

**A RAND NOTE**

**Efficacy, Effectiveness, Variations, and Quality:  
Boundary-crossing Research**

**Robert H. Brook, Kathleen N. Lohr**

**RAND**

The research described in this report was supported by a grant from the U.S. Department of Health and Human Services.

This Note contains an offprint of RAND research originally published in a journal or book. The text is reproduced here, with permission of the original publisher.

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution. Publications of RAND do not necessarily reflect the opinions or policies of the sponsors of RAND research.

**A RAND NOTE**

**N-3368-HHS**

**Efficacy, Effectiveness, Variations, and Quality:  
Boundary-crossing Research**

**Robert H. Brook, Kathleen N. Lohr**

**Supported by the  
U.S. Department of Health and Human Services**

**RAND**



## Efficacy, Effectiveness, Variations, and Quality

### Boundary-crossing Research

ROBERT H. BROOK, MD, SCD,\* AND KATHLEEN N. LOHR, PHD†

The title of this article telegraphs our judgment about the future of health services research as it relates to quality-of-care assessment: Returns from research will come more from integrating efficacy, effectiveness, variations in population-based rate of use, and quality of care into an operational model for policy, planning, and evaluation needs than from continuing to treat them as isolated subjects. We have been asked, however, specifically to consider future directions for *quality assessment* research and the uses to which its products should be put, and that is the main focus of this article. Nonetheless, our views of the directions research in this area should take are shaped by the hope that it will serve a

larger policy purpose and reach beyond the health services research audience.

Ten years ago, and certainly 20, conceptual and practical questions pertaining to these topics as individual subjects were important health services research concerns. Not all of them are so central today. Health services research has grown and matured; with it, these fields have ripened as well.

We can, *if we want*, accurately measure the efficacy of procedures or drugs, assay their effectiveness in the everyday world, examine per capita variations in their use, and even evaluate the quality of the care process in which use of these services is embedded. Indeed, in some cases we may know more about the appropriate way to treat or measure the impact of an illness than we do about its pathophysiologic causes: the correct "steps" of detecting and treating high blood pressure, for instance, are better understood than are the consequences of essential hypertension.

Health services research can profitably be focused on issues of efficacy, effectiveness, variations in use, or quality of care. Our argument, however, is that tomorrow's important issues require that information from all these areas be integrated into a "macro" model that will address continuing problems in the medical system. Changes in health systems may come through a regulatory model or a procompetitive model, but without an integrated flow of information about all four areas, evaluating or recommending alternative strategies of medical delivery will be difficult.

\* From the System Sciences Department, The Rand Corporation, Santa Monica, California, and the Departments of Medicine and Public Health, Center for the Health Sciences, University of California at Los Angeles, Los Angeles, California.

† From the Behavioral Sciences Department, The Rand Corporation, Washington, D.C.

Supported in part by the Health Insurance Study grant 016880 from the U.S. Department of Health and Human Services, Washington, D.C. Other support was provided by the Commonwealth Foundation, the Robert Wood Johnson Foundation, and the Hartford Foundation through their funding of the Health Services Utilization Study.

The opinions and conclusions expressed are solely those of the authors and should not be construed as representing the opinions or policy of The Rand Corporation, of sponsors of its research, or of any agency of the United States Government.

Address correspondence to: Robert H. Brook, The Rand Corporation, 1700 Main Street, Santa Monica, CA 90406.

### A Brief Definition of Terms

Efficacy, effectiveness, variations in use, and quality of care are familiar ideas, so familiar, perhaps, that imprecision as to their meaning in this context impedes thinking about them "in the whole." In this article, we mean the following:

"Efficacy" refers to "the probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use."<sup>1</sup> This benefit (i.e., what a technology *could* do, not what it does) should be judged as comprehensively as possible, reflecting positive and negative aspects of physical and mental health. We can currently weight these outcomes and account for patient preferences for various states of ill (or good) health; a decade or more of significant research into health status measurement has given us many such tools.

"Effectiveness" has all the attributes of efficacy except one: It reflects performance under ordinary conditions by the average practitioner for the typical patient. In quality-of-care terms, what your doctor or mine does for you or me in the daily course of events is measured in terms of effectiveness.

"Variations in use" most commonly refers to different observed levels of per capita consumption of a service, especially hospital care, office visits, drugs, and specific procedures. Generally, these variations are seen as especially significant when all the usual explanations for use, such as demographic, social, economic, and health status factors, have been controlled, leaving no obvious explanation for differences except those related to practice style of the individual provider. Typically, variation studies do not concern themselves with the outcomes of care (i.e., variations in effectiveness).

Avedis Donabedian, dean of the "quality-of-care" field, defined quality of care as ". . . that kind of care which is expected to maximize an inclusive measure of patient

welfare, after one has taken account of the balance of expected gains and losses that attend the process of care in all its parts."<sup>2</sup> Within a research framework, the quality of medical care is that component of the difference between efficacy and effectiveness that can be attributed to care providers, taking account of the environment in which they work. "Quality assessment" is the act of detecting and measuring that difference, including variations across regions and peoples. In practical terms, it is the measurement of the technical and interpersonal aspects of medical care.†

### The Past as Prologue

#### The Efficacy of Medical Practice

We do know something about the efficacy of medical practice, but not as much as we would like.§ For regulatory needs (e.g., establishing efficacy and safety of a new drug) and research reasons (e.g., testing a chemotherapy protocol in a cancer patient) the efficacy of many drugs and drug combinations has been tested. In other therapeutic realms, we tend to know more about the efficacy of costly or innovative surgical procedures or other technologic break-

† Lohr and Brook distinguish between quality assessment and quality assurance and define the latter as ". . . the formal and systematic exercise of identifying problems in medical care delivery, designing activities to overcome the problems, and carrying out follow-up steps to ensure that no new problems have been introduced and that corrective actions have been effective."<sup>3</sup> See Komaroff in this issue for a more complete discussion of quality assurance.<sup>4</sup>

§ In 1978, the Office of Technology Assessment (OTA) reviewed the efficacy of 17 different services, ranging from a variety of preventive services (e.g., Pap smears for cervical cancer, chicken pox vaccine) to diagnostic procedures (e.g., mammography, electronic fetal monitoring) and therapeutic interventions (antibiotics in otitis media, cast application in forearm fractures) and surgical procedures (e.g., hysterectomy, tonsillectomy).<sup>1</sup> Among other things, OTA concluded that many technologies are not adequately assessed for efficacy (or safety) before they come into widespread use. Even among practices that are deliberately studied, findings may be equivocal. The situation has not changed appreciably in the intervening years.

throughs than about traditional or ordinary practices, simply because the latter became widely accepted well before concerns with efficacy entered the policy or research domain and before appropriate research methods were available to investigate them. We know less about the efficacy of many diagnostic tests and procedures, and certainly less about tests and procedures used in combination or about using a given test or combination of tests more than once in a defined time period. Finally, we know next to nothing about the efficacy of "cognitive" practices, such as listening to, counseling, or reassuring patients.

Efficacy is often established with regard to patients who are not representative of those for whom the service is ultimately intended. To illustrate the point, consider the literature on two diagnostic procedures: colonoscopy of the lower gastrointestinal tract and endoscopy of the upper gastrointestinal tract.||

In terms of physician charges, colonoscopy ranks among the 50 procedures done most frequently for the Medicare population, yet no randomized controlled trial has tested the efficacy of this procedure in general, for a specific diagnostic purpose, or in the elderly. A few prospective studies have followed patients who have undergone this procedure. Almost all information on colonoscopy comes from retrospective studies carried out mainly by physicians who do the procedure.

Endoscopy is another popular procedure for which our knowledge of efficacy is meager. A few randomized controlled trials have tried to establish its efficacy for patients with bleeding from the upper gastrointestinal tract. However, for the majority

of patients who receive this procedure (i.e., patients with dyspepsia or nonbleeders), data about efficacy in terms of information gained or in relation to an upper gastrointestinal x-ray series or a trial of medication to relieve symptoms are nonexistent.

The lack of information on efficacy is so obvious that more and better clinical trials and other efforts to establish efficacy will be needed. Research will also be needed for developing a) better methods for conducting randomized clinical trials (e.g., experimental design, ways to measure and assign values to patient outcomes, disease severity, and comorbidity) and b) better quasi-experimental designs, when using randomized trials is not feasible. With regard to the latter, improved ways to exploit secondary data bases or insurance claims files and to follow patients who are potential candidates for a procedure will be needed.

#### The Effectiveness of Medical Practice

If we accept for the moment the supposition that the medical profession knows less about the efficacy of what it does than it would like to know, then that problem is compounded as regards the ordinary course of patient care because even less is known about effectiveness than about efficacy.

Granted, we know a good deal about the effectiveness of many preventive practices such as immunizations provided in a community setting; similarly, we know which antibiotics are efficacious against which organisms, so by implication we should know for which patients they are likely to be effective. We know the efficacy of setting broken legs, replacing hips, or carrying out intraocular lens procedures *when judged against the outcomes expected if one does nothing*. We might even be comfortable with saying that the efficacy of surgical procedures such as hernia repair, cholecystectomy, and appendectomy are established for patients with a classical clinical presentation, as are modes of care for patients with angina, hypertension, or diabetes.

---

|| As part of a current study on variations in use of services, extensive literature reviews have been prepared for six major procedures: coronary artery bypass surgery, cholecystectomy, gastrointestinal endoscopy, fiberoptic colonoscopy, carotid endarterectomy, and coronary angiography. Comments here and elsewhere on these services are taken from those publications.<sup>5,6</sup>

Having evidence about the efficacy of a given intervention does not mean we can translate that knowledge appropriately into guidance about the daily practice of medicine. Take, for instance, treatment of Stage I cancer of the cervix. Studies of the efficacy of radiotherapy for cervical carcinoma show that it provides the same cure rate as surgery but with lower complication rates,<sup>7</sup> suggesting it should be the recommended therapy. What is not known is whether the same holds true in the community, because differing skills of the average surgeon or the typical radiotherapist or varying attributes of the hospital may attenuate or magnify the outcomes expected on the basis of efficacy studies. In other words, we cannot be sure that what is learned in the best possible circumstances provides a reliable guide in ordinary circumstances. Just as we may be uncertain about the efficacy of much medical practice today, we may be equally uncertain as to effectiveness *even when* we know something about efficacy.

What is needed might be characterized as an epidemiology of effectiveness: some way of routinely collecting information that describes the outcomes of tests, procedures, drugs, and other services as they are customarily used in everyday practice. Current health services research resources would never permit the study of both efficacy and effectiveness of the wide array of services about which our knowledge is sparse, let alone investigation of why gaps between the two appear. Our view is that better returns from health services research dollars would come from examining the effectiveness of medical care and describing the magnitude of the difference between efficacy and effectiveness when efficacy is known.

The goals would be to reduce the range of uncertainty facing physicians in their day-to-day practice of medicine and to illustrate the limited utility of efficacy studies in the absence of effectiveness studies. If these goals were accomplished, perhaps a

longer-range outcome would be more research funding to study both efficacy and effectiveness.

Finally, concerns with effectiveness should shift away from services per se to *patients* and their problems, to *combinations* of services that may produce the best outcomes, or to *interactions* of variables (region of the country, type of hospital market area, type of patients, class of hospital, skills of physicians, attributes of other care providers, and so forth) that may modify or moderate the "average" effectiveness of the service under investigation.

#### Variations in Use of Services

Variations in use and outcomes of medical services are large and ubiquitous, as empirical evidence amply shows.\* Throughout the developed world, the per capita use of procedures varies by two- to 12-fold in ways that are not satisfactorily explained by population characteristics or differences in available resources. Rates of hospital admissions for medical reasons may also vary by eight-, ten-, or even 12-fold across hospital market areas.

Such large variations are seen irrespective of the mechanisms used to pay providers and of differences in national health care delivery systems. Moreover, some evidence from the United States, Canada, and other developed countries suggests that variations within geographic areas in per-person use or medical expenditures is as great as those across areas, reinforcing the notion that differences in practice style and organization at the level of individual practitioners may lie at the heart of variations in use of services.

Uncertainty about the appropriateness or effectiveness of a given service provided to

\* For an up-to-date compendium of articles on this topic, see the summer issue of *Health Affairs*.<sup>15</sup> Data on medical admissions were reported by Wennberg.<sup>16</sup> Articles included in the section on practice patterns provide more information on this topic.<sup>17</sup>



a given patient for a specific condition or complaint allows for differences of opinion and thus may account for much of the high rates of variation in use. Hence low variation rates may reflect a fairly high degree of consensus about effectiveness of a given procedure or service (although this conjecture remains to be proven). The very lowest rates, however, could represent underuse and failure to provide needed services. The highest rates, by contrast, may often represent overuse of services, the effectiveness of which is at best incompletely demonstrated or at worst very questionable.

These phenomena imply, virtually by definition, that the quality of medical care may vary markedly across population groups. What we do not know, at this juncture, is the range of variation in quality of care (let alone the explanations for it).

Let us be clear about the problem: We know little about the effectiveness of many aspects of medical practice, so we have little to go on in establishing quality-of-care standards. That is, we cannot easily judge the degree to which patient outcomes are optimized. We know that use of many services varies tremendously throughout the nation, often for no obvious reasons; amorphous notions about practice style, group consensus, and tradition, are thus offered as possible explanations. Without good measures of quality (i.e., expected levels of effectiveness), we cannot readily evaluate what low or, especially, high rates of variation mean in quality-of-care terms. Hence we cannot sensibly decide where medical care consumption can or cannot justifiably be constrained.

We emphasize that investigating the *clinical* appropriateness of such variation is crucial on two scores. First, documenting wide ranges of "unexplained" per-person use of specific services essentially tells us where to concentrate resources for research into the effectiveness and quality areas, on the reasoning that reducing physician uncertainty about expected benefits could lead

to better decisions about optimal therapeutic choices. Second, decisions about health care financing still must be made. In the absence of good clinical data that would justify use of specific services, cuts may be made arbitrarily in high-use and/or high-cost services. Good quality-of-care research will help guard against ill-advised reductions in the use of those services that are more effective (i.e., bring about demonstrably better patient outcomes) than their alternatives.

#### Assessing the Quality of Medical Care

Two decades of research support for quality assessment, much of it by NCHSR, have vastly improved the tools of our trade.<sup>1</sup> What has not yet happened is the application of those tools beyond rather narrow confines of assessing quality of care for specific diagnoses, particular groups of patients, or types of institutions. That is, quality assessment has not been integrated into larger models of health care delivery that would allow us to understand the meaning or ramifications of variations in the use of services or to specify the permissible ranges of effectiveness.

*Conceptualization of Quality of Care.* Much progress has come just in the definition and refinement of the concepts of quality of care. In this regard, the classic work by Donabedian was especially important. The constructs of structure/process/

---

<sup>1</sup> We could not possibly review here the vast literature in quality of care measurement that accrued during the 1960s and 1970s. Seminal work has been done by Brook, Donabedian, Egdahl, Greenfield, Kane, Kessner, Palmer, Payne, Rutstein, Sanazaro, Starfield, Ware, Williamson, and their several associates; selected citations appear in the bibliography.<sup>18-44</sup> For related work, see the article by Bergner on health status measurement.<sup>45</sup> We thank Jim McAllister of NCHSR for providing some information about work the Center has supported in the past. We have restricted ourselves to discussing the quality of *medical* care in this article, but it should be clear that many of the same points might be pertinent for dental, mental health, and other forms of care.

outcome are well established, as is the notion that the science and the art of care are interrelated but distinguishable.

*Evaluating the Medical Process.* Probably the greatest advances have come in ways to evaluate the process of medical care (i.e., what is done to or for a patient with respect to his or her particular disease or complaint). The concept of diagnosis- or problem-specific "tracers" has won wide acceptance; an alternative approach that might be characterized as applying "comprehensive criteria sets" to randomly selected patient records is also gaining prominence.

Twenty years ago few, if any, diagnosis- or symptom-specific criteria sets could be "pulled off the shelf." Today, one would not have to search far to find at least a few criteria sets for a variety of common conditions that afflict adults, children, or both. Work by Heather Palmer and her colleagues at Harvard on the Ambulatory Care Medical Audit Demonstration is a case in point. Unfortunately, as of this writing, no system is in place to generate such criteria on a continuing basis, to update them, or to validate them against proximate or ultimate outcomes.

"Criteria maps," which were developed through the UCLA EMCRO,\*\* incorporate decision logic (i.e., contingent or branching criteria).<sup>25-26</sup> They provide a powerful tool for quality assessment and quality assurance programs, and they may also lend themselves to development of protocols for use by both physician and nonphysicians. This promising technique has thus far been applied only to a few conditions; hence criteria maps (validated or not) are not widely available.

*Computer-based Information.* Our ability to carry out secondary analyses on large

data bases, such as insurance claims files, for quality assessment purposes has improved drastically over the years. We can, for instance, apply fairly detailed criteria to episodes of care for acute conditions, as shown in work based on data from the New Mexico EMCRO. As another example, we can evaluate whether hospital admissions following surgery have been caused by a previous operative complication. Furthermore, various computerized information systems such as Computer-stored Ambulatory Record (COSTAR) permit information on patient/provider ambulatory encounters to be entered directly in a computer file.

*Methods Effects.* Different quality assessment methods will produce different conclusions regarding the level of quality. Because we now know something about the effects of "methods" on results, however, sources of potential bias and error can be postulated and tested. Moreover, specific clinical models that relate process to outcome can be built, along the lines of ones developed for measuring the quality of care given to children with iron-deficiency anemia.<sup>41</sup>

*Outcomes of Care.* Our fundamental concern in quality assessment is the end result of care (i.e., the patient's eventual outcome or health status). Much progress has been made in measuring health status, and we note here only the broad areas of development.

First, we think about patient outcomes differently than we used to: Over the years, our understanding of what health means to people and of the factors that might interact (i.e., need to be controlled for in a research sense) to produce a good or poor outcome has improved tremendously. Insights into the multifaceted nature of health status prompted much methodologic work that produced reliable and valid outcome measures. These include functional status and physical capacity measures, mental health inventories, the Sickness Impact Profile, and measures of a person's percep-

\*\* EMCROs, or Experimental Medical Care Review Organizations, were established as part of an NCI/ISR-initiated program to foster local peer review in the early 1970s.<sup>31-33</sup>

tions and ratings of his or her general health.<sup>45-48</sup> Unfortunately, progress in measuring disease-specific outcomes, especially for people with a chronic disease, has not kept pace with measurement of overall health status.

Second, we know a good deal more about measuring a patient's satisfaction with care.<sup>49-51</sup> Some aspects of patient satisfaction relate directly to quality of care, although this construct is neither a necessary or sufficient measure of the outcomes of care. As in the field of health status measurement, much of this developmental work was supported by NCHSR.

Third, a great deal of progress has been made in developing measures of hospital case mix and severity of illness.<sup>52-59</sup> Although these measures may be relatively specific (e.g., those developed for assessing the seriousness of condition of patients admitted into intensive care units), others attempt to deal with the full range of problems causing a hospital admission. In all these cases, incorporating such indexes into research will permit better understanding of the links between processes and outcomes and better comparisons of outcomes across providers.

*Deficiencies in the Quality of Care.* Quality assessment research has identified major deficiencies in medical care; many persist to this day. Poor care is seen in ambulatory and hospital settings alike. It is just a matter of time before deficiencies in newer settings such as surgi-centers or free-standing emergency centers begin to be documented.

Some deficiencies are well known and of long standing; inappropriate use of antibiotics is a good example. Other suspected problems that are not well documented involve the applications and uses of psychotherapy, the high rate of nosocomial infections in hospitals, and inappropriate use of "intensive" technology such as intensive care units. The quality of nursing home care, roundly and deservedly criti-

cized in the past, will continue to be a salient problem as the population ages.

### Where Do We Go From Here?

#### Research Needs in the Quality Assessment Field

*Process and Outcome.* Quality assessment ultimately should serve to improve the level of health in this country. No government agency, hospital system, or medical practice can afford to judge quality of care solely on the basis of outcomes. To discover some adverse outcomes produced by poor quality of care, for instance, might require collecting data from hundreds of patients. Moreover, assessing patient outcomes alone provides little opportunity to influence medical practice patterns; conversely, modifying medical practice on the basis of normative process criteria has not always yielded demonstrable improvements in patient health status.

Thus one critical gap in the quality assessment field is our inability to establish the linkages between the process and outcomes of care. We often are unable to say with confidence that the services rendered to the patient bear some medically plausible relationship to the patient's subsequent health status.

Establishing the *clinical* validity of process measures (i.e., the degree to which process predicts outcome) is a significant area of future research. Disease models that include comorbidity, stage, process, and outcome measures need to be built, tested, and validated so that they (or their components) can be used to evaluate quality of care. This area will need continuing attention.

*Nonintrusive Outcome Measures.* There remains a need for outcome measures that do not rely on obtaining information directly from patients or physicians and that do not require medical record data. Research into so-called "nonintrusive outcome measures" may be aimed primarily at large secondary data bases, often insurance claims files such

as those included in the Medicare Statistical System, those maintained by Blue Cross/Blue Shield, or state Medicaid systems. Computer-based ambulatory or hospital medical records offer another potentially important source of data.

Developing nonintrusive outcome measures requires linking such files with other, population-specific statistical files (such as eligibility files or death certificates); it also requires testing simple and complex measures (e.g., death, complications of surgery, or recurrent signs and symptoms following hospitalization). Validating "candidate" measures calls for use of expert review, process-of-care data, and directly measured outcome variables.

Such work may show, of course, that outcomes really need to be measured directly, that is, nonintrusive measures as presently conceived are too nonspecific, insensitive, unreliable, or invalid for quality assessment or program monitoring. If that is so, the point needs to be made explicitly, so that further research can be directed toward improving how nonintrusive measures can best be used and so that major policy decisions about health care financing and delivery will not be made on indefensible quality-of-care grounds.

*The Evolution of Disease and Patient Outcomes.* The need continually to refine our definitions of patient outcomes is prompted by the evolution of our understanding of disease in clinical terms. For example, conditions once thought incurable are now curable; or, at a minimum, we can say that for some illnesses, death is no longer an acceptable outcome (e.g., cervical cancer, bleeding ulcer, malignant hypertension). The population is aging, and the elderly population is itself aging. With more attention being turned toward this population, diagnoses (e.g., senility, Alzheimer's disease, and depression) that could not be made reliably 5-10 years ago are now being made. These necessitate wholly new ideas about appropriate patient outcomes. The

advent of new technologies forces us down a similar path: as "experimental" procedures such as liver transplants or artificial organs become more fully accepted or disseminated into medical practice, we are forced to redefine the outcomes of care that the nation should legitimately be able to expect.

In some cases, this may essentially be an issue of efficacy, when such procedures or services are regionalized to only a few "best" centers. We should be alert, however, to the degree and speed with which medical practices tend to diffuse into the community, because then the question becomes one of effectiveness. In this regard, the need for good process and outcome measures to use in larger "technology assessment" efforts, especially those that are disease- or problem-specific, should be clear.

*The Art of Care.* Implicitly, much of what has just been discussed concerns the technical (i.e., technologic or mechanical) aspects of care. Perhaps as important are issues relating to the art of care and to the "cognitive" services that physicians and other care-givers provide. We need to know more about physician-patient communication and interaction and how, for instance, physicians can best communicate reassurance and hope about their patient's prognosis. The strong push for recognition of and appropriate compensation for cognitive (as opposed to procedure-oriented) care by some physician specialty groups argues for early research in this area. Finally, how the art of care or physician style affects patient outcomes deserves much more critical attention.

*An Epidemiology of Quality of Care.* One important area of research may be simply to develop an "epidemiology of the quality of medical care." Just as we will continue to explore and document variations in per capita use of services across the nation, we will need to know how quality is distributed in the population. Describing the distribution of quality of care in epidemiologic terms means, first, knowing how to measure

quality and, second, knowing something about the relationship of physician characteristics to the process of care and, eventually, to the outcomes of that care.

For example, certain types of health financing mechanisms (e.g., prospective payment in Medicare) may foster "regionalization" of complex or costly procedures into high-volume institutions. Restricting some services to high-volume facilities has been shown to be positively associated with better outcomes (e.g., lower death rates), but high-volume centers could also achieve that volume by performing such services inappropriately often.

Research should be aimed at improving our understanding of the epidemiology of quality and at answering a number of important policy questions. For example, would regionalization of common surgical procedures, or of invasive diagnostic tests, or of *noninvasive* diagnostic services result in better patient outcomes? What physician characteristics or institutional arrangements are directly associated with these better outcomes? Finally, are there interactions between patient characteristics or disease variables and provider variables that augment or detract from the effect of regionalization?

Furthermore, addressing questions of both equity and efficiency requires knowledge of the levels of quality of care a) among patients who use medical care and b) among all individuals in the population. In research or program evaluation terms, we should know how good the care is for "recipients" and for "eligibles," if we are to make reasoned choices about ways to finance and deliver medical care.

We make the point about studying both users and eligibles here because nonuse of medical care can itself be a quality-of-care issue. Nonuse is not by definition poor quality of care: avoiding unnecessary hospitalizations, invasive diagnostic procedures, excessive use of drugs, and so forth may be positive aspects to lower use of medical

care. For some population groups, however, nonuse of care may be an important problem, leading to poorer health status generally and poorer outcomes once care is sought.

We need more work on population-based models of quality assessment, such as that used by Nutting in the Indian Health Service.<sup>60</sup> These models will help us pay greater attention to the quality of care for many groups: populations enrolled in Medicare and Medicaid; persons in alternative health plans (especially those established to control the costs of publicly funded medical care); and individuals choosing preferred provider organizations. In general, the quality of care received through *any* innovative financing scheme that is based on economic incentives to curb the use of services merits attention.

*Costs and Quality of Care.* It is trite to say that the overriding problem of the day concerns the nation's health care expenditures, not the quality of the services received. We know a great deal about costs of and expenditures on medical care; we know a fair amount about what individual services may "cost" or at least what price they command in the market. We recognize there is a direct relationship between the price of medical care faced by the patient and the level of use of services. Finally, we know that the major public debates in the health field for the rest of the century will focus on medical care financing issues.

What we still do not know much about is the relationship, indeed the tension, between costs and quality—the so-called cost-quality trade-off. Further conceptual and empirical research will be needed on how to measure the direct and indirect costs and benefits of medical care, where benefits in this context refer to expected patient outcomes. Added to this is the need to be able to take into account patient "utilities" or preferences for alternative states of health and alternative levels of financial burden. Derived from this is the need to look at how to weight the preferences of different

factors (i.e., the utilities of patient versus family versus provider versus community versus society as a whole).

Choosing among alternative forms of care, when they differ in costs, requires knowledge about the expected benefits of those forms of care. Further, knowing when to provide more care and when to stop requires knowing something about the *marginal* costs and returns to care. In short, the domain of quality-assessment research of the future must be defined to include these cost-quality relationships, focusing on the question of what dimensions of health and patient outcomes to put on the benefits side of the balance scale.

### **An Integrated Model**

#### **A Simple View of the Policy Context**

We have briefly outlined important quality assessment issues that in our view should form the heart of the third decade of health services research. Reasonable people may differ with our selection and recommendations, believing, for instance that more (or less) resources should go to policy-oriented work or to refining methods and measures. Those debates we leave for another forum.

We would like here to return to the theme introduced at the outset of the article: research into quality of care (or any other topic, for that matter) can no longer be carried out in splendid isolation from other issues. We have suggested that a large research effort (perhaps the most important one) should attempt to integrate a) the efficacy of care where circumstances are most favorably disposed to produce good outcomes, b) the effectiveness of care in the daily course of events, c) population-based variations in the use of services and the intimately related variations in quality of care (measured in terms of patient outcomes), and finally d) levels of quality of care in the broadest possible terms. Proposing, testing, and investigating a model of this sort implies attention to the current

and future policy context of health care delivery and financing.

For purposes of this discussion, we would argue that the policy context involves continuing alarm about medical expenditures and a myriad of public and private sector efforts to control costs, increasing reliance on an as-yet unproven "competitive" approach to cost containment, a growing acceptance of the notion that medical care will be rationed (whether or not the care is needed), and a perception (strong in some quarters, dim in others) that members of society at greatest risk of harm from rationing (the elderly, perhaps the very young, the poor) need special protection as the financing issues work themselves out. Within this policy context, obviously, the interesting research questions are legion.

#### **Planning a Research Agenda**

Developing an explicit "hierarchy" of research needs will itself be an important step as the health services research community sets its agenda for the rest of the century. In the quality assessment area, criteria for making those choices should reflect the following points: a) the population groups most affected by actions elsewhere in the policy arena; b) the most common medical problems; c) the problems we think research could do the most to solve; d) the topics we know the least about; and e) the external factors we think will provoke the most serious future problems in the delivery of care (such as growth of transplant and artificial organ technologies or marked aging of the population).

*Clinical Trials.* One special arena of research calling for careful preparation concerns efficacy—more precisely, the planning and conduct of clinical trials. Better research tools and measurement techniques, including more reliable and valid measures of patient outcomes, are needed. Work in these areas can contribute to stronger bridges to the biomedical community over which ex-

isting and new knowledge can be more effectively passed. Research on how best to disseminate the products of such trials to the medical community and the policymaking branches of government should also be done.

*The "Macro" Model.* A second point is that the planning process for clinical trials must itself be integrated with an "operational" system that will track the use of services under investigation in the country as a whole. That is, clinical trials of the *efficacy* of a service should not be performed without prior attention to how *effectiveness* will be monitored, should the trial show positive results.

From the beginning attention should be given to whether outcome measures will be suitable for long-term monitoring. For instance, detailed physiologic variables that require direct patient contact and perhaps even invasive testing, although probably considered by clinicians to be the best efficacy measure, are likely to be unsuitable for measuring effectiveness and for long-term monitoring of the use of that service. Other less intrusive outcome measures should be sought from the beginning and validated against the physiologic variables tracked during the clinical trial. Perhaps the integration of clinical trials with evaluation of effectiveness and quality, coupled with greater attention to variations in population-based rates of use, will be the hallmark of the next level of activity in health services research.

#### Summary

Ideally, we need a fully integrated system of data collection and information dissemination. We need to be able to discard those medical procedures or services that are not efficacious and to quantify and understand the effectiveness of efficacious services. When efficacy and effectiveness diverge dramatically, then we need to understand issues relating to the quality of care rendered by the average physician.

Differences in efficacy and effectiveness are measured in terms of patient outcomes, but explanations of *why* patient outcomes differ partly reside in measures of the technical and interpersonal aspects of the care process (other things such as patient compliance more or less equal). When differences in the process of care do not explain differences in outcomes, we may look to structural characteristics of the medical delivery system as a whole for clues. Finally, we need to ensure that effective services are being provided to the appropriate groups of patients in ways that minimize inequitable variation in per person use of those services across geographic regions and population groups.

One crucial requirement for such a system is that research be designed to be internally valid, generalizable to the outside world, and important in a policy sense. To do this presupposes a commitment to integrate efficacy studies with measures of effectiveness, with measures of the structure, process, and outcomes of care, and with epidemiologic studies of variations in the use of services. In the long run, this research might serve to unite the epidemiologic and clinical approaches into one—a clinicoepidemiologic model of health—and produce the information and insights needed to help prevent the rationing of effective medical care.

#### References

1. Office of Technology Assessment. Assessing the efficacy and safety of medical technologies. Washington, DC: Congress of the United States, Office of Technology Assessment, OTA-11-75, 1978, p. 16.
2. Donabedian A. Explorations in quality assessment and monitoring. Vol. 1. The definition of quality and approaches to its assessment. Ann Arbor, MI: Health Administration Press, 1980.
3. Lohr KN, Brook RH. Quality assurance in medicine: experience in the public sector. Santa Monica, CA: The Rand Corporation, R-3193-HHS, 1984.
4. Komaroff A. Quality assurance. *Med Care* 1985;23:723.
5. Kahn KL, Fink A, Roth CP, et al. A review of the literature on fiberoptic colonoscopy: findings, complications, utilization rates, costs, efficacy, and indica-

- tions. Santa Monica, CA: The Rand Corporation, R-3204/4-CWF/PMT/HF (In press).
6. Kahn KL, Roth CP, Fink A, et al. A review of the literature on gastrointestinal endoscopy for diagnosis: findings, complications, utilization rates, costs, efficacy, and indications. Santa Monica, CA: The Rand Corporation, R-2304/3-CWF/PMT/HF (In press).
  7. Casciato DA, Lowitz BB. Manual of bedside oncology. Boston, MA: Little Brown & Co., 1983.
  8. Wennberg JE. Dealing with medical practice variations: a proposal for action. *Health Affairs* 1984;3:6.
  9. Sammons JH, Rubin R, Hackbarth G, et al. ReViews. *Health Affairs* 1984;3:33.
  10. Brook RH, Lohr KN, Chassin M, et al. Geographic variations in the use of services: do they have any clinical significance? *Health Affairs* 1984;3:63.
  11. Eddy DM. Variations in physician practice: the role of uncertainty. *Health Affairs* 1984;3:74.
  12. Schwartz JS. The role of professional medical societies in reducing variations. *Health Affairs* 1984;3:90.
  13. Vladeck BC. Variations data and the regulatory rationale. *Health Affairs* 1984;3:102.
  14. Caper P. Variations in medical practice: implications for health policy. *Health Affairs* 1984;3:110.
  15. Lohr KN, Lohr WR, Brook RH. Update. A London conference. *Health Affairs* 1984;3:139.
  16. Wennberg J, McPherson K, Caper P. Will payment based on diagnosis-related groups control hospital costs? *N Engl J Med* 1984;311:295.
  17. Hulka B, Wheat J. Patterns of utilization. *Med Care* 1985;23:438.
  18. Brook RH, Davies-Avery A, Greenfield S, et al. Assessing the quality of medical care using outcome methods: an overview of the method. *Med Care* 1977;15(Suppl):1.
  19. Brook RH. Quality of care assessment: a comparison of five methods of peer review. Rockville, MD: National Center for Health Services Research and Development, (HRA) 74-2100, 1973.
  20. Brook RH, Davies-Avery A. Mechanisms for assuring quality of U.S. medical care services: past, present, and future. Santa Monica, CA: The Rand Corporation, R-1939-HEW, 1977.
  21. Brook RH, Williams KN, Avery AD. Quality assurance today and tomorrow: forecast for the future. *Ann Int Med* 1976;85:809.
  22. Donabedian A. Explorations in quality assessment and monitoring. Vol. 2. The criteria and standards of quality. Ann Arbor, MI: Health Administration Press, 1982.
  23. Egdahl RH, Gertman PM. Technology and the quality of health care. Germantown, MD: Aspen Systems Corporation, 1975.
  24. Egdahl RH, Gertman PM, eds. Quality assurance in health care. Germantown, MD: Aspen Systems Corporation, 1976.
  25. Greenfield S, Nadler MA, Morgan MT, et al. The clinical investigation and management of chest pain in an emergency department: quality assessment by criteria mapping. *Med Care* 1977;15:898.
  26. Greenfield S, Cretin S, Worthman LG, et al. Comparison of a criteria map to a criteria list in quality-of-care assessment for patients with chest pain: the relation of each to outcome. *Med Care* 1981;19:255.
  27. Kane RL, Gardner G, Wright DD, et al. A method for assessing the outcome of acute primary care. *J Fam Pract* 1977;4:1119.
  28. Kane RL, Bell RM, Hosek SD, et al. Outcome-based reimbursement for nursing home care. Santa Monica, CA: The Rand Corporation, R-3092-NCHSR, 1983.
  29. Kessner DM, Kalk CE. A strategy for evaluating health services. In: Contrasts in health status. Vol. 2. Washington, DC: National Academy of Sciences, Institute of Medicine, 1973.
  30. Kessner DM, Snow CK, Singer J. Assessment of medical care in children. In: Contrasts in health status. Vol. 3. Washington DC: National Academy of Sciences, Institute of Medicine, 1974.
  31. Brook RH, Williams KN. Evaluation of the New Mexico peer review system. *Med Care* 1976;14(Suppl 12):1.
  32. Lohr KN, Brook RH, Kaufman MA. Quality of care in the New Mexico Medicaid program (1971-1975). *Med Care* 1980;18(Suppl 1):1.
  33. Lohr KN. Quality of care in episodes of common respiratory infections in a disadvantaged population. Santa Monica, CA: The Rand Corporation, P-6570, 1980.
  34. Lohr KN, Winkler JD, Brook RH. Professional standards review organizations and technology assessment in medicine. Santa Monica, CA: The Rand Corporation, R-2820-OTA, 1981.
  35. Palmer RH, Strain R, Maurer JW, et al. Quality assurance in eight adult medicine group practices. *Med Care* 1984;22:632.
  36. Palmer RH, Strain R, Maurer JW, et al. A method for evaluating performance of ambulatory pediatric tasks. *Pediatrics* 1984;73:269.
  37. Payne BC. The quality of medical care: evaluation and development. Chicago: Hospital Research and Educational Trust, 1976.
  38. Payne BC, Lyons TF, Neuhaus E. Relationships of physician characteristics to performance quality and improvement. *Health Serv Res* 1984;19:307.
  39. Rutstein DD, Berenberg W, Chalmers TC, et al. Measuring the quality of medical care. A clinical method. *N Engl J Med* 1976;294:582.
  40. Sanazaro PJ. Quality assessment and quality assurance in medical care. *Annual Review of Public Health* 1980;1:37.
  41. Starfield B, Scheff D. Effectiveness of pediatric care. *Pediatrics* 1972;49:547.
  42. Williams KN, Brook RH. Quality measurement and assurance. *Health and Medical Care Services Review* 1978;1:1.
  43. Williamson JW. Assessing and improving out-



comes in health care: the theory and practice of health accounting. Cambridge, MA: Ballinger Publishing Company, 1978.

44. Williamson JW. Improving medical practice and health care. A bibliographic guide to information management in quality assurance and continuing education. Cambridge, MA: Ballinger Publishing Company, 1977.

45. Bergner M. Measurement of health status. *Med Care* 1985;23:000.

46. Bergner M, Bobbitt RA, Carter WB, et al. The sickness impact profile: development and final revision of a health status measure. *Med Care* 1981;19:787.

47. Ware JE, Jr., Brook RH, Davies AR, et al. Choosing measures of health status for individuals in general populations. *Am J Public Health* 1981;71:620.

48. Brook RH, Ware JE, Jr., Rogers WH, et al. Does free care improve adults' health? Results from a randomized controlled trial. *N Engl J Med* 1983;309:1426.

49. Ware JE, Jr., Davies-Avery A, Stewart AL. The measurement and meaning of patient satisfaction: a review of the literature. *Health and Medical Care Services Review* 1978;1:1.

50. Ware JE, Jr., Synder MC, Wright WR, et al. Defining and measuring patient satisfaction with medical care. *Evaluation and Program Planning* 1983;6:247.

51. Ware JE, Jr., Davies AR. Behavioral consequences of consumer dissatisfaction with medical care. *Evaluation and Program Planning* 1984;6:291.

52. Horn S, Horn R, Sharkey P. The severity of illness index as a severity adjustment to DRGs. *The Health Care Financing Review*. 1984 annual supplement. HCFA Publ. No. 03194, 1984.

53. Young W. Incorporating severity of illness and comorbidity in case mix measurement. *The Health Care Financing Review*. 1984 annual supplement. HCFA Publ. No. 03194, 1984.

54. Kominski G, Williams S, Mays R, et al. Unrecognized redistributions of revenue in diagnosis-related group-based prospective payment systems. *The Health Care Financing Review*. 1984 annual supplement. HCFA Publ. No. 03194, 1984.

55. Gertman P, Lowenstein S. A research paradigm for severity of illness: issues for the DRG system. *The Health Care Financing Review*. 1984 annual supplement. HCFA Publ. No. 03194, 1984.

56. Wagner D, Knaus W, Draper E. Acute physiology and chronic health evaluation (APACHE-II) and Medicare reimbursement. *The Health Care Financing Review*. 1984 annual supplement. HCFA Publ. No. 03194, 1984.

57. Smits H, Fetter R, McMahon L. Variation in resource use within DRGs. The severity issue. *The Health Care Financing Review*. 1984 annual supplement. HCFA Publ. No. 03194, 1984.

58. Conklin J, Lieberman J, Barnes C, et al. Disease staging: implications for hospital reimbursement and management. *The Health Care Financing Review*. 1984 annual supplement. HCFA Publ. No. 03194, 1984.

59. Thompson J. The measurement of nursing intensity. *The Health Care Financing Review*. 1984 annual supplement. HCFA Publ. No. 03194, 1984.

60. Nutting P, Shorr GI, Burkhalter BR. Assessing the performance of medical care systems: a method and its application. *Med Care* 1981;19:281.

**RAND/N-3368-HHS**