auranofin*	*Added cost/score improvement*	*Added cost/added QALY*
Auranofin	$1,000/ 0.5 QALY	$2,000/ 0.6 QALY	$1,000/ 0.1 QALY	$10,000/ QALY
PKU screening				$4,000/ QALY
Heart surgery				$100,000/ QALY

Again, the health administrator would not only need to know what the gain in HAQ and QWB scores represents but also need some basis for judging whether $500 is a reasonable sum to pay for such gains. The QALY calculation may help somewhat here. Thanks to the way the QWB is designed, the auranofin treatment effect of 0.100 points or 10 percent of perfect health can be used in the cost-effectiveness ratio.

This step is shown in Table 9. We are assuming here that, over 12 months, auranofin treatment would generate $2,000 in total costs. We see that auranofin would produce 10 percent of a QALY for $1,000, equaling $10,000 for 1.000 whole QALY.

INTERPRETING COST-EFFECTIVENESS RESULTS

Should a health care program pay for auranofin? So far no cost/QALY ratio of alternative antiarthritis agents has been determined. If the ratio for screening for phenylketonuria is $4,000/QALY and for certain heart operations is $100,000/QALY, the health care program is presumably still obligated to deal with rheumatoid arthritis. Our hypothetical $10,000/QALY for a drug like auranofin gives some perspective but does

not itself lead to a decision about whether to reimburse the treatment costs.

It would seem more meaningful for a decision maker to learn simply that for $500 the average patient improves the equivalent of whatever it takes to increase on the QWB scale from 0.50 to 0.60, which is the score increase that we hypothesized earlier (in Table 7). This score change could be accomplished by a patient improving on the social activity dimension of the QWB scale (see Table 10). For instance, suppose a patient improved from "had help with self-care activities" (SAC 0.106) to full social activity (SAC 0.000). When the QWB score is fully calculated as shown in Table 11, it shows approximately a .100 gain in score. Thus, for $1,000 in extra cost, the average patient improves the equivalent of from needing basic care to full social activity, or some other equivalent improvement in function. This concrete example may help answer the cost-effectiveness question: Do you get good value for money?

TABLE 10 Performance Levels and Weights Used in Scoring the Quality of Well-Being Scale[a]

Level	*Weight*
Mobility (MOB)	
Drove car and used bus or train without help	0.000
Did not drive or had help to use bus or train	0.062
In house	0.062
In hospital	0.090
Physical activity (PAC)	
Walked without physical problems	0.000
Walked with physical limitations	0.060
Moved own wheelchair without help	0.060
In bed or chair	0.077
Social activity (SAC)	
Did work, school, or housework and otheractivities	0.000
Did work, school, or housework but other activities limited	0.061
Limited in amount or kind of housework, school, or work	0.061
Performed self-care, but not work, school, or housework	0.061
Had help with self-care activities	0.106

[a]The weight for the symptom/problem dimension is 0.170.

TABLE 11 Hypothetical Quality of Well-Being Results for an Individual Patient from Clinical Trials of Auranofin

Formula	1 - (MOB + PAC + SAC + symptom/problem)
Baseline score	1 - (0.062 + 0.000 + 0.106 + 0.170) = 1 - (0.338)
Total baseline score	0.662
Treatment score	1 - (0.062 + 0.000 + 0.000 + 0.170) = 1 - (0.232)
Total treatment score	0.768
Difference between baseline and treatment scores	0.768 - 0.662 = 0.106

LESSONS FROM EXPERIENCE

In conclusion, some points about cost-benefit and cost-effectiveness studies of new technologies may be usefully summarized:

- It is optimistic to count on a simple cost-reduction outcome, that is, on the technology reducing total treatment costs.
- Even if it seems logical that the technology should reduce costs, it may be difficult or impossible to prove it with time-series data.
- A systematic estimation study based on any data that are available may be needed.
- A clinical trial may provide evidence of cost reduction if the medical features of the protocol do not involve too many artificialities or alter real-world cost events.
- A clinical trial may provide evidence of quality-of-life improvement, which often means more than the traditional efficacy data required by regulatory authorities.
- Relating increased costs to increased quality of life in a cost-effectiveness analysis will give a dollar per unit-of-outcome ratio, which must be judged as desirable or not.
- If the unit of outcome is a score on a questionnaire, the meaning of the score must be communicated.
- If the unit of outcome is a QALY, the definition of QALY must also be communicated—no easy task.
- A dollar/QALY ratio will probably not be available for alternative treatments for the same condition.
- Even the dollar/QALY ratio requires a judgmental decision.
- More economic perspective is achieved with than without the cost-effectiveness analysis.

- Whatever the cost-effectiveness ratio, the quality-of-life results themselves can provide much evidence of how the technology helps patients.

The cost-reduction studies of cimetidine were definitely worthwhile. They have made a vital difference in reimbursement and related decisions in the United States and elsewhere. In this sense, they were commercially highly cost-beneficial. It is too soon to know whether the cost-effectiveness results for auranofin will inform similar decisions. The quality-of-life data alone, however, will be useful in communicating to providers the effect of auranofin on practical performance in daily life.

There appears to be a positive future for technology assessments like those above, if certain problems can be solved. We need to overcome the problem of the artificiality of data from carefully controlled clinical trials. We need routinely generated and computerized data on patients' health care and work loss over time. Finally, we need to understand what cost-effectiveness ratios can mean for practical decision making. These are not insignificant hurdles, but they can probably be overcome.

REFERENCES

1. Bodemar, G., and Walan, A. Maintenance treatment of recurrent peptic ulcer by cimetidine. Lancet 1:403–406, 1978.
2. Fineberg, H.V., and Pearlman, L.A. Surgical treatment of peptic ulcer in the United States. Lancet 1:1305–1307, 1981.
3. Fineberg, H.V. Discussion of paper by Jonsson. In Culyer, A.J., and Horisberger, B., eds. Economic and Medical Evaluation of Health Care Technologies. New York, Springer, 1982.
4. Norton, J. The Effect of Cimetidine on the Cost of Hospitalization for Peptic Ulcer Disease in Rhode Island (Update Report). Providence, Rhode Island, Health Services Research, Inc., 1982.
5. Horisberger, B. Epidemiological development of peptic ulcer and an evaluation of duodenal ulcers. In Culyer, A.J., and Horisberger, B., eds. Economic and Medical Evaluation of Health Care Technologies. New York, Springer, 1982.
6. Carlsson, P., and Jonsson, B. Macroeconomic evaluation of medical technology: A study of the introduction of cimetidine for the treatment of peptic ulcer. Center for Medical Technology Assessment Report 1986:1, Sweden, 1986.

7. Culyer, A.J., and Maynard, A.K. Cost-effectiveness of duodenal ulcer treatment. Social Science and Medicine 15C:3-11, 1981.
8. Geweke, J., and Weisbrod, B.A. How does technological change affect health care expenditures? The case of a new drug. Evaluation Review 8:75-92, 1984.
9. Rosenshein, M.S., Farwell, V.T., Price, T.H., Larson, E.B., and Dale, D.C. The cost-effectiveness of therapeutic and prophylactic leukocyte transfusion. New England Journal of Medicine 302:1058-1062, 1980.
10. Bush, J.W., Chen, M., and Patrick, D.L. Cost-effectiveness using a health care status index. Analysis of the New York State PKU screening program. In Berg, R., ed. Health Care Status Indexes. Chicago, Hospital Research and Educational Trust, 1973.
11. Fries, J.F., Spitz, P.W., and Young, D.Y. The dimensions of health outcomes: The health assessment questionnaire, disability, and pain scales. Journal of Rheumatology 9:789-793, 1982.
12. Kaplan, R., Bush, J.W., and Berry, C.C. Health status: Types of validity for an index of well-being. Health Services Research 11:478-507, 1976.
13. Kaplan, R.M., Bush, J.W., and Berry, C.C. The reliability, stability, and generalisability of a health status index. In Proceedings of the American Statistical Association, Social Statistics Section, 704-709, 1978.
14. Bombardier, C., Ware, J., Russell, I.J., Larson, M., Chalmers, A., Read, J.L., and the Auranofin Cooperating Group. Auranofin therapy and quality of life in patients with rheumatoid arthritis: Results of a multicenter trial. American Journal of Medicine 81:565-576, 1986.
15. Thompson, M.S., Read, J.L., Hutchings, H.C., Paterson, M., and Harris, E.D. The cost-effectiveness of auranofin: Results of a randomized clinical trial. Journal of Rheumatology 15:1, 35-42, 1988.

The View of a Government Medical Care Provider, Quality Assurer, and Technology Assessor

John A. Gronvall

THE VETERANS ADMINISTRATION HEALTH CARE SYSTEM

The Veterans Administration (VA) is all the things implied by the title of this chapter: a provider of care, an assurer of quality, and an assessor of technology. Each of these roles is carried out on a very large scale by the VA, within an organized system operated through a single management structure. In the context of this volume, focused jointly on quality of care and technology assessment, I believe my most important function is, figuratively, "to offer" the VA to you, the leaders of the scientific and medical community involved in quality assurance and technology assessment. The VA is an organization that can serve as an almost ideal testing ground for some of the ideas discussed during the course of the forum.

The VA is a remarkable and, to me, a remarkably interesting system. It is large. Annual inpatient episodes total 1.3 million, and outpatient visits exceed 18 million. Including the private care we pay for (as an insurer, paying for private sector care for certain eligible veterans), the VA is responsible for 1.4 million hospital patients and 20 million outpatients every year.

The VA's nationwide medical system currently provides integrated care through a network of 172 hospitals, 117 nursing homes, 228 outpatient clinics, and 16 domiciliaries. Our patient population is 3.5 million individuals each year, mostly male, all adult, with a relatively high proportion of elderly compared with the population as a whole. A largely uniform and centralized system of patient records is in place.

I have been a part of this extensive and, in some ways, little known system on a full-time basis for the last four years. As a physician in the

academic world, however, I have been related to it all my professional life.

As the VA's chief medical director (since January 1987), I am often asked what I want to accomplish. My short answer summarizes the ultimate goal of all my decisions and actions: "VA health care second to none." This is important to me, and applicable here, because it establishes as a standard the best quality of care that exists throughout the health care system, not quality standards that exist only within the VA.

As a corollary to that, I want the VA to be more active, more open, more widely recognized as a key participant in the health care system, advancing the concepts discussed in this volume.

As big as we are, the VA is in many ways the best kept secret in Washington, a situation that admittedly has its advantages. However, we do not want to be a secret to the health care community at large or to the thinkers and planners who are shaping the future of United States health care. I am convinced we belong in the thick of deliberations such as this, which is why I said I am here to offer the VA to you.

In thinking about this chapter, I identified the unique credentials that qualify me to address this topic. First, I know little about the topics under discussion; no part of my background prepares me to address them in any depth. I am, however, daily held highly accountable for any and all issues surrounding the quality of care provided by the VA and for the investment in the technology that supports medical care. My full comprehension of unrelenting public accountability is relatively recent. Washington is certainly one of the most intensely scrutinized environments. It has been a remarkable experience to feel the extent to which scrutiny of outcome and scrutiny of process, by the public and the public's representatives, affect our medical care programs.

I am expected to speak knowledgeably on quality-of-care matters on a moment's notice—in public and on the record. I now know well what Thomas Jefferson meant when he asserted that government has no inherent or independent power. As he envisioned it, and as we practice it today, a proper government must always remain accountable to the people who established it.

The public, clearly, has an undeniable right to know what it is getting for its money. Citizens have every right to be assured that the officials entrusted with carrying out health and medical programs are aware of the public's concerns and interests, regardless of whether these programs are operated by "public" or "private" organizations.

The size of the VA system adds an extra dimension to the job of managing it. The VA's Department of Medicine and Surgery is spread across the country and employs some 200,000 individuals. Besides providing direct patient care, we have extensive statutory responsibilities in the fields of biomedical research and health care education.

We operate an affiliation network that traces its beginnings to 1946 when the chief medical director issued the pivotal Policy Memorandum no. 2, which provided the underpinnings for the current system of affiliation of VA hospitals with most of the medical schools in the United States. The affiliation program is the foundation on which the VA developed the kind of medicine, research, and education programs that are national assets. Each year, more than 100,000 medical students, residents, nursing students, and dental students get part of their education in the VA system. Well over half of all doctors practicing in the United States had a significant part of their training in VA facilities.

In terms of research the VA spends $160 million annually to support some 2,500 medical research projects. Health services research and rehabilitation research account for another $22 million in VA funds. If we include the salaries of physician investigators, the total figure of direct VA support is well over $250 million. Our physician staff also brings in approximately $100 million annually in grant support from outside sources.

TECHNOLOGY ASSESSMENT IN THE VA

The VA, as a large integrated system, has in the past played a major role in many important studies and clinical trials. Tuberculosis, hypertension, coronary artery disease, rehabilitation, and prosthetics development are some of the areas in which the VA has made invaluable contributions to developing technology and defining appropriate care. A well-known example from recent years is the work of a VA researcher, Dr. William Oldendorf of the West Los Angeles Medical Center. His efforts formed the scientific basis for development of computerized tomography (CT).

The VA continues to be involved in assessing new technologies for their future potential, and it has an active task force on technology assessment working to define the broad scope of our evaluative efforts.[1] We are undertaking studies to compare magnetic resonance imaging with positron emission tomography as diagnostic aids. In Albuquerque we are collaborating with the Los Alamos National Laboratory in installing a very large experimental facility for magnetoencephalography.

Among the truly exciting projects under way are direct cortical stimulation for visual improvement and functional electrical stimulation (FES). FES uses computers to orchestrate delivery of electrical impulses to

[1]Editors' Note: Dr. Eleanor Travers, Chairman of the VA Task Force on Technology Assessment, is a member of the Federal Liaison Panel and Methods Panel of the IOM's Council on Health Care Technology.

the leg muscles of patients suffering from spinal cord injuries. Some of these patients are able, in fact, to stand and walk. It is a wonderful experience to see such an application of computer technology to a medical problem.

The VA clearly has an ongoing role in building the bridge between technological innovation and medical care. That role is an extension of past success.

QUALITY ASSURANCE IN THE VA

Quality assurance, on the other hand, is a more recent arrival on the scene—at least as a centrally organized and directed program. We are well along in implementing a systemwide peer review program, organized at the district level. (The VA has 28 medical districts, each including five or six hospitals.) This program, called Medical District Initiated Peer Review Organization (MEDIPRO), differs from private sector peer review organization (PRO) activities in that it focuses entirely on quality-of-care issues.

Another major VA quality assurance review resulted from a legislative mandate to compare VA surgical mortality experience with "prevailing national mortality and morbidity standards" (P.L. 99-166). While this legislation was being deliberated, we at the VA testified against its passage, saying that recognized national standards do not exist. The Congress decided to overlook that, and we have a mandate to compare VA surgical mortality experience with those yet-to-be-developed national standards.

The legislative process made clear to me that one can view this issue in more than one way. When I talked to Capitol Hill staff and members of Congress, I made the case that we could not do this because there are no national norms. Their response was just as simple: "Well, Doctor, the law we are writing will just require you to compare your surgical procedures with mortality outside the VA." When I would try to describe the limitations of mortality rates as measures of quality of care, I was again met with: "Now wait a minute. If someone has an operation and survives, it sounds good. When someone has an operation and dies, that's not nearly as good. Are you trying to tell me that there is no difference there that you can usefully record and report on?"

The rest of the conversation is immaterial. The law passed! We have submitted our first report. How did we do it? We did what you might expect: We convened panels of experts, mostly from outside the VA. These physicians and surgeons attempted to define acceptable mortality experience, based on their own experience and on the literature, for vari-

ous categories of surgical procedures. Then we compared VA data with those defined by the experts. I am pleased that, on a systemwide basis, we fell well within the acceptable limits. We are now in the process of reviewing individual facilities that exceeded the limits.

Parenthetically, we were surprised to find that the VA surgical procedure with the highest mortality (defined as death occurring within 30 days) was tracheotomy! It is not hard to figure out why, since tracheotomies are done on many patients who are terminal. This points out once again, however, that simple data can require elaboration before they become useful information.

Our efforts to review the VA's entire cardiac surgery program have been well publicized over the past couple of years. As a result of public concern about quality of care, we reviewed all 51 medical centers that were doing open heart surgery. This assessment led to closure of the open heart surgery programs at four of our medical centers. We were able to do this because the findings of the review were thoroughly grounded in quality-of-care considerations, which allowed us to fend off the more parochial political considerations that so often interfere with major decisions.

RELATIONSHIP OF QUALITY OF CARE AND TECHNOLOGY ASSESSMENT

Technology assessment has certain defining characteristics (summarized in Figure 1), all of which focus on the individual components of either a diagnostic or a treatment program. A patient moves sequentially through those episodes and interactions, and technology assessment can address the question of efficacy separately for each of those discrete steps. In doing so, we define the ideal or desirable outcome for each step and set the standards for appropriate or acceptable use.

FIGURE 1 Characteristics of technology assessment.

FIGURE 2 Characteristics of quality assurance.

In contrast, quality assurance (summarized in Figure 2) focuses on the whole patient and on the cumulative outcome of the patient's encounter with a doctor or hospital over time. The bottom line is whether the patient had a good result. If the patient dies, it really does not matter that all the individual pieces of the diagnostic and treatment sequences measured up except for one. That one piece or step failed and produced an undesirable overall outcome, and it is this aggregate outcome that is the concern of quality assurance.

Properly done, of course, quality assurance reviews can identify the failure point and provide the information necessary to correct it for the next patient. So far, quality assurance programs are, unfortunately, based largely on comparisons—among hospitals, among practitioners. Absolute standards remain elusive.

TECHNOLOGY ASSESSMENT AND QUALITY ASSURANCE

Is there a model in which technology assessment and quality assurance overlap? I would suggest that while technology assessment creates the objective basis for assessment of each diagnostic or therapeutic step in a patient's care, quality assurance mechanisms will determine what actual effect those individual steps have on accomplishing the ultimate outcome of all medical endeavors—improved health of the patient.

Excellence in health care is not yet fully describable through a set of objective criteria, but I believe we can get closer to it than we have thus far. I believe it is this area, where quality assurance and technology assessment meet (Figure 3), that offers the most fertile ground for future exploration.

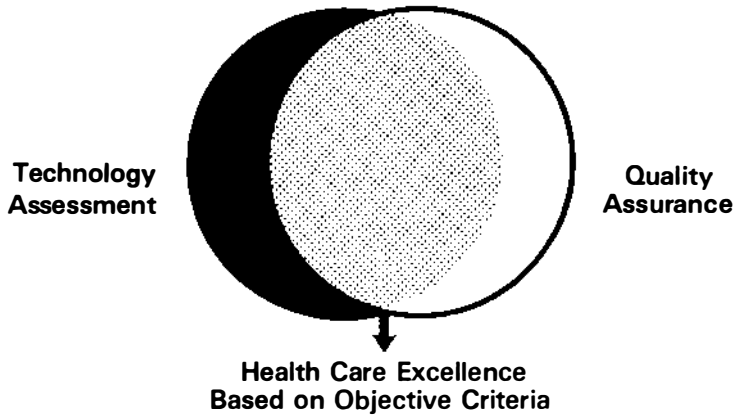

FIGURE 3 Relationship between technology assessment and quality assurance in the Veterans Administration.

We in the VA will continue to understand this relationship in our fishbowl. If our efforts to relate these two activities pay dividends, it will be good for our veteran patients. In addition, as is so often the case, it will be good for the U.S. public. In that context I again offer the VA health care system for your consideration.

THE ORIGIN OF THE EFFORT TO RELATE TECHNOLOGY ASSESSMENT AND QUALITY OF CARE

I want to close by sharing an item recently brought to my attention, the applicability of which seems to me uncanny, given the subject of this volume. It is often startling to find that someone else has already thoroughly thought through an idea we are still piecing together. It is at least encouraging to know that this deliberation is far from new or unique.

The following excerpts are from a remarkable paper that is nearly 130 years old. I invite you to read this elegant description of the problem of determining whether what we are doing in our hospitals is really any good.

> No one, I think, who brings ordinary powers of observation to bear on the sick and maimed can fail to observe a remarkable difference . . . in their duration and in their termination in different hospitals.
>
> It is sometimes asserted that there is no such striking difference in the mortality of different hospitals.

There is, undoubtedly, some difficulty in arriving at correct statistical comparisons to exhibit this.

For, in the first place, different hospitals receive very different proportions of the same class of diseases. The ages in one hospital may differ considerably from the ages in another. And the state of the cases on admission may differ very much in each hospital. These elements, no doubt, affect considerably the results of treatment.

In the next place, accurate hospital statistics are much more rare than is generally imagined.

The elements which really give information . . . are those which show the proportion of sick restored to health, and the average time which has been required for this object.

The proportion of recoveries, the proportion of deaths, and the average time in hospital must all be taken into account . . . as well as the character of the cases and the proportion of different ages.

Hospital mortality statistics give little information on the point.

These are the words of Florence Nightingale, and I recommend to you a paper entitled, "Notes on Hospitals: Being Two Papers Read before the National Association for the Promotion of Social Science at Liverpool, in October 1858" (Birmingham, Alabama, The Classics of Medicine Library, 1982, a facsimile reproduction of the work originally published in 1859).

I find it a remarkable and plain statement of just what kind of information would be useful in trying to figure out whether our hospitals are doing a good job. I have been fascinated to hear people calling John Bunker "the father of us all." Now, I suppose, we also know who our mother is!

The Challenges and Opportunities That Quality Assurance Raises for Technology Assessment

Sheldon Greenfield

I wish to explore the major theme of this volume: the interrelationship between quality assurance and technology assessment. I shall review a few major points and note the challenges and opportunities that each field has for the other.[1]

It is not hard for me to deal with the integration of these two spheres, because I have been interested in both areas for some years. In 1987 our group at the UCLA Center for the Health Sciences published a paper on the value of the lowly complete blood count; that paper became a chapter in the book referred to by Dr. Steinberg (1). We also wrote an article on quality of care for elderly patients that appeared in May 1987 in the *Journal of the American Medical Association* (2). Thus, I have been working simultaneously in both camps and not experiencing much difference between the two fields, although I acknowledge that in their extremes, the two fields have had disparate microcultures that only rarely interact.

THE INFLUENCE OF TECHNOLOGY ASSESSMENT ON QUALITY-OF-CARE ASSESSMENT

I would like to underscore Dr. Brook's original thesis, that these two fields desperately need each other. Indeed, many of the papers during the forum have supported the potential interdependence of the two fields.

My first point is that the field of quality-of-care assessment and assurance can be helped enormously by technology assessment. The challenge for people studying quality of care is to incorporate technology

[1]Editors' Note: Unless otherwise referenced, individuals referred to in this summary were speakers at the forum, and their papers can be found elsewhere in this volume.

assessment into quality of care. Drs. Morris, Berwick, and Roberts clearly pointed out that data needed to set standards of good medical practice are often unavailable and that, therefore, we do not always know what to ask doctors and organizations to do to optimize the quality of medical care. An example offered by Dr. Borgiel concerned patients with cough (see Figures 1 and 2 in "Assessing the Quality of Care in Family Physicians' Practices by the College of Family Physicians of Canada"). Some data are available on this topic. A technology assessment study on the value of chest x-ray for patients with cough presents a decision model for when to do a chest x-ray, in what kinds of patients, and under what circumstances. When the Canadian standards were set, was this topic reviewed? Very rarely are technology assessment experts brought in to help consensus panels set quality-of-care standards. As Dr. Morris pointed out concerning recent events in New York State, national searches for standards need to be done so that professionals in the hospitals can at least defend themselves and their practices on a rational basis.

I will cite another example in which our group attempted to develop criteria for the care of patients with breast cancer, particularly elderly patients (2). One section of the criteria deals with whether a bone scan should be performed in patients with stage 1 breast cancer. Three indications or criteria were set such that, if any are present, a bone scan should be ordered: if the patient has bone pain, if the patient has hypercalcemia, or if the serum alkaline phosphatase is elevated. This could be a very important issue in the patient's life because it could determine whether the patient had a more advanced stage of disease than could be detected on physical examination. To know whether this is a reasonable standard to hold physicians to, we would search the literature to see whether a bone scan should be ordered on everyone who comes in with early stage breast cancer or on just a relatively small number of designated individuals falling into certain high-risk categories (3,4). Performing scans in the latter group would reduce the number of bone scans to a relatively small number and, at the same time, preserve the quality of their care—everyone or almost everyone who needed to have a bone scan would have gotten one.

The quality assurance world is full of examples where better ideas about technology and technology assessment, either using already published studies or even using the approaches of technology assessment, would help in establishing criteria and standards that could be used in successful quality-of-care studies. Some people, including Heather Palmer, have done this carefully.[2] She and her colleagues have been very

[2]Editors' Note: The reference is to R. Heather Palmer, M.B., B.Ch., of the Harvard School of Public Health. Dr. Palmer, a forum participant, is known for her work in conducting randomized controlled trials of quality assurance approaches.

careful to solicit the kinds of opinions that would solidify their criteria. But in many quality-of-care standards, the criteria are specified by one or two individuals sitting down and simply saying it should be this way.

Another dimension that technology assessment can bring to quality is a way of thinking expressed by many, beginning with Dr. Farber. This approach to thinking and decision making relies on cost-effectiveness analyses in carefully describing the patient population, carefully defining the test or technology under specific circumstances, and introducing the notion of probabilistic thinking or, as Dr. Farber called it, directed thinking. In many quality-of-care reviews, lists are given of the things that we should do. What needs to be inserted into quality-of-care assessments are more appropriate and reasonable ways of thinking and making medical decisions that incorporate the principles embedded in technology assessments.

Experts in technology assessment can be criticized for applying systematic methods too well, because they lead to dependence on the sciences of decision analysis and cost-effectiveness, which may differ from actual clinical practice. This kind of rigorous thinking, however, can be very important in quality assessment. For example, the California Blue Shield efforts, the Clinical Efficacy Assessment Project (CEAP) studies, and the results of technology reviews that have been published would be very helpful in setting up reasonable standards for quality of care. It would help in situations as in New York State[3] where single incidents are investigated. I am not saying that a single, individual incident of poor care or death should not be investigated. The point is that physicians can be taught a new way of thinking—that medical decision making really depends on a probabilistic assessment of a situation, treatment, or diagnostic procedure and not entirely on what happens to an individual patient. The latter leads to a thorough investigation of outliers but does not deal with the way medical care should be practiced on the whole.

I would also say, in terms of quality of care being helped by technology, that technology assessment makes it easier to "sell" quality of care to physicians. We saw, more than 10 years ago, that physicians react badly to quality-of-care pressures. In my opinion, one of the reasons for the decline of the quality-of-care movement in the 1970s was that the methods, the tools, and the faith physicians had in what they were measuring were very weak, whereas physicians today seem to be adapting to technology assessment with a bit more equanimity.

Another recent example of technology assessment helping quality of care is one of our own studies for the CEAP done by Katherine Kahn and myself (5). We reviewed whether to do upper gastrointestinal (GI)

[3] Editors' Note: See the paper in this volume by Thomas Q. Morris, M.D.

endoscopy for patients with upper GI stomach pain, called dyspepsia. We eventually recommended that the diagnostic tests should be considerably reduced; this, naturally, upset the gastroenterologic community. We also recommended, however, that when certain high-risk patients needed further investigation, physicians should turn to endoscopy directly—the better diagnostic test—and avoid x-ray. We acknowledged that we valued the technology, and the strategy carried the day because it dealt with physicians' positive disposition toward the technology while at the same time rationalizing it. It would be easier now to perform a quality-of-care study on the management of patients with dyspepsia. The technology assessment community, particularly with respect to their approaches to medical decision making, can greatly improve quality-of-care assessments and put them on more solid ground.

THE INFLUENCE OF QUALITY ASSESSMENT ON TECHNOLOGY ASSESSMENT

A parallel question is how quality assessment can help technology assessment: What are the opportunities? What are the challenges? Some important aspects were brought out during this forum, one of which is that technology assessment has become too narrow or has stayed too narrow in using cost-effectiveness as its paradigm. The value of such studies is that they are very precise and very clear in terms of defining the population and laying out the consequences of a decision. The downside is that they are limited; they often deal with one decision or one technology in an extremely well-defined situation. They often do not deal with some of the aspects that Dr. Brook mentioned: the quality of life, patient preferences, and health status outcomes.

On the other hand, quality assessment is usually very broad. Instead of just attacking the issue of whether hypertensive patients need an intravenous pyelogram, or whether patients with headache need magnetic resonance imaging, quality-of-care assessment covers the total care of patients with headache, diabetes, or coronary artery disease. It potentially deals with all aspects such as complications, therapy, and diagnosis simultaneously. In fact, when quality of care uses outcomes [I will return to this later], it by definition takes a very broad view, and this broad view needs to be applied to technology assessment questions.

Let us return to the example of the patient coming to the doctor with a cough. Some people may think that the criteria for treatment of cough were too minimal, that they needed to be upgraded to include the more salient issues that a person with cough might present to the physician, such as whether to x-ray a chest and when. A quality assessment of

the type that Dr. Borgiel suggests would be an excellent way to study a technology in a natural setting. It would be of value to study those patients who had chest x-rays and those patients who did not and to contact them in their natural setting to determine the value of the technology (in this case, x-ray). This would be a case-control study to see if the proximate and distal outcomes, such as diagnostic yield and hospitalizations, might have been altered because of obtaining or not obtaining an x-ray.

We did a study a few years ago on routine hospital chest x-ray (6). The study should have been done as part of a quality-of-care assessment in the hospital. One of the sad things about quality assessment is that single studies in single hospitals are often used as the basis for general policy recommendations. Our study was done in one hospital, in one part of the country, among one small group of physicians, and it was probably not generalizable to all hospitals in the country. Yet, information on that topic was so scarce that the results were widely disseminated and accepted.

That kind of study, and many more technology assessments, should be repeated in various forms in quality-of-care studies in this country every day, particularly with patient contact and outcomes. Somebody else would be paying the bill for it, which, for example, is happening in New York State where interest in quality assessment activities is high. Embedded in these studies could be studies of technology. They might not be as rigorous as randomized trials, but they would provide extremely important, useful information, especially if replicated many times.

It seems as if one of the directives that has emerged from this forum is to let quality of care, that is, real life experiments, be used to advance technology as well as to improve quality.

One way this could be facilitated, as I mentioned, is to collect patient outcomes routinely in quality-of-care studies. Patients with cough could be phoned and/or sent questionnaires one month later to see how they are doing and what happened to them in the interim. It seems to me that there is enough interest in using outcomes for judging medical quality of care that they can even shed some light on the value of technologies. I am involved in two national studies that are using outcomes to determine the quality of care. One study is an office practice study, and the other is a hospital study in which we are comparing East Coast to West Coast hospitals. One purpose of both of these studies is to see if the outcomes are similar in different situations, that is, different hospitals, different systems of care, and different parts of the country. If, for example, the outcomes for patients with prostatectomies are about the same in two hospitals, we can examine whether the hospitals vary in cost and use of certain technologies. We might also be able to see whether a

particular technology, because of wide variation, should be studied in more detail if the outcomes were about the same; we might also be able to conclude from such a study that certain highly intense resource use does not necessarily lead to better outcomes. It may be possible to use outcomes to make statements about variations in technology use in specific studies such as these.

I emphasize, however, that the outcome studies must be done with extreme care. I agree with those who think that published hospital mortality data are very dangerous without proper adjustment. I also think, and I would disagree with Dr. Berwick on this point, that it is too extreme to say that there is a tyranny of outcomes. Rather, I believe that outcomes have the potential of being measured and adjusted adequately to make meaningful statements about quality of care and technology use.

These adjustments or controlling factors must be dealt with carefully, however. I will review a few of them. They are listed in Table 1 and are taken from a study by Dr. Brook, myself, and others (7).

TABLE 1 Methodologic Issues in Outcome Research

Process must affect outcome	Compliance and patient factors
Outcome measures must be good	Cost/feasibility
Severity	Standards
Comorbidity	Power: adequate number of patients
Optimal time window	

The first two points—process must affect outcome and having good outcome measures—are obvious, and I will not comment on them further. The next two factors are severity of the condition and comorbidity (that is, the remainder of the patients' conditions). This is often termed, generically, case-mix control, and it is especially important in situations where randomization is not possible or desirable.

Another major issue is setting the correct time window for which the outcome is ascertained following receipt of care. In the six-hospital study we are setting the time window outcomes of cholecystectomy at three months following surgery and total hip replacement at 12 months following surgery. The orthopedic surgeons tell us not to estimate the outcome of a patient with total hip replacement at an interval less than one year because it takes one year for patients to recover maximal function. If we had measured outcomes for all conditions after three months, we would have chosen time windows that were either insensitive or

nonspecific to the care delivered. Other issues are patient factors and compliance, the study size, and aggregation of the studies because of small numbers of adverse outcomes.

To summarize these issues without going into greater depth, it might be possible to use carefully selected and well-validated outcomes as useful measures of quality of care. This research agenda is important for the future because I see both processes and outcomes being used selectively, whichever seem more appropriate for a given situation.

In addition, some of these outcome studies will be useful for technology assessment. These could even include patient preference studies. Patients could be interviewed about how they felt about the technology and whether it made a difference to them. Some years ago, Sox et al. did a study of benign chest pain (8). Patients were randomized to receive or not to receive an electrocardiogram. They had different levels of patient satisfaction and, in fact, different levels of self-reported relief from pain when interviewed four weeks later. This study pointed out that patients have an opinion about technology and that it must be taken into account.

Quality-of-care studies, particularly ones using outcomes, can be of value in technology assessment and should be used whenever possible. When the American College of Physicians asked us to review endoscopy in patients with dyspepsia, we realized that we could not complete the review in the narrow sense of a technology assessment; we ended up by proposing a strategy for care of patients with dyspepsia. We included upper-GI series as well as endoscopy in the strategy. The study meant that we could do a review today of the outpatient office practice of management of dyspepsia using this strategy as the basis for setting criteria.

These kinds of quality-of-care assessments could also shed light on which technologies were useful and further the field of technology assessment. So, the second point amply brought out in this conference is that quality-of-care activities, particularly field experiments, can greatly inform technology assessment. Much of this work hinges on a stable funding mechanism for financing technology assessment at early stages. If we wait for the CEAP and similar groups to do technology reviews, even with data collection, too much time will have passed. Studies need to be embedded in day-to-day practice, for example, during quality-of-care assessment activities and ordinary data collection activities, as in Dr. Paterson's work. Technology assessments need to be performed at all levels, not just during reviews based on someone else's data that were designed to answer other questions.

CONCLUSIONS

As a final comment, we could risk losing the battle for quality assessment as we did 10 years ago by jumping in wildly and mandating

quality-of-care activities without a good understanding of approaches. I would like to see a task force in which technology assessors and the quality assurers would devise a plan that includes a set of reasonable standards for outcomes and processes and that incorporates regional or individual style variations based on both new and ongoing studies. Thus, we would avoid trying to judge things that we cannot judge right now, and we would avoid assuming that the methods available today will give us perfect measurements. Both fields would profit, and the health of the American people would benefit.

REFERENCES

1. Shapiro, M.F., and Greenfield, S. The Complete blood count and leukocyte differential count: An approach to their rational application. In H.C. Sox, Jr., ed. Common Diagnostic Tests, Use and Interpretation. Philadelphia, American College of Physicians, 1987.
2. Greenfield, S., Blanco, D.M., Elashoff, R.M., and Ganz, P. Patterns of care related to age of breast cancer patients. Journal of the American Medical Association 257:2766–2770, 1987.
3. McNeil, B.J. Value of bone scanning in neoplastic disease. Seminars Nuclear Medicine 14(4):277–286, 1984.
4. Thomsen, H.S., Lund, J.O., Munck, O., et al. Bone metastases in primary operable breast cancer. The role of serial scintigraphy. European Journal Cancer Oncology 6(1):25–32, 1980.
5. Kahn, K.L., and Greenfield, S. The efficacy of endoscopy in the evaluation of dyspepsia. Annals of Internal Medicine 102:520–528, 1985.
6. Hubbel, F.A., Greenfield, S., Tyler, J., et al. The impact of routine admission chest x-rays on patient management. New England Journal of Medicine 312:209–213, 1985.
7. Avery, A.D., Lelah, T., Solomon, N.E., et al. Quality of Medical Care Assessment Using Outcome Measures: Eight Disease-Specific Applications. Publication R-2021/2-HEW. Santa Monica, California, The RAND Corporation, 1976.
8. Sox, H.C., Jr., Margulies, I., and Sox, H.C. Psychologically mediated effects of diagnostic testing. Annals of Internal Medicine 95:680–684, 1981.

About the Authors

Jeremiah A. Barondess, M.D.

Jeremiah A. Barondess, the Irene F. and I. Roy Psaty Distinguished Professor of Clinical Medicine, has been a member of the faculty of Cornell University Medical College since 1953. Currently he serves as Co-Chairman, Council on Health Care Technology of the Institute of Medicine. Dr. Barondess is a Diplomate of the American Board of Internal Medicine and has served on its Board of Governors and Council on General Internal Medicine. He is also a Fellow of the American College of Physicians and has been chairman of its Board of Governors, a member of the Board of Regents, and president. He is a member of the Institute of Medicine. Dr. Barondess received his medical degree at The Johns Hopkins University School of Medicine in 1949; he served a residency in medicine on the Osler Medical Service of The Johns Hopkins Hospital and then at New York Hospital Cornell University Medical Center.

Donald M. Berwick, M.D.

Donald M. Berwick serves as Vice-President for Quality-of-Care Measurement, Harvard Community Health Plan. He is also an Associate in Pediatrics, Children's Hospital Medical Center, Pediatrician at the HCHP Kenmore Center, Lecturer on Health Services, Harvard School of Public Health, and Associate Professor of Pediatrics, Harvard Medical School. Dr. Berwick is a Phi Beta Kappa graduate of Harvard College, with an A.B., summa cum laude, in social relations. He earned his master's degree from Harvard University in public policy and his M.D., cum laude, from Harvard Medical School.

Alexander E. M. Borgiel, M.D., C.C.F.P., F.C.F.P.

Alexander E. M. Borgiel is an active member of a group of seven family physicians in Mississauga, Ontario, and serves as the Chief of Staff of The Mississauga Hospital (625 beds). Dr. Borgiel is a graduate in medicine of the University of Toronto. He is also Chairman of the Practice Assessment Committee of the College of Family Physicians of Canada, of which he is a Fellow.

Robert H. Brook, M.D., Sc.D.

Robert H. Brook is presently Professor of Medicine and of Public Health at the University of California at Los Angeles Center for the Health Sciences, where he directs the Division of Geriatrics and the Robert Wood Johnson Clinical Scholars Program. In addition, he serves as Deputy Director of the Health Program at The RAND Corporation, where he has been associated since 1974. Dr. Brook received his undergraduate education at the University of Arizona, his M.D. from The Johns Hopkins Medical School, and his Sc.D. from The Johns Hopkins University School of Hygiene and Public Health. He served from 1972 to 1974 with the U.S. Public Health Service at the National Center for Health Services Research. He is a member of the Institute of Medicine.

Saul J. Farber, M.D., M.A.C.P.

Saul J. Farber is The Frederick H. King Professor and Chairman, Department of Medicine, at New York University of Medicine. He also serves as Director of Medicine at both University Hospital and Bellevue Hospital. He is Dean for Academic Affairs for the School of Medicine at New York University, where he is currently serving as Acting Dean and Provost. Dr. Farber has been the Chairman of the American Board of Internal Medicine and President of the American College of Physicians, which awarded him its Distinguished Teacher Award. He has also served as President of the New York Heart Association and the Association of Professors of Medicine. He was Organizing Chairman and Chairman of the Federated Council for Internal Medicine. He is a member of the Institute of Medicine.

Sheldon Greenfield, M.D.

At the time of the forum, Sheldon Greenfield was Professor of Medicine at the Schools of Medicine and Public Health of the University of California at Los Angeles. He was also Co-director of the RAND/UCLA Center for Health Policy Study. He is now Professor of Medicine at New England Medical Center and a director of its Institute for the Advancement of Health and Medical Care. Dr. Greenfield is a member of the Society for Research and Education in Primary Care in Internal Medicine (SREPCIM) technology committee, a member of the Robert Wood Johnson Foundation's Advisory Committee for the Program on Research and Development in Functional Status, and a member of the board of The Henry J. Kaiser Family Foundation Program in Faculty Scholarship in General Internal Medicine. Dr. Greenfield is a graduate of Harvard Col-

lege and the University of Cincinnati School of Medicine. He completed his internal medicine residency at the Beth Israel Hospital in Boston where he served as Chief Resident in Medicine.

John A. Gronvall, M.D.

John A. Gronvall has served as Chief Medical Director of the Veterans Administration since 1987, having been Deputy Chief Medical Director and Deputy Director of the Department of Medicine and Surgery's academic affairs office. From 1971 to 1982, he was Dean of the University of Michigan Medical School, as well as Professor of Pathology. He has also been affiliated with the University of Mississippi School of Medicine as Associate Dean and Associate Director of the Medical Center. He was chairman of the Association of American Medical Colleges in 1978–1979. Dr. Gronvall received his undergraduate and medical degrees from the University of Minnesota. His residency in anatomic pathology and neuropathology was at the University of Minnesota Medical School, and he is certified by the American Board of Pathology (Anatomic Pathology). He is a member of the Institute of Medicine.

William N. Hubbard, Jr., M.D.

In a long and distinguished career William Hubbard served The Upjohn Company in a number of capacities: member of the Board of Directors, Vice-President and General Manager of the Pharmaceutical Division, Executive Vice-President, and President. He retired from the last post in 1984. He has also had numerous university appointments, including John Wyckoff Fellow, Instructor, Assistant Professor, Assistant Dean, and Associate Dean at New York University College of Medicine and Associate Professor and Professor, Internal Medicine, Director of the University of Michigan Medical Center, and Dean, University of Michigan Medical School. Currently, Dr. Hubbard serves as the Chairman of the Council on Health Care Technology of the Institute of Medicine. He is also a member of the Boards of Trustees of the W. K. Kellogg Foundation, Columbia University, and Family Health International and member of the Boards of Directors of Industrial Technology Institute (where he serves as Vice-Chairman), Johnson Controls, Inc., Consumers Power Company, and The Upjohn Company. He is a member of the Institute of Medicine. Past activities include serving as Regent and Chairman of the Board of Regents of the National Library of Medicine, President of the Association of American Medical Colleges, Chairman of the Board of Directors of the Pharmaceutical Manufacturers Association, and as a

member of numerous committees and panels for federal and state agencies and organizations. Dr. Hubbard received his undergraduate degree from Columbia University and attended the University of North Carolina School of Medicine and New York University College of Medicine, where he received his medical degree in 1944. His medical residency was served at Bellevue Hospital, New York University.

Kathleen N. Lohr, Ph.D.

Kathleen N. Lohr has been a Senior Professional Associate at the Institute of Medicine since February 1987. She now serves as the Study Director for a congressionally mandated study to design a strategy for quality review and assurance for Medicare. For the Council on Health Care Technology, she is the director of a project sponsored by The Henry J. Kaiser Family Foundation on advances in health status assessment and the principal staff for the council's Methods Panel. Before joining IOM, Dr. Lohr spent more than 12 years with The RAND Corporation. Previously, she was associated with the Bureau of Health Manpower of the U.S. Department of Health, Education, and Welfare and with The Johns Hopkins University School of Medicine and School of Hygiene and Public Health. She earned her B.A. with distinction in sociology and an M.A. in education from Stanford University and a Ph.D. in public policy analysis from the RAND Graduate Institute.

Thomas Q. Morris, M.D.

Thomas Q. Morris is President of The Presbyterian Hospital in the City of New York and Professor of Clinical Medicine at the College of Physicians and Surgeons of Columbia University. Dr. Morris is a member of the Board of Trustees of The Mary Imogene Bassett Hospital in Cooperstown, New York, and of the American University of Beirut, a Director of the Greater New York Hospital Association, and Chairman of the League of Voluntary Hospitals and Homes. He chaired the Quality of Care Task Force of the Hospital Association of New York State. Dr. Morris earned his M.D. degree from the College of Physicians and Surgeons of Columbia University in 1958, after receiving a B.S. degree magna cum laude from the University of Notre Dame. He completed his medical residency on the Columbia Medical Division of Bellevue Hospital and fellowship training in the Department of Medicine of the College of Physicians and Surgeons and on the Medical Service of The Presbyterian Hospital.

John H. Moxley III, M.D.

John H. Moxley at the time of the forum served as Senior Vice-President of American Medical International, Inc. Currently he is President of Meta Medical, Inc. Before that, he was Assistant Secretary of Defense for Health Affairs; from 1973 to 1980, he was also Associate Professor of Medicine, Vice Chancellor for Health Sciences, and Dean of the School of Medicine at the University of California at San Diego. He was Dean of the University of Maryland School of Medicine in Baltimore from 1969 to 1973. Dr. Moxley received his undergraduate education at Williams College in Williamstown, Massachusetts, and his medical degree from the University of Colorado School of Medicine. He is a Diplomate of the American Board of Internal Medicine, a Fellow of both the American College of Physicians and the American Federation for Clinical Research, and a Member of the AMA Council on Scientific Affairs.

Morton L. Paterson, Ph.D.

Morton Paterson has been associated with SmithKline Beckman Corporation since 1970, first as a marketing research analyst, then as manager of long-range planning and of market evaluation and marketing research. He is presently Director, Cost-Benefit Studies, a position he has held since 1984. Dr. Paterson received his undergraduate education at Wesleyan University in Connecticut and earned his Ph.D. from Yale University in 1962. He served with the United States Army in Berlin from 1956 to 1958.

Kenneth A. Platt, M.D.

Kenneth A. Platt was a practicing physician near Denver, Colorado. He served his community and his profession since the completion of his army career in the Korean War in 1953. His professional appointments included the presidencies of his county and state medical societies, chief delegate from Colorado to the American Medical Association for 15 years, and founder of The Colorado Foundation for Medical Care in 1970. He was the Foundation's President in 1971–1972 and its Medical Director since then. He was a member of the Institute of Medicine.

Richard A. Rettig, Ph.D.

Richard A. Rettig joined the staff of the Institute of Medicine, National Academy of Sciences, in January 1987 as Director of the Council on Health Care Technology. From 1981 through 1986, Dr. Rettig was Professor of Social Science and Chairman, Department of Social Sciences, and Director, Master of Public Administration Program at the Illinois Institute of Technology. From 1975 through 1981, Dr. Rettig was a senior social scientist with The RAND Corporation, with research interests in health policy. During 1971–1975, he was Associate Professor of Public Administration in the College of Administrative Sciences, Adjunct Associate Professor of Political Science, and Adjunct Associate Professor of Preventive Medicine at Ohio State University. Before that, he was Assistant Professor of Public Administration, the Graduate School of Business and Public Administration, Cornell University. Dr. Rettig's government experience includes service with the U.S. Bureau of the Budget and the New Jersey Department of Higher Education. He received his bachelor's degree from the University of Washington in 1958 and the Ph.D. degree in political science from the Massachusetts Institute of Technology in 1967.

James S. Roberts, M.D.

James S. Roberts joined the staff of the Joint Commission on Accreditation of Hospitals (now Healthcare Organizations) as Assistant Vice-President for Accreditation, and became Senior Vice-President for Research and Planning in October 1986. Currently, he is Clinical Assistant Professor of Medicine and of Psychiatry at the Northwestern University School of Medicine. Previously, he was Medical Director of the Genessee Valley Group Health Association and Administrator of the Joseph C. Wilson Health Center in Rochester, New York. He also served as Clinical Assistant Professor of Medicine and of Preventive Medicine and Community Health, University of Rochester School of Medicine and Dentistry. From 1975–1976, Dr. Roberts held administrative posts at the U.S. Department of Health, Education, and Welfare Bureau of Quality Assurance (now the Health Standards and Quality Bureau); between 1971 and 1973, he served in the U.S. Public Health Service, National Center for Health Services Research and Development, as a project officer for the Experimental Medical Care Review Organization (EMCRO) Program. Dr. Roberts received his B.S. degree from Washington State University in Pullman, Washington, in 1965, and was a member of Phi Beta Kappa. He earned his M.D. degree at Washington University School of Medicine in

St. Louis in 1969. Dr. Roberts began his internship and residency at the University of Rochester, Strong Memorial Hospital in Rochester, New York, and completed his medical residency at the University of Washington, Seattle, in 1974–1975.

Ralph W. Schaffarzick, M.D.

Ralph W. Schaffarzick is presently Senior Vice-President and Medical Director, Blue Shield of California, where he has served since 1969. In addition, he has been a member of the faculty of the Stanford University School of Medicine, as (among other appointments) Assistant Clinical Professor of Pharmacology and Therapeutics and Clinical Professor of Medicine and Health Services Research. Dr. Schaffarzick has served on committees of the California Medical Association, projects sponsored by the State of California's Board of Medical Quality Assurance, and on the Technical Advisory Group on Coverage Policy of the Department of Health and Human Services Health Care Financing Administration. He received both his undergraduate and medical degrees from Stanford University.

Earl P. Steinberg, M.D., M.P.P.

Earl P. Steinberg is an Assistant Professor of Medicine at The Johns Hopkins University School of Medicine and holds a joint faculty appointment in the Department of Health Policy and Management at The Johns Hopkins University School of Hygiene and Public Health. He is also Director of The Johns Hopkins Program for Medical Technology and Practice Assessment. In July 1984 Dr. Steinberg received The Henry J. Kaiser Family Foundation Faculty Scholar Award in General Internal Medicine, an award given "to support exceptionally talented young faculty in general internal medicine." Dr. Steinberg received his A.B. degree summa cum laude from Harvard College, his M.D. from Harvard Medical School, and his Master of Public Policy degree from the Kennedy School of Government at Harvard. His residency training in internal medicine was done at the Massachusetts General Hospital.