Assessing Clinical Competence

Victor R. Neufeld, M.D., FRCP(C)
Geoffrey R. Norman, Ph.D.
Editors

Foreword by William B. Spaulding,
M.D., FRCP(C), FACP

SPRINGER PUBLISHING COMPANY
New York
This section introduces the reader to the reasons for writing this book. The first chapter traces three historical perspectives that have been concerned with clinical competence. These are the influence of the profession, the perspective of society, and the impact of research and development. These three “force fields” converge to raise important questions about our current understanding of clinical competence, how we measure it, and how we apply our understanding. The chapter closes with a brief description of the scope and style of the book.

In Chapter 2, we review various approaches that have been used to come to an understanding of what is meant by clinical competence. Several distinct perspectives are presented. These include the “armchair” approaches, task analysis and other formalized methods used by specialty boards and licensing examinations, and the indirect viewpoint on competence provided by a variety of research methods—epidemiological, psychological, and quality of care.
Clinical competence is a battered child: a child because it is a relatively new area of interest, with most of the research and development having been done in the past 10 to 15 years; battered because it has been mistreated by researchers, who cannot agree on what it is and how to measure it. Consider these examples of confusion about clinical competence:

Ask any number of clinical educators or professional groups to define the term *clinical competence* and somewhat different responses will be obtained from each of them. Until recently, there has been little attempt to study what a clinician actually does, and it is even rarer to find performance assessment procedures which are systematically based on new insights into clinical methods.

Measures of clinical competence have been described in the literature which appear to have been adopted prematurely, without the benefit of step-by-step methodological development. In some cases, the measures are in the early "pilot project" stage or are simply untested ideas.
Systems and institutions have arisen outside of the medical profession, each of which are concerned with some aspect of clinical competence. These responsibilities may include the selection of applicants to a particular training program, the program itself, and the licensing, certification, or recertification of individual physicians. Frequently, the boundaries of responsibility are unclear. More importantly, it is unusual for questions such as these to be asked: Is the system working? What is the evidence that an intended result is being achieved? Given a particular approach, how does it compare with a more standard or alternate regimen?

This book aims to clarify some of these issues. Primarily, it is written for clinical educators—the people in academic institutions and other professional organizations who are the users of the research on clinical competence. The book should also be useful to methodologists who are doing the research in this area.

Historical Perspectives

For many decades, the medical profession has struggled with the question of competence. The motivation to do so has come from a variety of sources. Three of the most important are the influence of the profession itself; the influence of society; and the influence of research and development. These three are interrelated and overlap to some extent. We shall first look at them separately, and then examine the combined effects of these “force fields” on the profession at the present time.

The Influence of the Profession

The medical profession, through its various institutions, has used a number of different strategies to assure the competence of clinicians. A recent book (Samph and Templeton, 1979) reviews the role of various American professional and government institutions in assuring competence. While this panorama of involved groups is perhaps more complicated in the United States than in other countries, a similar pattern could be described in Canada, Great Britain, and other countries of the world. Some of these patterns will be described later in the book.

In general terms, the profession has used three strategies to
produce competent clinicians: the accreditation process, the licensing and certification process, and specific training programs.

The Accreditation Process. The earliest approach used in North America, accreditation was based on the assumption that if the setting in which physicians are trained met acceptable standards, then the performance of the individual professionals should be acceptable. The American Medical Association as early as 1905 devised a rating system consisting of a number of categories for assessing medical schools. Since those early days, the accreditation process has become vastly more complex and has spread along the “continuum of medical education” to include postgraduate (residency training) programs, then lately extending into programs in continuing medical education.

Paralleling the accreditation of training programs has been the accreditation of health care institutions, particularly hospitals, but also ambulatory care settings. Now called “quality assurance,” this activity, particularly in the United States, has swollen into a large and complicated bureaucracy, stimulated by the legislation which gave rise to the Professional Standards Review Organizations (PSRO), and is having a significant impact on the struggle to understand clinical competence.

From a methodological point of view, the question could be asked: What is the evidence that the accreditation process, in its various forms, assures the competence of individual clinicians; and does it have any significant effect on health outcomes? The evidence to date is not particularly encouraging.

Licensing and Certification. The second strategy used by the profession to assure clinical competence is that of licensing and certification, focusing on the performance of the individual practitioner in contrast to that of an institution. The history of the process in the United States is well described by Derbyshire (in Samph and Templeton, 1979). It is a fascinating story. There has been much change from the time, a number of decades ago, when senior retired clinicians on state medical boards grilled candidates with archaic questions (“describe mercury inunction”), to the current use of computer-marked comprehensive multiple-choice examinations.

The certification story is similar. From virtually no standards other than some sort of acceptable apprenticeship, there has been a steady evolution to prescribed training programs and structured certification examinations. This trend is influenced by
the development of competence in individual trainees, including the selection system, the curriculum itself, and the evaluation system.

Medical schools in North America (and, to some extent, residency programs) are in a luxury situation. There are many more candidates than places, currently about three candidates per place. Many of the unsuccessful applicants are probably quite acceptable and may eventually become adequate physicians. It has been observed that important qualities required for successful completion of a training program and for professional practice are already present in applicants to medical school. Given this situation, an opportunity exists for devising selection methods and procedures which will most optimally match an applicant's ability and interest with the style and objective of a particular training program and eventually with the demands and opportunities of a professional practice. In the interests of educating future physicians well, there should be an optimal match between candidates and programs.

A curriculum is designed to prepare a certain kind of “product.” The term curriculum usually encompasses the formal or structured learning experiences of students—the lectures, laboratories, clinical tutorials and seminars, elective experiences, and direct participation in patient care under supervision. There is also considerable evidence that the “informal curriculum” is important in shaping future professionals. These informal factors include interaction with peers, unplanned and often unrecognized “role modeling” of supervising residents and attending physicians, unobtrusive influences of the medical school as a social entity, the personal and family demands on an individual, and the general social climate of the time (Funkenstein, 1977).

Of the many components of a training program, perhaps none is quite as influential as the manner in which student performance is assessed. Generations of medical students have discovered that the “bottom line” of getting through medical school involves understanding the evaluation system and complying with the behavior pattern that it demands. It is in the evaluation system that the “real” objectives of any program are displayed, and the truly important values become apparent.

These considerations again suggest a number of questions: To what extent does a training program in fact influence the eventual performance of a clinician? Which component is the most important: the selection system, the structure of the curriculum, the “informal” learning environment, or the evaluation system? To
the phenomenon of increasing subspecialization. The American Board of Medical Specialties, for example, now consists of 23 separate boards, each with its own approach to certification. The Royal College of Physicians and Surgeons of Canada has 36 individually recognized medical and surgical specialties. Over the years, each specialty group has come to define the required training, usually stated in terms of a prescribed number of years and rotations, but rarely in terms of performance objectives.

There is a wide range in the format of certification examinations. Some are simple, using only multiple-choice examinations. Some are complex; an example is the certification examination of the Canadian College of Family Physicians, which includes written examinations, simulated patients, and simulated “office orals” in which the examiner is himself a patient, married to an equally complicated scoring system. An additional recent development in Canada is the use of institutional reports from residency program directors to contribute directly to the certification decision (Valberg and Firstbrook, 1977).

Finally, there is the phenomenon of relicensure and recertification. Virtually every group within the profession of medicine has in the last few years wrestled with the issue of assuring continuing competence of individual practitioners. The strategies employed range from encouragement to attend continuing education events through to mandatory recertification and relicensure. The methods proposed for the latter again are wide ranging, from self-assessed office medical records to computer-based patient management problems.

Again questions need to be asked from the methodological perspective: Which competencies are being tested in today’s examination for licensure and certification? Is it generally agreed that they are the most important? How good are the measures employed? Is there any evidence that the increasingly complex machinery for relicensure and recertification has any actual influence on the continuing competence of a physician? Do our current certification procedures discriminate accurately—that is, do they pass the sufficiently competent, but hold back those that are not ready for independent practice?

*Training Programs.* The third strategy employed by the profession to assure competence is the structured educational process itself. We have already referred to the accreditation of institutions, which attempts to ensure a basic level of quality in a training program. Several components of an education program can affect
what extent does each of these components relate to educational objectives, explicit or implicit? And do the objectives of a training program reflect the range of competencies required in actual professional life?

**The Influence of Society**

Modern society exerts an influence on its professions through formal mechanisms and through informal influences, such as the general climate in which potential physicians grow up, including the level of public awareness about health and disease. Examples of formal mechanisms are specific laws and policies governing professional behavior and the legal arrangements reflected in malpractice suits.

*Formal Mechanisms.* The legal statements which define the limits of professional responsibility tend to be "responsive"; that is, they formalize a system which is already in place. It is unlikely that they have a direct influence on the performance of individual practitioners.

Perhaps more influential are government policies that relate to "cost containment" and to "quality assurance." A current example is the PSRO legislation in the United States where a large and complicated bureaucracy has been constructed, aimed at controlling hospital costs through the application of guidelines on duration of hospital stay. A recent Canadian example involves the government of Ontario. In an attempt to control costs, the Ministry of Health has declared maximum hospital bed ratios per given population and is actually involved in closing down some smaller hospitals. In many countries, the government participates directly in medical insurance, resulting in a direct influence on physician fee structures and salary levels. This level of participation in the health system has a direct effect on what individual physicians and groups of physicians actually do in professional practice.

Another formal mechanism that influences physician performance is the malpractice phenomenon, prominent in the United States, where the legal tradition differs somewhat from that of Canada and Western Europe. Large sums of money are paid by American physicians for malpractice insurance, and the clinical behavior of physicians is directly influenced by the spectre of malpractice. This is evident in the practice of "protective medicine"—for example, the excessive ordering of laboratory and radiological investigations in an emergency room.
Informal Mechanisms. Related to the malpractice situation is the widely accepted observation that the public is more interested in and knowledgeable about health and disease. Patients, quite appropriately, demand explanations about an illness, its prognosis, investigation, and treatment. Further, we wish to participate in our own health management—a concept that expresses itself in various ways: self-health manuals, participation by partners in childbirth, holistic health clinics, consumer participation in health councils, and the proclamation of statements of patient rights. For the individual practitioner, this means spending more time on patient education and involving the patient and his family more in direct health care activities.

The recently published studies of Funkenstein and his colleagues (1977) report on the perception and career choices of Harvard medical students over a 20-year period, incorporating five social eras, and show that the values and to some extent the career choices of the graduates were influenced by the prevailing social environment.

These formal and informal influences that affect the practice situation and individual professional activity present for methodologists a number of challenges. For example, can reliable and valid methods be developed to select students who represent the social spectrum of society? To what extent do the factors influence physician performance? Can some of them be measured, and can physician behavior be altered in a way that makes health care less costly but still effective?

The Influence of Research and Development

A third stream of influence on the performance of clinicians is research and technological development, as seen in the availability of sophisticated technology to practicing physicians and in the rise of new fields of research which relate to clinical competence. Some of its facets will be described briefly.

Virtually every physician is affected by the availability of increasingly sophisticated technology. Through automated blood testing systems, laboratory results which were not ordered are frequently brought to the physician’s attention. They are answers to unasked questions, which, in turn, lead to new questions; and these new questions frequently are unrelated to the patient’s initial illness. Another example of rapidly advancing technology is represented in diagnostic imaging. Consider the available tests in a large hospital for investigating a suspected pancreatic mass.
Which devices should be used and in what sequence—a barium meal, angiography, or whole-body computerized tomography? The answers are by no means clear to practicing physicians.

A different example is the use of the computer in testing. Computers have been used for storing and retrieving multiple-choice questions (the so-called item bank); computers are also used increasingly as a basis for simulating clinical thinking. Probably the very availability of the technology was the main reason for developing the simulations, rather than any compelling logic to use a computer for purposes such as this.

There has been an increased interest in research in medical education. In the United States, this interest dates back approximately 20 years, when educational researchers joined faculties of medicine and other medical professional organizations. Stimulated by the desire to enhance teaching and evaluation, specific methodological techniques were applied to the area of clinical competence. There was also an increased awareness of concepts of educational measurement such as validity (in its various forms), reliability, and objectivity, which were used to examine existing methods of assessing competence and to design new assessment tools.

A related research area might be called “health care research,” a broad and heterogeneous area of investigation which includes the work of clinical epidemiologists, operations researchers, health economists, and biostatisticians. Much of this work is focused on the problem of developing health outcome measures and applying these measures to the pursuit of specific health care questions. The progress that has occurred in the “quality assurance” movement is attributable mainly to the work of this group of researchers. Both health care research and educational research have had a profound influence on definition and measurement of competence. These areas form the basis for the remainder of this book.

A Convergence of Forces

In this review of three “force fields” influencing clinical competence over the past two decades, we have discussed the activities of the profession itself, the less direct influence of society, and the influence of research and technological development. All have their effect on the clinical activity of an individual practitioner. These three developments appear now to be converging, producing much of the urgency and some of the confusion about our current understanding of clinical competence. Here are some examples:
In its concern for demonstrating to the public that physicians are competent, the profession has started down the path of recertification and relicensure. In some jurisdictions, these concepts have been taken to the stage of government mandate. But from a methodological point of view, this trend would appear to be premature. The same types of instruments used for initial certification are now being used (in the absence of anything better) for recertification. The irony is that these tools have not yet been demonstrated to possess adequate validity for certification. Quite apart from the whole question of whether recertification is an appropriate idea philosophically, it is highly doubtful from a methodological point of view whether we are ready for it.

The computer-marked multiple-choice examination is now the most widely used evaluation instrument; it is also a favorite format in certification. This approach seems to have been accepted as an adequate tool for these critical decision points in the career of an individual physician. It is generally agreed that the tool is useful for sampling the knowledge base of an individual, but not much more than this. This overreliance on one instrument has occurred despite repeated pleas from thoughtful clinician-educators that many other attributes should be included in a definition of competence. The same plea is heard from the public. A survey sponsored by the Ontario Medical Association, but conducted independently under the direction of Mr. Edward Pickering, reported, for example, that patients identified the main deficit in the medical profession in the areas of "human relations" and access to care (Pickering, 1973). This imbalance is particularly troublesome in the absence of any convincing evidence of a correlation between the demonstration of an adequate medical knowledge base and the actual performance of a physician (Wingard and Williamson, 1973).

From the health care research literature has emerged the sobering evidence that many health outcomes are not directly attributable to what physicians do (McKown, 1976). Through the use of such design strategies as the randomized clinical trial, many of the treatments we once thought to be beneficial turn out, in fact, not to be, and some result in more harm than good (Haynes
et al., 1978). Observations such as those should significantly influence the means by which the competence of physicians, both individually and as a group, are assessed. And yet it is extremely difficult to change behavior once a certain trend is established. As an example, consider the probably excessive use of coronary artery bypass surgery. Governments and private institutions are evidently prepared to fund cardiovascular surgery units—sometimes several in the same city. Patients ask for the procedure. And the profession, armed with the technical know-how, proceeds with this therapeutic strategy, in advance of convincing evidence of its specific effectiveness for certain conditions.

These situations demand clear and critical thought. Our hope is that this book will help the clinician-educator sort out some of the “fallout” from the convergence of these developments of recent history.

Scope and Style

For the most part, this book is limited to a discussion of the clinical competence of physicians since most of the available research has used physicians or student-physicians as subjects. Comprehensive reviews of clinical competence of other health professionals are available (Morgan and Irby, 1978). We think, however, that much of what we will be saying will also be useful to other health professionals.

A methodological perspective pervades this review. We are primarily concerned with the empirical basis for our understanding of clinical competence, the methods used to assess it, and the systems which assure its development or continuance. It follows that this book is not primarily a “how to do it” cookbook, or a sales pitch for a particular approach or device. Rather, it attempts to look as objectively as possible at the available evidence, with comments based on specific predetermined criteria.

There is a special emphasis on education and on research studies which were conducted in a learning environment. In contrast to such areas as general health care research and the rapidly growing field of “quality assurance,” there is a paucity of method-
logical reviews where clinical education is the primary focus. The contributors to this book are deeply involved in education and are seeking to apply critical thought to the learning setting and to their behavior as educators. Learning is fundamentally related to competence. We view educational measurement and evaluation as part of the learning process. And competence, both initial and continuing, is the end product of effective learning.

The contributors are all from the Faculty of Health Sciences at McMaster University, where for the past ten years or so some new ideas in medical education have been explored, and a number of the methods and approaches which will be described have been tested. It was thought that the flow of ideas in the book would be facilitated by collegial discussion and feedback; this is obviously easier when all the contributors work in the same place.

Part I includes this general introduction and a chapter in which we review various approaches that have been used to come to an understanding of what is meant by clinical competence. Several distinct perspectives are presented: the "armchair" approaches to definition of competence illustrated by textbooks on clinical diagnosis, task analysis, and other formalized methods used by specialty boards and licensing examinations, and the indirect viewpoint on competence provided by a variety of research methods—epidemiology, psychology, and quality of care research.

Part II, Methods, identifies and describes criteria to be applied to any method with purports to measure clinical competence. Each chapter then focuses on one instrument (or class of instrument). Included is a synthesis of the strengths and weaknesses of the tool under discussion and suggestions for further research.

The review of individual measures is followed in Part III, Applications, by consideration of how the measures can be used. Three chapters are devoted to looking at the application of tools to special components of professional performance where it is likely that more than a single method is appropriate. These are the use of diagnostic tests, the assessment of technical proficiency, and the measurement of physician-patient interaction.

The final section, Part IV, Implications, takes the form of a synthesis, under four headings: education, research, certification, and health care. Each chapter attempts to provide a "where do we go from here" statement, based on the perspectives presented in the earlier sections of the book.
References


Defining Competence: A Methodological Review

GEOFFREY R. NORMAN

What is clinical competence? The present chapter reviews the various methods used to define competence and points out the strengths and limitations of each. The focus is the clinical encounter—the activities of the physician in diagnosing and managing a patient's problems. Although other aspects of physician activity such as attitudes or relationships with colleagues and other health professionals, participation in professional associations, and scholarly or research activities may be important components of professional activity, they are excluded from the present discussion.

The Dimensions of Competence

Senior (1976) has made a distinction between "competence" and "performance," the former meaning what a physician is capable of doing, and the latter meaning what a physician actually does in his day-to-day practice. The distinction parallels the separation between the efficacy and effectiveness of a drug—the former is an indicator of the performance of a drug under ideal circumstances, with carefully chosen patients and...
near-perfect compliance, and the latter refers to the performance of the drug under actual conditions, with less than perfect compliance and a more heterogeneous patient population.

Competence in a specialty can be defined along a number of dimensions. For example, the American Board of Internal Medicine (1979) outlines four dimensions which must be considered in defining competence: (1) the relevant abilities of the physician—knowledge, technical skills, and interpersonal skills, (2) problem-solving tasks—data gathering, diagnosis, continuing care, etc., (3) the nature of the medical illness—the problems encountered by the physician, and (4) the social and psychological aspects of the patient’s problem, especially those which relate to diagnosis and management.

Other specialties and certifying bodies have used variations on a matrix approach to the definition of competence (Burg et al., 1976; McDermott et al., 1977). The American Board of Medical Specialties (1979) suggests a conceptual framework with three principal dimensions: (1) the tasks of the physician (history taking, use of tests, etc.), (2) the clinical situations or subject matter of the discipline, and (3) prerequisite abilities (knowledge, problem-solving ability, attitudes, etc.). Although these schema suggest that several dimensions must be considered in a definition of competence, many approaches to a detailed specification of competence in a specialty deal only with one or two dimensions.

Methods of Defining Physician Activity:
An Overview

There are two main sources used in defining the activity of the effective clinician—the writings of individual experts, which commonly appear in the introductory chapters of textbooks on clinical diagnosis, and the consensus of groups of experts (American Board of Medical Specialties, 1979) which have been compiled as detailed and structured taxonomies and published by specialty boards. There is also, however, an extensive research literature in disciplines ranging from psychology to epidemiology relevant to the definition of competence. The primary goal of this research may not have been to define competence, but some of it has a direct bearing on describing relevant dimensions of competence: the thinking processes of clinicians, the manifestations of these processes in observed behaviors, the nature of problems confronting the clinician, and the nature of the interaction between the
physician's activities and the patient's problem resulting in a change in health status.

A further breakdown of these methods is defined below. In the remainder of the chapter we shall review each method in greater detail.

1. **Reflective/philosophical.** Historically the earliest definition of the profession, this "method" involves consideration of the qualities deemed important by a single recognized authority, and generally appears in the context of philosophical writings or textbooks on clinical diagnosis.

2. **Task analysis.** Andrew (1976) has defined task analysis as "a process whereby the activities that physicians engage in...are documented and described in such a way as to make explicit the purposes of the activity, the procedures that the physician utilizes in performing the activity, and the outcomes or expected products of the activity." She identifies three streams in task analysis: the consensus of expert opinions; the critical incident approach; and the use of activity logs.

3. **Descriptive studies.** Observational and epidemiological studies in primary care and some specialties focus on the type of problems seen by the physician. This method is fundamental in defining the subject matter or clinical problems seen by the physician.

4. **Studies of diagnostic thinking.** Medical diagnosis has been studied by psychologists as an example of "problem-solving" activity. Several research paradigms have been used in investigating clinical reasoning, including information-processing studies, in which physicians are observed and questioned in the course of solving a clinical problem; clinical judgment studies, in which physician judgments are described by mathematical models; and decision analysis, using Bayes' theorem in conjunction with the probability of various outcomes and their expected value in developing an optimal approach to diagnosis or management of a particular problem.

5. **Consumer opinion.** These studies of patient attitudes and expectations lead to a definition of competence from the client's perspective.

6. **Epidemiologic and quality of care.** With these approaches competence is defined from a health care perspective,
Ideally in terms of those aspects of performance which can be shown to result in improvement or deterioration of health status.

A detailed review of each approach follows.

**Reflective/Philosophical Statements**

From the time of the Greek philosophers, the leading physicians of the day have put forth personal statements bearing on the principles and practice of medicine. Plato states, “Now the most skillful physicians are those who, from their youth upwards, have combined with a knowledge of their art, the greatest experience of disease; they had better not be of robust health and should have had all manner of disease in their own person.” Hippocrates is cited as saying that the physician should have a good working knowledge of astrology! Osler, in an address to the graduating class at the University of Pennsylvania in 1889, defines three primary qualities of the clinician: the “Art of Detachment,” which refers not to scientific objectivity, but to personal detachment from the pursuits and pleasures of youth; “The Virtue of Method,” scientific method and practice management; and the “Quality of Thoroughness,” not in history-taking and physical examination, but in scientific knowledge. More recent treatises on the practice of medicine cannot claim the literacy or asceticism of Osler’s admonitions but are otherwise similar. For example, Tyrer and Eadie (1976) have described one aspect of clinical reasoning in the following manner:

As more and more evidence of various sorts is accumulated, the physician, by reasoning in relation to his knowledge, is able to formulate possible diagnoses in his mind and then narrow down the list of suspect diseases until a definite diagnosis can be made. Should insufficient or invalid evidence of facts be obtained . . . a correct diagnosis is unlikely. . . . Should the knowledge of the clinician be defective . . . it would be difficult for him to reach a correct diagnosis . . . or should his reasoning be defective he may be unable to isolate the correct diagnosis from the forest of alternatives presented to him. . . .

Tumulty (1973) defines the skills of the clinician in the following manner:

An effective clinician must have a number of skills. He must be a scientist. He must be knowledgeable about the natural course of com-
mon and uncommon clinical disease. He must be able to harvest clinical evidence from all available sources. He must be a keen analyst of these gathered facts. ... The clinician must have the facility to communicate with [the patient] and his family members.

These descriptions of the effective clinician are not confined to textbooks on clinical medicine; similar statements can be found in medical journals (Dudley, 1970; King, 1967; McWhinney, 1972). Some of these writings have had a major impact on medical education and clinical practice and, although philosophical and possibly idiosyncratic, cannot be lightly dismissed.

Critique. Reconciling these definitions with the more systematic descriptions derived from the formal or research approaches is difficult. These writings have some bearing on both prerequisite abilities and problem-solving activities. Yet each treatise, considered alone, would be a poor prototype on which to develop an evaluation method. Although common threads can be developed, the writings are a reflection of the viewpoint of a single individual. There may be agreement among practitioners that the writings are a reasonable representation of clinical practice, and presumably the extent to which an individual author can capture the attitudes of his peers is related to the popularity of his writings. But each description is, of itself, incomplete and lacking in content validity and often presents a utopian view of how medicine should be practiced.

Task Analysis Approaches

In contrast to the philosophical stance of the reflective approaches, task analysis is rooted in a much more pragmatic base. The three streams identified by Andrew (1976) vary in the extent to which they use actual incidents of behavior or subjective judgments but the goal is the same—to describe the components of competence required by an adequate but not necessarily exemplary practitioner of the specialty. The description is deliberately comprehensive, and it attempts to deal with all four dimensions of competence—abilities, processes, problems, and interactions with the patient.

The Consensus of Experts. Experts in a field are frequently used in defining the content of a discipline. They may do so explicitly as members of a committee whose goals are to formulate educational objectives and define the components of competence of a
specialty, or implicitly as members of test committees designing examinations for licensure or certification. Explicit competency statements have been developed for a number of specialties, such as the American Board of Pediatrics (1974), The American Board of Internal Medicine (1979), and the College of Family Physicians of Canada (1973). The process used by these various specialty boards is similar. Initially, a committee of acknowledged experts in the field is designated; the group may vary in size from fewer than 10 to 50 or more members. The committee or subcommittees then develop detailed specifications of competence along predetermined dimensions (for example, subject matter, abilities, tasks). These statements are then collated, reviewed, and discussed until a consensus is reached.

Andrew (1976) has documented a number of potential biases in this approach. The first problem is one of sampling: If the group of experts is too small, or if they do not represent a sufficiently broad perspective, consensus may be difficult to achieve and may have little relationship to the actual practice of the specialty. A second bias is that competency statements should derive from the independent judgments of the experts, in order to capitalize on the diversity of interests represented in the committee. After initial generation of competency statements, however, dominant members of the committee may alter the final consensus. Another problem is simply a reflection of the magnitude of the task. It is a mammoth undertaking, requiring hundreds of hours of professional time, and may deter many specialties from developing such statements. A last problem is that the definition which emerges is a statement of the ideal. The extent to which the statement of competence reflects how the specialty should be practiced, rather than the actual practice of the specialty, is unknown.

Critical Incident Approaches. The "critical incident" approach (Flanagan, 1954) overcomes some of the defects present in the previous method. A large number of practitioners are asked to identify, from their own experience, incidents which appear to have either a positive or a negative effect on the quality of care delivered to the patient. The method was used by the National Board of Medical Examiners (Flanagan, 1960) to define clinical competence for their examination. The study involved 600 practitioners and a total of 3,300 critical incidents, described as "clinical situations in which they had personally observed interns
doing something that impressed them as examples of conspicuously good or poor clinical practice." These incidents were then grouped and classified into nine major categories.

The technique was also employed to develop a certification examination for the American Board of Orthopedic Surgery (Miller, 1968), this time involving 1,700 critical incidents contributed by 1,100 specialists, and grouped into 94 subcategories. Sanazaro and Williamson (1970) used the technique to study a number of primary care specialties involving a total of 2,000 physicians and 12,000 incidents. Cowles (1965), in a variation of the method, classified a total of 2,300 critical comments retrospectively derived from written evaluations of clinical clerks and scored each comment in an attempt to develop an objective means of evaluating verbal statements.

These studies and the resulting examinations have had a major impact on the development of clinical evaluation methods. The National Board study continues to be used as a guideline for developing new evaluation methods, and the orthopedic examination format has been copied, with modifications, by other specialties, notably the College of Family Physicians of Canada (Lamont, 1972). The method overcomes many of the problems of the expert committee approach. The criterion group is very large and is deliberately chosen to be representative. Incidents are not deleted in the final summary, but are simply categorized, removing personal biases from this stage of the process.

What problems exist with the method? The first is inherent in the method: By definition, there is no attempt to determine what constitutes adequate practice, only very good and very poor practice. As a result, a standard of acceptable performance cannot be interpolated easily and one is left with judgments about how many good incidents are good enough and how many poor incidents are too many. The second problem is one of validity. Although, at first glance, one might dismiss the possibility of bias because of the large samples involved, the potential for bias still exists. The method is based on actual performance, but the performance is judged through the eyes of the criterion physician. As one example, thoroughness emerges as a principal feature of good intern performance in the National Board study, yet thoroughness on history and physical has been shown to be unrelated to diagnosis and management, and thoroughness of laboratory workup at hospital admission has been shown to be inversely related to quality of care (Barrows et al., 1978; Sackett, 1978).
This method, then, has played a historically significant role in defining clinical competence and has several methodological advantages. Nevertheless, a potential for different types of bias remains.

Procedural Logs. To avoid the biases inherent in the previous methods, a more direct definition of competence must derive from the actual activities of competent physicians. One approach has been to gather data on day-to-day performance through the use of activity logs, in which physicians maintain a detailed record of their daily activities over a period of time. One difficulty with this method is a logistical one of ensuring accurate recording by physicians as the events occur, rather than retrospectively. Problems also exist in developing an adequate and representative sample, and avoiding volunteer bias. Finally, in contrast to the critical incident method, there is no way of determining from the log which activities constitute good or competent behavior, which are poor, and which activities fall in the mid-range.

Critique. To the extent that there is a conventional approach to the definition of competence, the three methods summarized in this section epitomize this approach. Most discussions about defining competence focus on these methods. We have described the strengths and weaknesses of each method in the preceding discussion, but there are inherent problems common to all three methods. One methodological problem is sampling. The sampling must meet two possibly conflicting criteria; on the one hand, it is necessary to sample at random a large number of practitioners to ensure that the resulting definition is representative; by contrast, the individuals whose opinions are to culminate in the definition of competence must be accepted exemplars in the field and therefore not randomly chosen. The second criticism of these methods is that, like the philosophical statements, they are derived, not from actual samples of clinical practice, but from aspects of practice which are, in one way or another, perceived as important by the practitioners. The extent to which the competencies described in the final document are a reflection of practice is contingent on the judgment of the practitioners, and such judgments are liable to a number of biases which may affect the validity of these products. Finally, as Lloyd (1979) has indicated, the most serious limitation of such approaches is that they have a tendency to become ends in themselves, rather than means to ends.
Descriptive Studies: 
Observational and Epidemiological

All of the task analysis approaches in the previous section have the explicit goal of a comprehensive definition of competence. In contrast to this formalized process, there is an extensive literature, primarily in family medicine and general practice, which takes a more pragmatic view of clinical competence, defining the entity in terms of the daily activities of competent practitioners. These studies have three main foci: (1) detailed observation of a small sample of physicians over an interval of several days, (2) epidemiological studies of the types of problems seen by physicians, classified by complaint and/or diagnosis, and (3) studies of defined activities, such as history taking and physical examination.

Observational Studies. Two intensive studies using similar methodology have been conducted in this area, by Peterson et al. (1956) in North Carolina and by Clute (1963) in Canada. Both studies used ethnographic methods, in which the investigators visited physicians in their practices, held discussions with them, and observed their activities over a period of several days.

The studies presented a wealth of data describing many aspects of general practice, ranging from the use of cytologic smears and tuberculin tests to the occupation of the physician's father. The studies can be criticized as objective descriptions of clinical competence in general practice, however, since performance in history, clinical examination, and use of investigations was judged against an external criterion. For example, Peterson states, "Almost no good histories dealing with anemia or hypertension were observed," suggesting that the histories were being judged against criteria which may have been inappropriate for family practice. Peterson and colleagues had accepted that "most, if not all, physicians performed complete physical examinations on all new patients... That such was not the case soon became evident." In light of what we now know about the limited value of screening tests, the general practitioners may have been practicing more rational medicine that was credited to them.

Clute's study involved 44 general practitioners in Ontario and 42 in Nova Scotia. Again, the data gathered from the study was comprehensive and included previous education, office facilities, work load, membership in professional societies, income, and social activities, in addition to a detailed breakdown of the kinds of problems seen and quality of care delivered. Once again,
however, a large portion of the book is devoted to assessing the quality of care delivered, and although the authors devote considerable space to a rationale for their scoring methods, they might still be interpreted as arbitrary.

Price et al. (1971) conducted a series of studies of physicians in Utah over two decades. A 1964 study involved the assessment of 80 criterion measures on each of 800 physicians ranging from grade point average in the first two years of medical school to the dollar value of office equipment. The 80 variables were then used to provide some empirical basis for defining the domains of competence. The investigators developed a variety of methods to combine and weight these variables into an overall measure of competence. Of particular interest to the present chapter, those variables which were judged most important to competence had no direct relationship to patient care but were based on peer judgments—for example, the number of times a physician was nominated as an outstanding contributor by colleagues. These investigators went a step further in defining competence in a second study. Critical comments about physician behavior were obtained from 100 physicians used in the first study. These were then rated by ten different consumer and professional groups, ranging from nurses and medical students to hippies. Although these comments do not provide a detailed specification of competence on which an examination could be structured, they do provide insight into domains considered important by providers and consumers.

**Studies of Patient Problems.** A ubiquitous feature of general practice research is a preoccupation with defining the types of problems seen by the practitioners. Limited studies have been conducted in Canada (McFarlane, Norman, and Spitter, 1971; McFarlane, O'Connell, and Hay, 1971) and Britain (Fry, 1966; Hodgkin, 1973) and two major studies have been done in the United States (National Center for Health Statistics, 1978; Marsland et al., 1976). More similarities than differences have emerged from these studies, conducted in widely different geographic locations. The most common problems dealt with by general practitioners are acute upper respiratory infections, followed by cardiovascular disease and emotional problems. The prevalence of emotional problems shows greatest variation from study to study, presumably more as a function of the physicians involved than their patients.

The implication of these studies for educational programs directed at training primary care physicians is significant. The
family physician is competent to the extent that he can manage the problems he is likely to encounter: management of emotional problems, the problems of detection and compliance in the 10 percent of his adult practice who are hypertensive, and so forth. The futility of attempting to train such a physician in the academic medical center is elegantly demonstrated by White (1961).

*Studies of Practice Activities.* Other studies in general practice have focused on other aspects of the patient encounter—for example, the proportion of patients with new or old conditions who received only history, history and limited physical examination, investigations, and so on (Hull, 1972; Morrell, 1972). A complete, or even extensive physical examination is a relative rarity in general practice, occurring in less than 10 percent of visits.

*Critique.* The three classes of research we have reviewed in this section do not, of themselves, define clinical competence. They do, however, yield insights into the performance of physicians. As more studies are initiated in a variety of domains, it is increasingly evident that physicians frequently do not practice medicine in the manner they have been taught in medical school, and their actual performance may not bear a very close resemblance to how they believe they practice, or think they ought to practice. Thus, there may be considerable distortion when competence is defined “from the armchair” using the methods described previously.

The study of presenting problems is perhaps the best approach to a definition of the clinical situations encountered by the physician, which in turn defines the relevant subject matter. However, the transfer from a description of practice activity to a definition of competence is not completely straightforward. The criteria imposed on practitioners by Clute and Peterson comprised a definition of competence which in retrospect appears to have been arbitrarily high. Yet the actual performance of physicians cannot be presumed to define the lower limits of competence in the absence of some external criteria. In addition, although prevalence studies of presenting complaints and diagnoses assess one dimension of the subject matter of a discipline, they do not directly translate to a definition of the appropriate content areas. The physician must be competent to deal with rare but potentially serious conditions and make the appropriate decisions to treat, consult, or refer. These situations are not adequately addressed by simple considerations of disease prevalence.
family physician is competent to the extent that he can manage the problems he is likely to encounter: management of emotional problems, the problems of detection and compliance in the 10 percent of his adult practice who are hypertensive, and so forth. The futility of attempting to train such a physician in the academic medical center is elegantly demonstrated by White (1961).

Studies of Practice Activities. Other studies in general practice have focused on other aspects of the patient encounter—for example, the proportion of patients with new or old conditions who received only history, history and limited physical examination, investigations, and so on (Hull, 1972; Morrell, 1972). A complete, or even extensive physical examination is a relative rarity in general practice, occurring in less than 10 percent of visits.

Critique. The three classes of research we have reviewed in this section do not, of themselves, define clinical competence. They do, however, yield insights into the performance of physicians. As more studies are initiated in a variety of domains, it is increasingly evident that physicians frequently do not practice medicine in the manner they have been taught in medical school, and their actual performance may not bear a very close resemblance to how they believe they practice, or think they ought to practice. Thus, there may be considerable distortion when competence is defined “from the armchair” using the methods described previously.

The study of presenting problems is perhaps the best approach to a definition of the clinical situations encountered by the physician, which in turn defines the relevant subject matter. However, the transfer from a description of practice activity to a definition of competence is not completely straightforward. The criteria imposed on practitioners by Clute and Peterson comprised a definition of competence which in retrospect appears to have been arbitrarily high. Yet the actual performance of physicians cannot be presumed to define the lower limits of competence in the absence of some external criteria. In addition, although prevalence studies of presenting complaints and diagnoses assess one dimension of the subject matter of a discipline, they do not directly translate to a definition of the appropriate content areas. The physician must be competent to deal with rare but potentially serious conditions and make the appropriate decisions to treat, consult, or refer. These situations are not adequately addressed by simple considerations of disease prevalence.
Studies of Diagnostic Thinking

Medical diagnosis has been of interest to psychologists for some time as an example of human problem-solving activity. Clinicians have been studied using a variety of experimental paradigms, and, in turn, this research has affected our view of the nature of clinical reasoning, eliminating some of the mystique, but frequently substituting a plethora of simple models lacking in credibility and seemingly unrelated to each other.

The studies reviewed in this section do not presume to describe all aspects of clinical competence; their focus is on the mental strategies used by clinicians to seek out information, generate diagnostic hypotheses, weigh the information against these hypotheses to arrive at a conclusion, and develop a management plan with due consideration of the costs and benefits associated with alternative therapies. Excluded from the discussion are the domains of technical skills, interpersonal skills, and prerequisite knowledge. Instead, the focus is on understanding what many would view as the essential feature of clinical competence—the mental problem-solving strategies involved in using the available data to arrive at a diagnosis and management plan.

The studies fall into three general classes (Elstein and Bordage, 1979) characterized by the goal of the research and the experimental methods employed. We shall focus on two in this review.

1. Information-processing approaches. The goal of this research is to understand how humans solve problems. The method involves detailed observation of human behavior and of verbalized thought processes, while the subjects approach diagnostic problems. The goal is a theory which describes both the process of problem solving and the relationship between process and outcome.

2. Decision analysis and Bayesian approaches. This research approach is further removed from the study of human thought. The goal is to seek an optimal means of combining the data to arrive at a correct or best outcome using information about the probability and utility of various possible end states.

Information Processing Approaches. Studies of clinicians using this paradigm have "the goal . . . to define precisely the processes and states that a particular subject is using to solve a particular problem, and to be able to list—for example, in the form of a
computer program—the exact sequence of operations used” (Kleinmuntz, 1968). Several studies have been conducted in medicine, and all use similar experimental techniques. The subject physician is presented with a clinical problem, either as a simulated patient or orally by the investigator, and is asked to approach the problem in his usual style. His thought processes are captured, either by thinking aloud as he works through the problem, or by reviewing a videotape subsequently. Subjects have included neurologists (Kleinmuntz, 1968), criterion internists (Elstein et al., 1978), neurology residents (Barrows and Bennett, 1972), family physicians (Barrows et al., 1978; Feightner et al., 1977), medical students (Neufeld et al., 1981), and residents in family and internal medicine (Scherger, 1980).

A common finding of these studies is the formation of multiple diagnostic hypotheses very early in the encounter, frequently within a few seconds of obtaining the chief complaint. These hypotheses serve an essential role in determining both the process and the outcome of the encounter. Elstein et al. (1978) and Barrows et al. (1978) found that the hypotheses strongly dictated the subsequent search for information. Most of the data were gathered to confirm one hypothesis or another. Furthermore, using a simple three-point weighting of each finding against each hypothesis, the investigators in both studies were able to accurately predict the clinician’s ultimate choice of diagnosis, suggesting that the clinical reasoning involved in arriving at a diagnosis may not be as complex as we had supposed. Barrows et al. (1978) found that the number of hypotheses and many other measures of the process, in particular the amount of significant data gathered, were unrelated to diagnosis and management. What was important in diagnosis was whether the clinician had entertained the correct diagnosis as an hypothesis during the encounter. These investigators also examined differences by specialty type (family medicine or internal medicine) and educational level, from entry into medical school to graduation. With the exception of the content of the hypotheses, most features of the encounter, such as number of hypotheses or amount of data, were essentially the same for all subgroups. These results suggest that the process of problem solving may be a general feature, and the important consequence of education and/or experience is the acquisition of content.

Although the studies can be faulted on methodological grounds, since they frequently involve a small sample of subjects, and only a few clinical problems, they are based on actual perfor-
mance. Furthermore, they have begun to explore the mental problem-solving strategies of clinicians, an area of competence which formerly could only be inferred from the introspective, and frequently erroneous, accounts of experienced clinicians.

What are the implications of these studies for a definition of competence? The findings may be viewed by some as disquieting. The notion of the clinician as the passive and objective observer of illness finds little confirmation in these studies. Similarly, the artful balancing of positive and negative information to arrive at a diagnosis appears much simpler than has been formerly construed.

**Bayesian Diagnosis and Decision Analysis.** The methods discussed above have an obvious role to play in defining the nature of clinical reasoning and clinical competence as actually practiced by clinicians. But also relevant is an extensive literature based on the manipulations by computers of disease probabilities, conditional probabilities of various signs and symptoms, and the value of various treatment outcomes. The literature has roots in business and psychology and uses various labels—computer-aided diagnosis, Bayesian diagnosis, and decision analysis (McNeil, 1975). It is not intended to understand the clinician's thinking processes. Rather, the method uses, as input, information derived from actual disease probabilities, and its goal is to determine an optimal diagnosis or management decision based on the available data (optimal in the sense of maximizing the probability of particular diagnostic end states, or the utility, to patient or physician, of particular treatment outcomes).

This literature is viewed with suspicion by many clinicians for a number of reasons (Elstein, 1976; Feinstein, 1979); it is highly mathematical and not readily understandable, and the computer models attempt to treat all patients with a given condition by a single model, the antithesis of individualized clinical care.

Although how these methods can be applied to a definition of clinical competence is unclear, the findings of a number of studies in the area are a disquieting reminder of human fallibility in clinical reasoning. Computer diagnosis has been shown to be, at worst, as accurate as physician diagnosis, and at best, when actual probability data are available, considerably better than human diagnosis (Beach, 1975; Leaper et al., 1972). Further, as the literature on decision analysis grows, the method is evolving as an ideal criterion against which actual physician performance can be judged.
Critique. Will the application of these methods lead to the replacement of the clinician's reasoning with the numerical manipulations of a computer? Despite the fallibility of human judgment, we feel such an outcome is unlikely. Methods such as decision analysis may, however, have an educational role to play in improving clinical judgments through explicating the range of available alternatives and the value of possible end points (Elstein, 1976). The apparently contradictory conclusions of different clinicians in the same situation may then be shown to have a rational basis in differing subjective probabilities or different values placed on a range of possible outcomes.

Similar comments might be made about the clinical judgment and information-processing approach, which will not, of itself, lead to a definition of competence. What it has accomplished is the illumination of a central area of competence—problem solving or clinical reasoning—which was previously considered unobservable and which lay primarily in the realm of the art of medicine. Although specific measures of these areas of competence do not exist in any practical form as yet, the insights provided by the research could potentially result in the development of such measures in the near future.

Patient and Consumer Opinion

It is the nature of professional practice that the client is not in a position to judge the competence of the individual practitioner, and decisions about competence are made by legislative bodies such as state boards and medical associations. The patient indeed has no valid means at his disposal to determine the technical skills of a surgeon or the diagnostic acumen of an internist. Yet there is one area in which the patient is informed and is the best judge of competence—the domain of doctor-patient relationship and interpersonal skills.

Korsch and co-workers (1968) studied the relationship between the physician's interviewing techniques and patient satisfaction using 800 tape-recorded encounters with pediatric patients and follow-up interviews in an emergency room setting. Significant relationships were found between satisfaction and (1) the friendliness and sympathy of the physician; (2) the physician's communication skills; (3) physician explanations of the nature and cause of the illness; and (4) his or her dealing directly with the patient's concerns and anxieties. Dissatisfaction was related to unfulfilled patient expectations for lab tests, X-rays, physical examinations, or drugs.
Hulka (1971, 1975) has developed a methodology to directly assess physician awareness of patient concerns using a problem-specific scale completed by both physician and patient. Physician awareness has been related to a number of physician and practice characteristics, such as specialty certification and residency training; however, the relationships are complex and not consistent, nor is there a consistent relationship between physician awareness and health outcomes.

Critique. Like the research on problem solving, these studies do not provide a comprehensive framework for a definition of competence; rather, they serve to elucidate one area which previously resided in the domain of the “art” of medicine. At the present time, it is unclear precisely how patient satisfaction is influenced by the expectations of the patient or other patient variables, and how the dimensions of physician performance related to patient satisfaction can be characterized. Nevertheless, it is apparent with the rise of consumer advocacy in North America that increasing attention must be paid to this domain of competence.

Epidemiologic and Quality-of-Care Approaches

Although the research approaches we have reviewed in this chapter are focused on the performance dimension, these approaches tend to be of a descriptive nature. By contrast, research on quality of care attempts to measure physician performance on specific conditions against an absolute standard of acceptable care. This literature has been reviewed by Williamson (1976) and others.

A feature of all research conducted in this tradition is a focus on actual performance—the actual quality of care delivered to a series of patients with one or more conditions in a hospital or clinic setting. Quality of care may be measured against standards set by some external body, similar to the criteria used in the Clute (1963) and Peterson (1956) studies cited earlier, so-called process measures, or may be based on measures of patient morbidity or mortality resulting from the care received—referred to as “outcome measures.” These outcomes are related to various measures of process in the health care system to indicate which features of the process of health care are significantly related to improved outcomes.

Some examples will illustrate these distinctions. A study by Moses and Mosteller (1968) examined the postoperative death rates from cholecystectomy and gastric resection in various hospi-
tals. After correcting for a variety of patient variables, the death rate in teaching hospitals was three to five times lower than in nonteaching hospitals. Similar results were found by Lipworth, Lee, and Morris (1963) in an examination of case fatality rate for a variety of conditions in teaching and nonteaching hospitals. Morbidity associated with appendicitis (peritonitis or abscess) was used in a study of Goran and Gonnella (1975) of various sources of health payments.

Process and outcome measures were both used in a study by Brook and Appel (1973) of 296 patients in a single hospital. They also used implicit judgments by faculty of quality of care in comparison with explicit criteria developed by the same panel. In terms of assessment of clinical competence, a major concern emerging from these studies is that frequently very little relationship was found between performance and such conventional measures of clinical competence as board certification or years in practice. Furthermore, the study of Goran et al. (1973) directly compared performance on management of urinary tract infection with an evaluation of competence using patient management problems and found that clinic teams did very well on the written examinations and very poorly on the actual clinic management of similar problems.

Critique. Studies of actual performance have a clear role to play in defining clinical competence, if for no other reason than that the great discrepancy between these measures and conventional evaluation methods strongly suggests that they assess a different, and undoubtedly important, domain of competence. Nevertheless, these methods have certain disadvantages: They are extremely expensive, and they are usually constrained to examination of a particular problem area. As such, they cannot be used to contribute to a broad definition of competence.

Summary and Conclusions

No single method can adequately define the prerequisite knowledge, skills, and attitudes required of a competent physician in a particular specialty. Each method has certain limitations. These may derive from biases inherent in the method, such as the critical incident studies or consensual approach, or may be a result of the limited focus of the method, such as the information-processing studies.
Regardless of the method chosen, the goal is the detailed specification of the necessary knowledge or skills of the competent practitioner. These individual components are usually arranged or sorted into domains of competence, such as knowledge, technical skills, or interpersonal skills. The detailed specification of the individual elements within each domain is beyond the scope of this book and is a task we leave to the members of objectives committees and specialty boards. The domains of competence are of immediate concern, however, since each evaluation method to be considered in subsequent chapters may be useful for assessing only some of the dimensions of competence. We must begin by outlining these areas in order to critically examine the appropriate role of each evaluation method.

The scheme we will use for the purposes of organizing this book is shown in the Table 2.1. It is no more or less optimal than

| TABLE 2.1 |
| A Categorization of Clinical Competence |

| The following abilities are required in encounters between a physician and individual patients: |

1. **Clinical Skills**
   - The ability to acquire clinical information by talking with and examining patients, and interpreting the significance of the information obtained.

2. **Knowledge and Understanding**
   - The ability to remember relevant knowledge about clinical conditions in order to provide effective and efficient care for patients.

3. **Interpersonal Attributes**
   - The expression of those aspects of a physician's personal and professional character that are observable in interactions with patients.

4. **Problem Solving and Clinical Judgment**
   - The application of relevant knowledge, clinical skills, and interpersonal attributes to the diagnosis, investigation and management of the clinical problems of a given patient.

5. **Technical Skills**
   - The ability to use special procedures and techniques in the investigation and management of patients.
any other categorization reviewed in this chapter and serves, not to define competence, but to provide a reference point for the discussions of specific methods which follow. The scheme will be introduced again in the discussion of educational implications in Chapter 16, at which time we shall suggest the potential role of each of the evaluation methods we have reviewed in assessing competence.

REFERENCES


