In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Patricia Cutshall

Department of Adult Education

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date 30 June 1978
Abstract

Health sciences literature from 1970 - 1977 which pertains to monitoring and maintaining competence of health professionals is abstracted and reported. Recognizing that competence is defined by the criteria and standards used to evaluate its presence, this thesis uses Donabedian's evaluation model to organize the bulk of the literature reviewed. While the model applies to competence of the health care system, this paper limits discussion to competence of health professionals. It considers literature concerned with definition, evaluation, and restoration/maintaince action related to knowledge and skills possessed by practitioners (structure), clinical performance (process), and results of care (outcome). Additionally it provides information about the historical development of and current trends in credentialing mechanisms intended to certify competence, explores in some detail the issue of mandatory continuing education, and discusses contemporary social and political phenomena which influence and are influenced by competence monitoring activities. The concluding chapter offers observations on the current state-of-the art with respect to evaluating and maintaining competence and makes recommendations for further research, development and implementation of competence monitoring activities consistent with the North American social context. Finally, several implications for continuing educators' roles and responsibilities vis a vis monitoring and maintaining health professionals' competence are discussed.
# Table of Contents

## I Introduction
- Purpose .................................................. 2
- Procedure .................................................. 2
- Definition of terms ........................................ 3
- Organization of the review ................................. 4

## II Assessing Potential for Clinical Performance: Structure ..... 7
- Self-Assessment Procedures ................................. 7
- Randomly Available Test Instruments ...................... 8
- Self-Assessment Procedures for Select Groups .......... 9
- Formal Testing of Practitioners ............................ 12
- Tests for Full Range of Practice .......................... 13
- Testing in Specialty Areas ................................. 16
- Special Purpose Assessment Procedures ................. 17
- Non-traditional Testing Procedures ....................... 17
- Individualized Physician Profile ........................ 18
- Tests of Clinical Decision Making ....................... 18
- Tests of Interactive Skill ................................ 21
- Summary and Comment ...................................... 22

## III Assessing Clinical Performance: Process .................... 26
- Systematic Assessment of Total Performance ............. 26
- General Discussions of Systematic Performance Evaluation ... 27
| Performance Evaluation Using Implicit Standards | 28 |
| Performance Evaluation With Explicit but Imprecise Standards | 33 |
| Performance Evaluation Using Explicit Standards | 37 |
| Systematic Assessment of Situation-Specific Performance | 40 |
| Situation-Specific Process Evaluation: Individual Practice | 42 |
| Situation-Specific Process Evaluation: Agency Settings | 45 |
| Situation-Specific Process Evaluation: Regional Systems | 50 |
| Unplanned Evaluation of Performance | 59 |
| The Place of Work | 60 |
| The Profession | 61 |
| The Licensing Authority | 63 |
| Summary and Comment | 68 |

| IV Assessing Results of Clinical Performance: Outcome | 72 |
| General Measures of Health Status | 72 |
| Situation-Specific Outcomes Evaluation | 74 |
| Summary and Comment | 80 |

| V Assessing Combinations of Structure, Process and Outcome | 82 |
| Concurrent Evaluation of Structure-Process-Outcome Combinations | 82 |
| Concurrent Evaluation of Structure, Process, Outcome | 83 |
| Structure-Process Combinations | 85 |
| Agency Based Process-Outcome Evaluation | 88 |
List of Tables

I Types of Disciplinary Action Taken with American Physicians 1963 - 1967 ............................................. 65
II Causes for Disciplinary Action Taken with American Physicians 1963 - 1967 ............................................. 65
III Advantages and Disadvantages of Measurement of Structure ...... 115
IV Advantages and Disadvantages of Measurement of Process .......... 115
V Advantages and Disadvantages of Measurement of Outcome ........ 116
VI Evidence of Desired Change in Reported Evaluation-Action Projects .................................................. 131
VII Arguments in Favor of Mandatory Continuing Education .......... 171
VIII Arguments in Opposition to Mandatory Continuing Education ..... 172

List of Figures

1 Interaction Process Analysis (Robert F. Bales) ...................... 23
2 Change Agent Behaviors Which Contribute to Evaluation-Action Programs ........................................................ 188
INTRODUCTION

Cost, availability, and quality of health care are currently among the most contentious and problematic issues of the seventies in North American society. "Solutions" such as consumer control of the delivery of those services and continuing education in health care for both the consumers and providers are subjects of continuing and vigorous debate. Both problems and solutions have increasingly been translated into very real events.

Most notable of these are several recent legislative activities of far reaching significance: sweeping changes in the regulation of health professionals as has occurred with the Professions Code in Quebec (400) and with the Health Disciplines Act in Ontario (229); the establishment of Professional Standards Review Organizations in the United States as an amendment to the Social Security Act (515,516,518); the enactment in several states of "mandatory continuing education" laws; and, paradoxically, a slow down in legislation establishing new categories of licensed health personnel, this last recommended by the U.S. Department of Health, Education and Welfare (513). Within the private sector, popular literature reflects a growing conviction that responsibility for health and health care cannot be left solely to health professionals (82,250).

Reactions of health care providers to all this activity are vigorous. The health sciences literature of the seventies is notable for
the volume of attention given problems connected with ensuring that health care is acceptable in terms of cost, availability and quality. With respect to quality, the competence of health care practitioners is one of the more significant determinants and is the subject of this paper.

**Purpose**

The purpose of this paper is to identify, through a literature review, recent developments in the health care field which are related to strategies for assessing and maintaining competence of health professionals.

**Procedure**

Literature from the health field was surveyed with special attention to items from medicine, dentistry, nursing and pharmacy. The scope of the search was limited to literature of American and Canadian origin published from 1970 to June 1977 and to concern with competence of the practicing professional (as opposed to the professional at entrance to practice). A Medline search and review of periodicals for the 1970 - 73 period on the same subject by Neylan (362) was utilized as a base and was extended to include research abstracts and government documents. Periodical literature, research abstracts and government documents from the 1974 - 77 period were reviewed. Chief sources for data retrieval
were Dissertation Abstracts, ERIC, Medline and PSRO Information Clearing House, operated by Capital Systems Group, Inc. under a contract with the United States Department of Health, Education and Welfare. This survey should be considered a representative sample of the literature available, not an exhaustive search.

Definition of Terms

Accreditation - the process by which an agency or organization evaluates and recognizes a program of study or an institution as meeting certain predetermined qualifications or standards.

Audit - a systematic approach to evaluation of care which results in documentation by health professionals of how far their care conforms to their own standards of adequacy or excellence.

Certification - the process by which a non-governmental agency or association grants recognition to an individual who has met certain pre-determined qualifications specified by that agency or association.

Competence - the quality of being functionally adequate in performing the tasks and assuming the roles of a specified position with the requisite knowledge, ability, capability, skill, judgement, attitudes and values (177).

Continuing Education - learning experiences, formal or informal, designed to enlarge and/or update the knowledge and skills of practicing health professionals.

Credentialing - the recognition of professional or technical competence. The credentialing process may include registration, certification, licensure, professional association membership, or the award of a degree or other academic recognition in the field.

Criterion - a variable expressed in such a way that its presence or absence as evidence of adequacy can be ascertained, eg describes penicillin, blood pressure is 150/90. When used as a single measure, a criterion may represent a standard.
Licensure - process by which an agency of government, or its delegate, grants permission to persons meeting predetermined qualifications to engage in a given occupation and/or to use a particular title; or, grants permission to institutions to perform specified functions within their jurisdiction.

Peer Review - an all-inclusive term for health care review efforts involving the evaluation by practising health professionals of the quality and efficiency of services performed by other health professionals of the same discipline.

Registration - the process by which an agency or organization maintains a public list of individuals who meet predetermined criteria specified by the agency which maintains the register. Where the right to register and be registered is protected by law, registration may be considered licensure.

Standard - a criterion, or degree of measurement of a set of criteria, which may be considered adequate for a stated purpose such as an award, discipline, further education. A standard may be expressed as a single criterion or may be expressed as a point on a percentage or other scale of measurement with respect to the same criterion in a sample of cases or with respect to a set of related criteria in one or a sample of cases.

Variable - an attribute or type of behavior, eg physician's prescribing pattern or patient's blood pressure, which is accepted as an indicator of adequacy and which is expressed or used in such a way that a measure of the variable which will be accepted as evidence of adequacy is or must be additionally specified.

Organization of the Review

This paper reports relevant literature by addressing four major issues related to evaluating and maintaining competence of health professionals: definition, evaluation, action for restoration/maintenance, and social context. The first of these, definition, is fundamental to the other three. Theoretically competence connotes adequacy in a role. Realistically however, the variables and standards used to judge its pre-
sence or absence become its operational definition; i.e. competence is defined through the act of evaluation.

There are many ways that components of competence can be categorized for evaluation. Several models for evaluation in health care and education have been proposed (113,120,126,144,146,281,425,473). Because of its simplicity, relevance, and considerable influence on health care, the model proposed by Donabedian (144,146) is used as a frame of reference in subsequent chapters.

Donabedian suggests that all elements in the health care system can be classified in one of three major categories: structure (resources), process (activities), or outcome (results). Each can also be isolated, defined, and measured against a standard. Structure includes such variables as funds, abilities of personnel, space; process involves the behavior of practitioners, patients, and administrative systems; outcome focuses on whether services were used, the patient got better, or the community benefitted. When the scope of concern is narrowed from the health care system to the competence of practitioners, the three areas constrict to knowledge and skills (structure), clinical performance (process), and altered health state of the patient (outcome).

The literature reviewed is organized in eight chapters:

Chapter 1 - Introduction
- establishes purpose of the review, defines terms, describes the procedure used and the organization of the paper.

Chapter 2 - Assessing Potential for Clinical Performance: Structure
- considers developments related to judgements about and action on the knowledge, skills or attitudes practitioners bring to the job.
Chapter 3 - Assessing Clinical Performance: Process

- surveys developments in the definition and evaluation of performance during the giving of care and in activities undertaken to correct sub-standard performance or maintain and improve satisfactory performance.

Chapter 4 - Assessing Results of Clinical Performance: Outcome

- Considers efforts being made to judge and improve competence of health professionals by focussing on the results of care.

Chapter 5 - Assessing Combinations of Structure, Process and Outcome

- describes combined approaches to evaluation and follow up action which have developed largely in response to the disadvantages inherent in focussing on structure, process, or outcome in isolation.

Chapter 6 - Selection of Evaluation And Action Strategies

- identifies a number of areas where the literature is of sufficient volume and variety to permit tentative generalizations helpful to decision making in planning for evaluation of competence and follow-up action.

Chapter 7 - The North American Social Context

- discusses literature concerned with psychological considerations related to evaluation and with social and political developments relevant to evaluating and maintaining competence of health professionals.

Chapter 8 - Observations and Recommendations

- Offers some general observations and recommendations to continuing educators.

The variety and quantity of relevant literature has resulted in a complex structure for most of the chapters. It will be useful, before reading each chapter, to study the chapter's topical outline in the table of contents. In this way the organizational plan for each chapter, its contents, and their relationship will be more apparent.
Chapter II

ASSESSING POTENTIAL FOR CLINICAL PERFORMANCE: STRUCTURE

This chapter will examine recent developments in assessing the knowledge, skills and attitudes which practitioners bring to a clinical situation. Of concern are evaluation approaches which attempt to define and test desired professional behavior. The assumption which underlies a "testing" approach to assessment of competence is that those professionals who can perform in a desired manner under test conditions are likely to do so in practice. These approaches will be discussed within three broad categories: self-assessment procedures, formal testing of practitioners, and non-traditional testing procedures.

Self-Assessment Procedures

Use of self testing as a vehicle for estimating present competence and for motivating and guiding continued development of competence has probably received most attention from the medical profession. Caplan (86) enumerates several advantages of self-assessment: the physician has evidence of his knowledge deficits and therefore is motivated to learn; related learning is self-paced, private and non-threatening. Paper and pencil testing can readily provide information about recall of factual information and clinical problem solving, thus the fundamental bases of professional practice are tested. Finally, the choice of a corrective
learning mode is open and can accommodate to learner preferences and available resources.

Randomly Available Test Instruments

Probably the least individualized assessment tools available to practitioners are "test yourself" features which appear in professional journals. The "Self-Assessment Mini Program" which appears in the International Journal of Dermatology (375) is a good example of a regularly appearing journal series which briefly presents a clinical problem, asks relevant questions, discusses answers, and lists references for further study. "Nursing Decisions: Experiences in Clinical Problem Solving" (366) provides an example of this approach found in the nursing literature. "Thyroid Disorders" (213) is a variation on this approach; this is a self-study program, complete with objectives, which begins with provision of clinical information and then provides a testing and scoring service. Participants who complete the test successfully are awarded continuing education credit under the American Nurses Association's continuing education accreditation mechanism.

In addition to material available to practitioners in journals, at least one publishing house (464) has published an entire self-assessment series in which each volume focusses on a particular clinical specialty area. While the majority of these are for physicians or nurses, the series includes one volume on clinical pharmacy and a few on the various paramedical technologies. In addition to test questions and answers, the reader is given a rationale for the answers and direction for further reading. The publisher gives no information on the process used for initial development of the material, nor on whether there have been
attempts to establish reliability or validity. Performance norms, if they exist, are not given, and standards for adequate performance are not suggested.

All of these aids to learning assume that the practitioner will select those relevant to his practice, that he has access to suggested support materials if required, and that he is able to learn effectively to correct knowledge deficits. There is no evidence that this does or does not occur. Depending on chance to get the right self-assessment tools to the right practitioner at the right time is likely a highly inefficient application of this approach.

Self Assessment Procedures for Select Groups

The development of self-assessment tools useful to specifically defined practice groups presumably increases the likelihood of relevance to practice and of timely availability. The American Association of Respiratory Therapists (380) and the American College of Dentists (342) have both developed self-assessment procedures specifically for their members. The dental series is four tests of 150 questions each which covers all content areas of dentistry. Use of the tests, including confidential report of results, costs dentists $50.00 each. In a move to stimulate interest in the self-assessment test, the college developed a fifty item instrument which could be administered within an hour. This is currently being given in dental society meetings where participants score their own papers. Apparently response is favorable despite the fact that no direction for further study is given.

The major use of self-assessment procedures has been by medical specialty societies. Rubin (445) reports that 14% of all practicing
Ophthalmologists participated in the first self-assessment examination sponsored by the American Academy of Ophthalmology and Otolaryngology in 1971. This test covered the specialty's eight basic knowledge areas and took four hours to complete. Participants received confidential results, information on norms by time elapsed since graduation from medical school, and study references for each question. Although by more recent standards the approach used was somewhat crude, Rubin's account of the Academy's experience with planning, publicity, test administration, and user reaction is instructive. He reports that many participants failed to understand the strengths, limitations, and appropriate use of exam results. One question the report raises is, "is it more helpful to survey all knowledge sub-areas equally (as this test did) or to weight the exam to reflect typical case load?" One might conclude, as the Academy did, that a survey of the universe of knowledge applicable to that practice area is most helpful, but this avoids the question, "is there any point in sharpening knowledge on something you are not likely to use?" This decision has been left to the practitioner.

The American Academy of Neurology (324) has designed its "education exam" primarily for residents and newly graduated neurologists to assist them in identifying weak areas needing further study before taking Board examinations. Evaluative data from the 1972 and 1974 exams show the tests to be reliable, discriminating, and valid. Scores are released to residents, their training directors, and to practitioners. There is no follow-up and no guidance for further study.

Far more structured as an educational tool is SESAP, the Surgical Education and Self-Assessment Program offered by the American College of Surgeons (467). This is described as a self-contained,
comprehensive aid to continuing education containing over seven hundred assessment items, black and white and color illustrations, and an authoritative critique (essay answer) of each item with references and reprint service.

The American College of Physician's Medical Knowledge Self-Assessment Program (MKSAP) is designed not only for a specific group, but also for a specific purpose (78,436,539). Originally a self-assessment program much like those described above, the third and fourth versions of this program were designed to assist physicians to prepare for the American Board of Internal Medicine recertification exam. Enrollment for MKSAP III was thirty thousand, and the recertification exam was passed by 95% of those enrolled who took the exam. MKSAP IV included a syllabus in the form of a concise study guide of current information (what experts believed to have been important developments of the last five years); test materials for eleven subject areas with factual and clinical problem solving questions in each; evaluative data (participants can self-score or have this done by computer); and comments, reference books, and illustrations for reviewing one's answers. Evaluation of MKSAP has been limited to quantifying its popularity and participants' success on recertification exams; no effect on practice has been demonstrated nor is this planned.

It may be concluded that, with the exception of medicine, the health professions have not as yet shown great interest in the self-assessment approach to maintaining competence. In summarizing the extent to which self-assessment is used in medicine however, Leveridge (296) states that over twenty medical societies now offer programs to members; he projects that this practice will grow. While individual programs
vary, most include testing of knowledge required for specialty area performance, direction for further study, and provision of the necessary materials for study. Those programs which score tests for members appear to go to great lengths to protect participants' anonymity and most use information on aggregate tests results to provide direction in planning for other types of continuing education.

There seems to have been little effort to ascertain whether self-assessment activity has a positive effect on practice behaviors or on health outcomes. Rather, providers and users have tended to be content to assume that relevant knowledge increases as a result of participation and that this, in itself, is sufficient justification. Butt (78), in discussing evaluation of self-assessment programs, states "some things that count can't be measured and some things that can be measured don't count."

**Formal Testing of Practitioners**

In contrast to self-assessment, formal testing procedures are less likely to focus on motivating and aiding improvement of essential knowledge and more likely to be used for a summative evaluation of a practitioner's knowledge, skills, or attitudes. These assessment procedures are characterized by controlled administration, and often concern for security, to ensure that the evaluatee cannot cheat. Information on this group of assessment procedures will be considered in three categories: tests which survey the full range of professional practice within the discipline, tests which focus on a specialty area of practice, and a small group of other formal special purpose assessment procedures.
Test for Full Range of Practice

Broad spectrum testing of practicing professionals is almost non-existent, perhaps reflecting the reality that once past screening for initial licensure, the typical practitioner's scope of practice narrows immediately. Indeed, the first of the three assessment procedures in this category mentioned in the literature is often used prior to licensure.

The examination provided by the prestigious National Board of Medical Examiners (NBME) is intended to evaluate a physician's readiness for independent practice. While participation is voluntary, the status rewards are high and many states will accept success in the "national boards" in lieu of a licensure examination (105). Formed in 1916 with the objective of offering high quality exams in the practice of medicine for external agencies' discretionary use, the NBME exam evolved as a three part assessment procedure: I, basic sciences; II, clinical subjects; III, lab and bedside practical exam (128). Parts I and II have undergone considerable refinement as a result of research and development in cognitive testing and today are largely constructed in written objective item format.

In 1968, because of the difficulties in controlling clinical situations and achieving reliability in scoring, part III underwent a major revision. The NBME chose the critical incident technique as a way to define what will be tested. Over three thousand incidents were collected from over six hundred doctors who supervised interns. The incidents fell into nine major areas: history, physical, tests and procedures, diagnostic acumen, treatment, judgement and skill in giving care, continuing care, physician-patient relations, and responsibilities of a
physician. From these, three separate methods were devised for use in part III: (a) motion pictures and multiple choice items to test observation and diagnostic skill, (b) patient management problems to test clinical problem solving, (c) presentation of combinations of patient data, e.g. lab test results, x-rays, recordings, tracings to test clinical information. While without question this examination assesses a broad range of physician behavior, it should be remembered that attitudes and interactive skills are tested only indirectly and motor skills not at all in the NBME format.

The New York State Regents External Degree Program in Nursing (B.S.N.) moves beyond the constraints of the NBME exam and assesses performance in the clinical setting (509). Still under development, this program will make available to practicing nurses a baccalaureate degree in nursing acquired through successful challenge of a sequential battery of cognitive tests of general subjects, relevant clinical subjects, and three performance exams: clinical performance, which involves application of nursing process in caring for several child and adult patients in the hospital setting; health assessment performance, which tests ability to obtain, analyze, and summarize essential clinical information; and professional performance, which tests ability to use a variety of resources and strategies in assisting clients and their families in a variety of settings. The clinical performance examination is the only one of the three performance exams developed and tested to date. It is also used as part of the assessment procedure for the New York State External Degree Program in Nursing (Associate Degree) which is preparatory to licensure.
Because controlled use of the clinical situation as test material is unique, the procedure is worth examining in some detail. The exam consists of one out of two possible clinical laboratory tests (e.g. demonstration of the correct preparation of medications), followed by examination in two out of a possible three adult patient care situations and in one out of a possible two child care situations. In each of the patient care situations, the evaluatee must comply with pre-established standards for planning and evaluating care and must demonstrate all of the specifically defined "critical elements" of nurse behavior. There must be no jeopardy of patient physical or emotional safety and no violation of asepsis. There are twenty possible areas of care which can be tested (295). Lenburg (294) describes the general procedures which were used to develop the examination and discusses as well the procedures used for orienting hospital staff and training evaluators. Reliability is satisfactory. By July 1976, thirty-eight applicants had completed the exam and 92% of these had passed State Boards (the national average is 81% for associate degree graduates and 88% for baccalureate degree graduates).

Libby (303) has developed a broad assessment approach for evaluation of pharmacy technician performance; this has been used only for research. He developed a combination instrument that included a tool for attitude testing, a fifty item cognitive test, and five skill appraisal items to determine proficiency at thirteen defined technician performance requirements. Of most interest was his finding that both technicians and the pharmacists with whom they worked overestimated the technicians' ability to perform at an acceptable standard. The work has not been replicated.
Testing in Specialty Areas

Current literature reveals little of what is probably the existing level of activity in testing competence in specialty areas. No doubt this is because specialty certification is not new to the health scene and therefore likely to attract little attention.

Benson (46) describes the American Board of Internal Medicine three level certification process. This consists of three years of residency training, ongoing evaluation of clinical competence during training, and cognitive testing. While the Board is exerting some effort to improve training evaluation procedures, the mainstay of evaluation is the certification exam. Performance in this correlates highly with clinical evaluation ratings in those sub-specialties which have used structured rating systems during the residency period. The Board has developed a matrix of three dimensions: tasks of the physician, e.g. take a history; requisite abilities (cognitive, affective, psychomotor); and clinical problems, e.g. diseases and organ systems involved. This has provided the framework against which the exam is constructed. Recertification will be by written exam also; the Board has considered and rejected peer review, chart audit, and oral and computer exams as not presently meeting the criteria of validity, reliability, and feasibility.

The certification exam of the Canadian College of Family Physicians is somewhat bolder in its approach and broader in scope (320). It uses multiple choice and patient management problems as well as film strips, videotapes, orals, role playing and simulated office visits using professional actors. No information is given on reliability or validity testing.
Certification is not so widespread among the other professions. Both dentistry and nursing maintain specialty certification programs which are described in the literature in terms of their purposes, procedures, and component parts 23,24,30,91,148,546). Details about test construction and evaluation are not provided however.

Special Purpose Assessment Procedures

The literature provides information about only three additional assessment procedures; two of these were developed for highly specific purposes. Singleton (475) describes assessment procedures developed to confirm successful completion of a nurse practitioner educational program. Johnson (264) administered a nutritional knowledge test to physicians to identify the need for medical school curriculum changes, continuing education, and consultation from nutritionists. The third procedure, the use of a skills inventory, is described in general terms by Cantor (84). These employer-produced checklists are most often self-administered by the employee and are used as a basis for clinical placement and inservice planning.

Non-Traditional Testing Procedures

A final group of testing procedures, primarily in developmental stages, will be described in this section. As might be expected, these are largely non-traditional evaluation approaches and are not extensively used. Included are the description of a highly individualized testing system, a group of tests of clinical decision making, and a final group
of assessment procedures which look at interactive skills.

Individualized Physician Profile

The University of Wisconsin Individual Physician Profile (395, 476) is an experimental program which attempts to provide a self-assessment instrument which is relevant to the practice of the individual physician. Diagnoses from the subscriber physician's practice are coded according to the International Classification of Diseases, Adapted (ICDA) and fed into a computer which contains a test bank of some four thousand items coded for diagnostic category and degree of difficulty. The computer categorizes the diagnoses and prints out a multiple choice exam which reflects proportionately the categories of diagnoses present in the practice. The greater the frequency of a particular clinical problem in the physician's practice, the more complex the related questions become.

Once the exam is completed and scored, a computer analysis of the physician's areas of strength and weakness is sent to him along with a set of recommendations drawn from a computerized Education Resource Index, which includes everything from conferences to home study materials. Earlier in the project, personal educational consultation was used to interpret test results and aid in planning for further education, however this has apparently been abandoned.

The major disappointment of the program is its underutilization and cost in time and money. Evaluation has shown that users find the program of value in ascertaining educational needs.

Tests of Clinical Decision Making

The most well known of this group is the patient management
problem (PMP), wherein a clinical problem is described in some detail and the evaluee is asked to make sequential decisions about its investigation and management. While the National Board of Medical Examiners has used this assessment approach as a testing method since the late sixties, the PMP is still evolving.

Barro (41) has described Christine McGuire's current work (322) on patient simulations at the University of Illinois as different from that used by the NBME, in that less beginning information is given and a branching progression through the problem gives the evaluee information only in response to his diagnostic questions before requiring him to make further diagnostic for therapeutic decisions. This interaction with the problem is presumably a more accurate simulation of reality — indeed the test taker may even "kill" his patient thus ending his test problem prematurely. In McGuire's simulations, all possible responses are classified and weighted by experts to reflect the relative value to the patient of the therapist's decision. An Efficiency Score (per cent of helpful responses) and a Proficiency Score (per cent of actions agreeing with experts' decisions in selecting helpful and avoiding harmful actions) are combined to form a Competence Score. Reliability of this approach is high. Validity has been checked by chart audit of real patients; although fewer chart items were recorded than were selected on comparable PMP's those that were selected in PMP's were found in charts.

Whether performance on PMP's correlates with more direct performance measures is unknown, although Palva and Korhonen (378), in a study involving performance on a PMP related to drug-induced agranulocytosis, found that outcomes on the simulation correlated with actual outcomes in a local medical service. Specifically, 10% of the student
treated patients in the simulation would have died from improper treatment; experience with the local medical service showed an identical mortality rate over a four year period. This rather unique study was characterized by naturally occurring controls that would not ordinarily have been available to validate the PMP. The local medical service was staffed almost exclusively by graduates of the training program, and the disease entity is sufficiently rare that graduates could not have learned by experience in the intervening period. Moreover, agranulocytosis mortality is directly attributable to physician decision making and action, and not subject to other variables. The authors therefore concluded that for this clinical problem, performance on the PMP was predictive of clinical performance.

The MERIT Project (Model for Evaluation and Recertification Through Individualized Testing) was developed with financial assistance from the American Board of Internal Medicine and has received considerable attention (180,386). The subject physician develops a personalized practice profile by using a diary technique for one hundred consecutive patients to collect data on chief complaint, exam and lab results, treatment, long term management and outcomes. These sample cases are boiled down to three representative patient care simulations which the physician "treats" through a computer terminal; the interaction is conducted in natural language. Performance is scored on five dimensions: thoroughness of examination, diagnostic accuracy, efficiency, appropriateness of management, and danger and discomfort to patient. Feedback is immediate.

Validation of the MERIT procedure has been accomplished using an audit of office records for critical concepts. Those concepts found in actual patient management were also applied to the physician's management
of the computer cases; correlations were acceptable. Development cost is an obvious problem with the MERIT approach. Each simulation takes six hours author time, twenty hours for trial and revision, and fifty hours for programming. Although MERIT was developed as a possible recertification procedure, the high costs associated with test development and security measures make it more useful as an educational tool. Additionally, the use of a personalized rather than a standard test for summative evaluation could be a contentious issue. There seems to have been little exploration of its potential as a diagnostic tool for planning the practitioner's continued education.

Another decision making test, Rimoldi's test of diagnostic skills, is also described by Barro (41). This simple instrument presents the evaluee with a written patient situation and a set of small cards, each of which shows on one side a history, physical or lab question which he may choose to "ask" by turning the card over to display the information sought. Diagnostic skill is scored according to the number of questions asked, their usefulness, and sequence. No doubt the latent image and computerized application of PMP's are seen as easier to manage than packs of cards, thus accounting for the fact that this system appears to be little used. Nevertheless, it is appealing in the simplicity of materials required.

Tests of Interactive Skill

In her review of physician assessment procedures, Barro (41) has identified a few little-used but innovative approaches. She notes that Elstein in 1972 used actors to provide a history (physical assessment data is simply provided in writing); the physician is assessed by
observers for his interviewing and counselling skills using a criterion check list. This allows some control of the variables with which the physician must deal in the actual clinical situation. The amount, type and order of information gathered is subject to scrutiny. Reliability and validity of this approach have not been established.

Barro also summarizes several attempts at evaluating patient-practitioner communication: Gozzi (1969), Korsch (1968), Davis (1971), and Addler and Enelow (1966) all based their assessment attempts on Bale's Interaction Process Analysis. Gozzie and Korsch counted "blocks" and "facilitators" in an episode of communication (and demonstrated a reasonable degree of reliability and of validity) to produce a performance index. Davis developed a more elaborate coding system for twelve categories of communication behavior. This adaptation of Bale's system was reproduced in Barro's review and typifies attempts at communication measurement under test conditions (see Figure 1).

Summary and Comment

This chapter has reviewed approaches to evaluating the knowledge, skills and attitudes which practitioners bring to the patient care situation. The three broad topic areas covered were self-assessment procedures, formal testing of practitioners, and non-traditional testing procedures.

Self-testing devices available in professional literature are popular as continuing education measures. They assume a practitioner will select test material relevant to his area, will have access to re-

**Figure 1**
lated learning materials, and will be able to learn from them. There is not evidence that this happens. In an effort to overcome these potential liabilities, over twenty (296) of the specialty medical societies in the United States have developed self-assessment procedures for their members' use. While these vary in their approach and degree of sophistication, a typical program will test the physician, provide feedback about areas of strength and weakness, direct him in further study, and provide relevant learning materials. Aggregate test results are used to plan continuing education programs for groups. There seems to have been little effort to ascertain whether self-assessment activity has a positive effect on practice behavior or on health outcomes. Rather, providers and users have tended to be content to assume that given information about knowledge deficits, the practitioner will be motivated to learn, and that this in itself is sufficient justification.

Formal testing procedures for practicing professionals are limited to summative evaluation procedures used primarily for credentialing at an advanced level. There are few attempts to evaluate a broad range of practitioner behaviors, reflecting the reality that the scope of practice generally narrows to a specialty area within the discipline soon after licensure occurs. Although active specialty certification programs exist, the details of test development and construction for these programs are infrequently reported in the literature. Occasionally formal testing procedures are developed for special purposes such as research. It does not appear that formal testing procedures are used to assist practitioners with professional development.

The literature also reflects a continuing and active interest in innovative approaches to testing health professionals. Most of this
developmental work appears to be taking place within the medical profession.

It may be concluded that to the extent that assessment of the structure of health professionals' competence exists, the professions are inclined to be thorough in their testing of lower levels of cognitive functioning, are beginning to find effective ways to assess ability in clinical decision making (which is considered by many to be the most critical area of practice) and are just beginning to explore testing psychomotor skills or attitudinally related professional behaviors. Perhaps this last is more understandable, as consensus on values is difficult to achieve in society at large. It is to be hoped that the New York State clinical nursing evaluation experience will prove that it is possible to reliably measure and make judgement about the adequacy of a wider range of professional behaviors which the practitioner brings to the job than has been commonly assumed to be possible in the past.

Accepting that reliable evaluation of capability for certain aspects of professional practice is possible, and that tests for evaluation of a fuller range of clinical behaviors can be developed, it is important to recognize that even the most desirable of these assessment techniques are but proxy measures for real-world clinical performance. For this reason evaluation of the structure of competence may always be more useful in formative than in summative evaluation.
Chapter III

ASSESSING CLINICAL PERFORMANCE: PROCESS

Current concern with the issue of competence derives from a desire to assure that practitioners will do the best thing possible for their clients. The most direct assessment of competence will judge the practitioner's behavior, i.e. what he does, as he carries out his professional functions. This chapter examines recent literature concerned with assessing the knowledge, skills, or attitudes which practitioners exhibit in clinical practice. There are three major sections: systematic assessment of total performance, systematic assessment of situation-specific performance, and unplanned evaluation of performance.

Systematic Assessment of Total Performance

While it is an overstatement to say that the full range of any practitioner's performance is ever assessed, this section will review reports of assessment procedures which evaluate a broad spectrum of professional practice behaviors. Many of these schemes are open ended; i.e. any behavior which can be observed or measured will be considered. Others identify the areas of behavior to be assessed, but these areas collectively include a wide range of possible professional behaviors. This literature is categorized and discussed under four headings: general discussion of systematic performance evaluation; performance evalua-
tion using implicit standards; performance evaluation with explicit, but imprecise standards; performance evaluation using explicit standards.

General Discussions of Systematic Performance Evaluation

There are few general statements or guidelines for performance evaluation which provide substantial information to evaluators. Perhaps the most comprehensive overview of performance appraisal found in this literature search is provided by Haar and Hicks (210). These authors describe the advantages and disadvantages of the most common types of assessment approaches: essay technique, field review (several people review performance and arrive at consensus), forced choice (the rater must choose descriptive statements of behavior most and least like the evaluatee), collections of critical incidents, ranking of evaluatees in relation to each other, graphic and numerical rating scales, check lists, quantitative work standards, and assessment against pre-established work objectives (management by objectives). The twenty-four references, mostly from management literature, provide further information for the reader. Marriner (328) discusses the pros and cons of many of these same techniques; she also describes such common rater errors as that of halo, logical association, proximity, central tendency, and leniency, and suggests techniques to overcome these. Detailed discussion of how to evaluate clinical performance using the critical incident technique is presented by Weinstein (527); check list, by McPherson (323); and rating scales, by Dohner (140).

Several writers (9,210,274,328,487) outline desirable attributes for any performance evaluation system. Collectively, they conclude that assessment techniques must be selected or designed for the purpose
for which they will be used and must be congruent with the agency's philosophy and goals; standards must be clear and unambiguous; there must be careful documentation; evaluators should be trained in the use of the instrument; in avoiding common rater errors; and in problem solving and interviewing; management must support and enforce evaluation policies; assessment should be ongoing to provide for immediate and meaningful feedback; and the primary focus should be the evaluatee's continued growth. Albrecht (9) emphasizes the importance of evaluatee participation in the evaluation process if behavior change is to occur. Stevens (487) gives considerable attention to summative as well as formative evaluation, and lists common reasons for failure of the evaluation process when it must result in discipline: poor documentation, inappropriately severe discipline, use of standards not related to the evaluatee's responsibilities, and failure to use due process.

Performance Evaluation Using Implicit Standards

The most unstructured approach to performance evaluation is to leave entirely unspecified either the parameters of behavior to be evaluated or the standards of acceptability. A strategy which provides slightly more structure is to describe the parameters or variables to be assessed, but to leave the standards unstated. Literature concerned with both of these approaches is discussed below.

The Massachusetts Department of Health (330) reports evaluating competence of employee physicians by the simple expedient of commissioning an extra-mural site review team of three respected physicians to assess adequacy of medical services in the several hospitals owned by the
state. The review team is apparently at liberty to decide which areas of performance to evaluate and to decide what level of behavior is acceptable. The team is reportedly able to identify superior, adequate, marginally adequate and incompetent physicians.

Allowing the performance variables and standards to remain unstated before review seems to be used frequently in psychiatric services. Two reports (297,360) from mental health centers describe using a multi-disciplinary case conference format in which the therapist's management of a case is evaluated and recommendations are made. Both reports indicate that this approach retains the flexibility necessary to accommodate different therapeutic styles, is highly educational, is useful in identifying problem areas, and achieves savings by reducing unnecessary or prolonged use of services. Because no measurement against stated standards take place over time however, it is difficult to support these claimed advantages with evidence. One of these authors (360) has indicated that where it is necessary to shift focus from education and consultation to judgement of adequacy, i.e. when performance falls below the implied minimal standard, there emerge a number of problems which are difficult to resolve. Kerstein and Weissman (276) were able to use a similar approach when they reviewed the charts of thirty-six non-hospitalized suicide attempters. They found roughly one-half of these cases had been adequately treated and identified deficiencies both in residency training and psychiatric services delivery.

The practice of specifying the variables to be evaluated and leaving standards unstated is also commonly seen. Hadamard et al. (211) report analyzing consultation requests from hospital employed physicians in order to assess the adequacy of their performance. In this instance,
the parameters to be considered were specified using a form developed elsewhere to evaluate candidates for board accreditation in internal medicine; they addressed such broad variables as history taking, completeness of physical exam, etc. Standards of acceptability were left unspecified. Physicians took seven minutes each to analyze two hundred sixty consultation notes and judge performance. Their judgements correlated with subjective ratings given the hospital doctors by their superiors, thus leading the study group to conclude that the approach has merit.

Another study team (461) analyzed transcriptions (from dictation) of hospital employed physicians' initial assessments and treatment plans. Again broad areas of behavior to be considered were specified, but the standard of adequacy left to the evaluator. Deficiencies were identified by category of physician (e.g. clerks were high in omissions from history, junior residents handled pneumonias poorly, etc.) This approach was sufficiently sensitive that senior medical staff invariably scored better than their juniors. Reliability tests were not done.

The literature indicates that the task of identifying performance variables for evaluation has been approached in a variety of ways. Kimball (274) suggests that the job description be analyzed to define and describe broad categories of nurse behavior, e.g. leadership. She then goes on to suggest that performance can then be evaluated by collecting anecdotal descriptions of behavior, classifying these by category, and estimating the degree of acceptability of each area of behavior. The University of Miami (305) approached this task by defining broad behavioral categories from a curriculum framework and rating medical students against a four point scale. Test-retest reliabilities are reported as
good, correlations between self and peer ratings were high, as were correlations of ratings with grades. It is suggested that this approach could be used for rating performance of hospital medical staff. The question of whether ratings would correlate with other measures of performance is discussed.

Other approaches to defining categories of behavior are also suggested in the literature. Harris (219) developed an observation guide for nurse practitioners by starting with a literature review; then organizing the material using a conceptual framework "dimensions of care," e.g. assessment; and finally having these validated by a nationally recognized expert group. In another study, Ward (524) identified fifteen areas of health needs and problems, specified related competencies, then used nurse practitioners in fifty states to validate the information. Return data were also correlated with practice, population, and economic settings.

The American Nurses Association's Standards of Nursing Practice (20), which specify the dimensions which must be considered when selecting measurable criteria for assessment of performance, utilize the steps of nursing process as a conceptual framework. These were derived primarily from expert practitioner opinion. The standard statements exist in sets: a generic set which is applicable to all nursing care situations, and a number of specialty sets which are applicable to such areas as maternity nursing, mental health nursing, etc. They are intended to and have been used as a point of reference for evaluation systems.

Lamberton et al. (287) report using the ANA standards in developing peer review skills in a masters level family nurse clinician program. This program provides structured opportunities for nurses first
to develop personal comfort through incidental contact with behavioral objectives and through informal giving and receiving of feedback in relation to these, then to enter a more formal phase of case presentation and group review in the presence of faculty, and finally to practice one-to-one unsupervised peer review using patient records. As with those instances described above, the elements of performance to be considered are specified, but standards of acceptable performance are left implied. This article would be of interest to any professional group concerned with developing attitudes and skills in evaluation.

Gudmunsen (206) also used the ANA Standards to organize a questionnaire to rate nurses' clinical performance. The rather high correlation between performance scores on this tool and State Board (licensing exam) scores of study participants indicates that this tool may have some validity and perhaps merits further development and testing, since it is more easily administered than direct clinical observation or records review. The questionnaire (self and supervisor opinion) technique has also been used and replicated with some degree of success in nursing by Dyer (150,152). These last two evaluation approaches seem more suitable for research than for ongoing evaluation of clinical competence.

In devising a way to assess physician performance in pediatric cardiology, Johnson et al. (263) used yet another approach to identifying areas of clinical performance. Two eminent cardiologists, for ninety minutes, dictated all of the knowledge, skills, and attitudes they expected to see in a competent practitioner. These were later edited for importance, categorized, and validated by medical center staff against similar statements by the relevant specialty medical society. The
authors acknowledge several limitations in the tool, but found it useful in providing direction for curriculum planning and counselling of medical students, interns and cardiology residents. The tool seemed to be acceptably sensitive. Finally, Donabedian (144) used historical and contemporary views of the physician role to produce an exhaustive list of desired physician behaviors which might be properly evaluated.

Performance Evaluation With Explicit But Imprecise Standards

There appears to have been at least some effort not only to develop fairly complete descriptions of behavioral elements to be evaluated, but also to develop measurement systems permitting a statement of standards of acceptability which, while not entirely precise, are more explicit than implied.

The Tate Instrument (353) was developed in the mid-sixties and is still widely used and apparently influential as an evaluation approach. Tate collected critical incidents of staff nurse behavior which had been recorded by two hundred forty head nurses and classified them into five major categories: knowledge and judgement, conscientiousness, skill in human relations, organizational ability, and observational ability. The rater observes and records significant nurse behaviors (critical incidents) and, using these as an information base, estimates the evaluatee's level of performance in each category against a visual scale which ranges from very unacceptable behavior to outstanding behavior. The rater is assisted in this by being provided with "benchmark" descriptions of nurse behavior, arranged at given points along the scale, which are thought to be representative of the behavior at that level. For instance, the "less than acceptable" benchmark on the human relations scale
might be "this nurse could be expected to ignore patient's comments regarding fear of what he will say under anaesthesia". Benchmark descriptions are chosen only when many nurses agree on their place on the scale. The critical incidents recorded by the evaluator are used for formative evaluation and then discarded. The completed rating scale is retained as part of the summative evaluation record. On tests, reliability varied from $r=0.29$ (observational ability) to $r=0.70$ (human relations). The categories and behavioral descriptions used for the scales are now dated, but the approach is appealing in its simplicity. The work apparently has not been tested against other instruments.

Elenbas and Jacoby (157) used a similar approach to develop a behaviorally based rating scale for pharmacy residents. The final version of the instrument measured twenty-two characteristics within four broad categories (drug information, patient work-up and monitoring, communication and professional interactions, service responsibilities), with sixty-eight benchmark behavioral description being provided. The authors indicate that the tool is applicable to almost any clinical pharmacy situation: there is no information as to whether or not the tool has been tested for validity or reliability.

Other measurement scales have been developed according to frequency of performance, complexity of performance, or degree of supervision required. Multnomah County Oregon public health nurses (261) identified ten facets of interpersonal communication, then asked peers to estimate anonymously according to the frequency, e.g. 20% of the time, with which these behaviors were demonstrated. Results were useful for counselling, although acceptability of this system was not totally positive. Results were also used for input, along with supervisor rating of practice and self-assessment, for decisions about promotion.
Bidwell and Froebe (52) described levels of nurse performance within several behavioral categories on both clinical and administrative tracts using Bloom's taxonomy as a way to organize tasks within a category according to increasing levels of complexity. Lower positions within the nursing services demanded performance of less complex tasks worth one or two points, while higher positions demanded more complex analysis and synthesis behaviors and carried higher point values. Although this tool rather distorts the meaning and intent of the taxonomy, nevertheless a progression in the complexity of skills is apparent in the task lists which are reproduced in the article. The scoring system allows for a yes-no response for simple tasks and for a greater range of response for the more complex tasks. If the rater chooses not to award full point value assigned to a given task, this decision must be justified.

The University of Michigan Medical Center (49) used literature and job descriptions to identify and define "major nurse performance factors", and then assigned to each category three to five levels of performance according to the degree of supervision required. Each level was assigned point values. The more independently a nurse can perform a given function, the greater number of points are awarded. To be rated at a given level, the behavior described in that level must be regular and recurring in the nurses's practice. The total points earned determines the size of the merit pay increase, and there is a minimally acceptable level below which counselling, discipline or termination occurs. This is an interesting application of the merit principle as a basis for decisions about salary increase.

A final approach which makes measurement standards more explicit used the nurse rater's knowledge of the range of performance possi-
ble. Slater (220,223) developed an eighty-four item scale arranged in six sub-sections: psychosocial individual, psychosocial group, physical, general, communication, and professional education. The five point scale against which the evaluatee's action is compared for measurement is the care known by the rater to be given by first level staff nurses. The rater is asked to think of the worst and the best nurses in this category which she has known. Using these as reference points, she is then instructed to enter the names of nurses she has known at all five points on the scale. The instrument has considerable flexibility in that the scale can be expanded to nine or eleven points; can be used in any clinical setting (there are a number of cue behaviors which both provide concrete examples of behaviors which fall within each item and which can be varied to suit the clinical setting); and can be used for retrospective or concurrent ratings. In the latter case a two and one-half hour observation period is necessary. The scale has been shown to be sensible and surprisingly reliable. It is apparent that reliability would suffer with a rater whose past clinical experience is limited or atypical.

The QUALPACS Scale (522) uses the same approach to measurement and employs many of the Slater Scale items. QUALPACS, however, is designed to measure nursing care received by a given patient or set of patients and thus is suitable in most instances only for judgements about the competence of a service group such as the nursing unit staff. Measurement of its sixty-eight items must be made by a nurse using direct unobtrusive observation of patients receiving care over a two hour period and by indirect means such as chart review and discussion with personnel. The total time to review care received by one patient is three hours, but the rater can observe two patients simultaneously. QUALPACS
achieves an acceptable level of reliability and has been tested to a limited extent for validity. An interesting independent study by Smolinski (481) showed that when twenty-eight nursing acts were derived from the QUALPACS Scale and submitted to seventy nurse-patient pairs in two hospitals, there were not significant differences in care as perceived by patients compared to judgements of nurses who used this instrument. Both QUALPACS and the Slater Scale have proven useful in identifying educational needs, measuring change after educational or administrative intervention, and in identifying management changes necessary in the delivery of service (162,220).

Performance Evaluation Using Explicit Standards

While evaluative approaches described in this section still rely heavily on a rater judgement, they are constructed so that the behavior to be considered and the decision about level or standard of performance can be made in a fairly precise and readily quantifiable fashion.

The simplest of these instruments are check lists of what are thought to be the critical practitioner behaviors. The evaluator simply notes presence or absence of each behavior, and the level of performance is stated as a composite or percentage of performance items exhibited.

Physical therapists (238) developed such a system for use in peer review, and occupational therapists (216) employed a similar approach. Both groups identified weak areas of performance and found this a good base for educational planning; the occupational therapy group reported documented improvement in the quality of performance.

Rubin (444) describes one hospital's experience with evaluating
nurse performance using explicit statements of desired behavior first through the use of chart audit and later through the use of other data collection techniques including direct observation, patient interview, and incident report reviews. The index (per cent of requirements met) serves as an indicator of unit achievement; like QUALPACS, this is patient-centered and thus one can evaluate competence of the nurses collectively who cared for that patient, but not that of an individual nurse. All indices of less than 100% are reviewed and typically lead to recommendations for both administrative change and inservice education.

Phaneuf's Nursing Audit (347,390,391,392,393) is a widely used chart audit tool applicable to most settings where the patient record is an integral part of providing comprehensive and continuing nursing care for patients. The audit is applied retrospectively to a ten per cent sample of all discharged patients or fifty cases, whichever is greater, and takes fifteen minutes per chart. The numerous process items to be evaluated were developed using as a framework seven basic areas of nurse performance originally proposed by Lesnik and Anderson; scoring options are yes, no, uncertain and not applicable. Each performance area is weighted by importance (e.g. "observation" is twenty per cent of total score, "reporting and recording" is ten per cent, etc.) so that total score reflects emphasis on those areas deemed most important. Although items are constructed to reflect a broad spectrum of nursing functions, data collection is limited to information which can be obtained from the chart. Deficiencies identified pertain to all nurses who contributed to the chart. The author cautions that these must be viewed in the context of the agency's resources and service policies when corrective actions are being planned.
Pelley (384) reports on the development of a method for concurrent review of nursing care which used Abdellah's twenty-one problems as an organizing framework. Geared for minimum demand on resources (one-half hour per patient), the audit reviews for evidence of specific nurse behaviors in the care plan, record and at the patient's bedside and can be used on day or evening shift. The audit has been found useful for immediate feedback and influence on nurse behavior. Performance indices are computed, and consistently weak areas become top priority for in-service. Ramirez (404) describes how one hospital instituted a thirteen item audit to assess quality of written nursing care plans. A sample of five plans per nursing unit every other month was audited for a qualitative index (per cent of applicable criteria met). This was averaged with a quantitative index (per cent of patients on unit with plans) to yield a unit score. This procedure, in combination with feedback and education, has been shown to be effective in increasing the quantity and quality of nursing care plans.

The Hospital for Sick Children in Toronto (526) developed a far more sophisticated process measuring instrument for nursing service which still makes minimal demands on resources. This institution chose to attempt isolation of a few variables that were highly indicative of care in preference to the usual route of developing a comprehensive tool where several items are frequently highly related and provide little additional information. Development involved using expert opinion to identify criteria, and practitioner response to validate their importance; as a last step criteria were examined and those found to be redundant, invariant or irrelevant were deleted. A six point scale was used. When the instrument (SAVE - Selected Attribute Variable Evaluation) was validated
against QUALPACS however, it was found to be overly generous and not sensitive. Items from both QUALPACS and SAVE were subjected to several criteria (frequency of occurrence, sensitivity to range of quality and change in quality, ability to reflect a defined area of care etc.) to yield twenty-two items. Sub-sets were then constructed, based on correlations within the overall instrument, and all sub-sets now show a correlation of .90 with QUALPACS. SAVE requires about one-half hour of review time per patient and is reported to be particularly effective when used for immediate feedback to and counselling of practitioners.

Jackson and Smith (254) employed the concept of indicator behaviors (selected behaviors thought to predict performance for a category of behavior) to evaluate community pharmacist performance. Competence was defined using five categories of pharmacist functions derived from an American Pharmaceutical Association task force report on pharmacists' roles. Two trained "shoppers" and a clinical investigator rated pharmacists' performance on criteria designed to represent the five categories. Although a composite index of the quality of performance could be calculated, the fact that no attempt was made to manipulate the weights of the criteria (which varied considerably in number for the various categories) makes the quality index questionable. Nevertheless this approach might be a helpful one to identify continuing education needs of community pharmacists.

Systematic Assessment of Situation-Specific Performance

Recently, increasing attention has been given not to whether performance as a whole is adequate, but rather to whether performance was
or is adequate in relation to a specific clinical problem. Usually the situation chosen for review will be care of patients in a specific diagnostic category (e.g. pneumonia), receiving a given procedure (e.g. laminectomy), falling within a given age group (e.g. the hospitalized toddler), or exhibiting a particular functional problem (e.g. immobility). The evaluation study is usually limited in time, rather than ongoing. Typically a study topic (e.g. pneumonia) is chosen, performance criteria (e.g. the physician will order a flat plate of the chest) are specified, all or a sample of applicable cases which fall within the study topic are surveyed for compliance with performance criteria, and an index (per cent of compliance with each criterion) is computed. Varying standards of acceptability will have been set, depending on the purpose of the study and the perceived importance of the criterion. Usually follow-up action is taken to correct a less than acceptable index, and theoretically a second study is done to test the effectiveness of corrective action.

This evaluation-action procedure and its variations are known most frequently in the jargon of the health sciences literature as a "quality assurance program". Although borrowed as a concept from the field of industrial quality control, Clement Brown (70,71) is most often credited with its popularization in health. Because of its recency and constantly changing characteristics, the literature in this section is particularly difficult to categorize in any meaningful way. One could use, as a organizing concept, a time frame (prospective, concurrent, and retrospective review procedures); a source of data frame (direct observation, opinion based on recall, charts and records, record abstracts, and payment claims); purpose (cost control, improvement of quality, summative
evaluation of quality); type of follow-up action (administrative action, education, counselling, claims rejection, discipline, etc); or the administrative structure which organizes the evaluation of care (taken from simple to complex, these would be individual practice, agencies, and regional or other systems organized to provide health care).

None of these organizing frameworks emerges as outstandingly useful or simple. Further, the available literature lacks a common language and is sufficiently vague as to defy classification, necessitating, in many cases, an educated guess as to where the particular report fits. Since most of these evaluation-action programs are described within the context of a particular administrative structure however, the discussion which follows is organized in three broad categories: situation-specific process evaluation in the individual practice, situation-specific process evaluation in the agency setting, and situation-specific process evaluation in a regional system.

Situation-Specific Process Evaluation: Individual Practices

Not surprisingly, there is almost no literature available in this category. Until recently, the independent practitioner has been accountable only to an unsophisticated and not very knowledgeable payor and subject to the scrutiny of no one save his patient. Moreover, health practitioners have received little encouragement or training in systematic appraisal of their own practices. Caplan (86) has stated that independent practitioner audit is both desirable and achievable; the medical profession in particular is pioneering developments in this area.

Lyons and Payne's research study (313) of the degrees of relationships among thirteen diagnostic categories of office care performance of ninety-three primary care physicians is an early effort at situation-
specific performance evaluation in the individual practice setting. Using conformance to predetermined criteria as indicated in patient records, mean scores of individual physician performance within diagnostic categories were computed and correlated. An overall average inter-diagnostic correlation of plus .25 was obtained with clusters of correlations apparent. The investigators concluded that while there appears to be some homogeneity of performance among diagnostic categories, this is not enough to conclude that there are "high and low performers" or to preclude fruitful continuing education activity.

The Assessment of Performance project (ASPERF) of the American Society of Internal Medicine is an outstanding and unique example of effort on the part of practitioners collectively to provide individual physicians with a tool for assessing performance using situation-specific performance criteria (214,215). The basic format for the data collection instrument is the sequential series of steps used in clinical decision making: history and physical, problem identification, investigative management, diagnosis, treatment for short-term goals, continuing care, treatment for long-term goals, and contraindicated management.) The project staff has identified the patient problems and diagnoses most commonly seen by internists. For each of these clinical situations and for each step in the clinical decision format, experts have identified performance criteria critical to that problem. Each criteria set is coded by patient problem and by using the International Classification of Diseases, Adapted (ICDA) or its hospital counterpart (HICDA).

The ASPERF process is initiated by the physician who wishes evaluation of his practice; staff survey his records to see if they are sufficiently complete to use for evaluation and to see if his practice
profile indicates sufficient criteria sets are available to sample that practice. If these requirements are met, records are surveyed by ASPERF staff using the established criteria sets. The practitioner is then given feedback about his performance, guidance in planning corrective or enriching continuing education, and an appointment for reassessment. ASPERF is as yet unproved as a professional development program. Nevertheless, it is promising self-help aid for the internist and a model which other professional groups could profitably examine.

There have been two studies which could contribute to evaluation in the individual practice setting. One district of the American Psychiatric Association in California surveyed the practices of its members to establish treatment norms locally by diagnosis, age, sex etc. (290). The investigators feel that records-based evaluation using regional norms allows for confidentiality to be maintained and permits the necessary range of treatment modalities to be included. The project seems more destined to provide data for claims review than for sound appraisal of practice however.

A more promising project is that undertaken by a consortium of eight American medical organizations to select and validate performance criteria for health supervision of children in four age groups and for ambulatory management of urinary tract infection, bronchial asthma, and tonsillopharyngitis. The several detailed reports of this ambitious project (327,373,499) are worth pursuing because of the lessons to be learned from the methodology used. After establishing an initial list of performance criteria, a national panel of three hundred eighty-eight expert academicians and practitioners evaluated the items' relevance to outcome and acceptability for peer and self-assessment. The same criter-
ia were then tested in a broad spectrum of practice settings. Some of the important lessons learned were that many of the criteria selected could not be supported by citing research, that physicians consistently overestimated the amount of documentation present in patient records, that an elaborate validation process did little to improve criteria, and that there was a high level of agreement between academicians and practitioners.

Situation-Specific Process Evaluation: Agency Settings

Many general descriptions or interpretations of agency-based "quality assurance" are available in the literature. Most are simply enthusiastic re-explanations of the same process, occasionally with some embellishment.

Davidson's interpretation (118) is typical in its treatment of the subject. He recommends a five step process of patient care appraisal in which an agency committee:

1. chooses a study topic.
2. develops process criteria, i.e. desirable actions the practitioner should take when faced with the problem.
3. measures the degree of compliance with the criteria in a sample of cases.
4. analyzes results, then plans and takes action to increase compliance if this is not acceptable.
5. measures degree of compliance again to see if the desired change actually occurred.

A large scale effort to encourage adoption of this process in Washington State was mounted in the early seventies as a continuing education project which involved cooperative effort by several organizations.
in a multi-faceted program involving written and audiovisual materials, workshops and consultations (117). It was hoped that the process would be both educational itself, especially during step 2, and would identify valid continuing education needs. Del Bueno (124) emphasizes the importance of considering in step 4 the cost implications of the various alternatives for action. Dohner (139) elaborates on step 4 by insisting that any action to correct deficits should at the least include feedback of results to all staff, provision for discussion among staff, and agreement on specific recommendations for action.

Slee (478) suggests that step 3 should be a screening device only and that step 4 will be most profitably accomplished if deviations from the desired behaviour are analyzed on a case by case basis. A time frame should be used and deadlines for correction of deficits imposed. He also provides a check list for administrators to use in order to determine whether their quality control systems are adequate. In a separate article (479) he points out that this process can be applied in four different types of review: claims review of individual cases to determine medically justifiable services, concurrent or utilization review of individual cases to determine justifiable services, retrospective chart review of homogenous groups of patients to compare typical service given with standards set by staff, and retrospective review of individual cases that deviate from established standards. He believes the purpose of these last two are primarily educational, while the first two are useful primarily for cost control.

The similarities in all these approaches stem from the position of the Joint Commission on Accreditation of Hospitals in promoting quality assurance through accreditation requirements. JCAH standards are
directed toward quality protection through retrospective medical audit, continuing medical education based on audit, utilization review and provision for the credentials committee to have access to audit findings (398).

In addition to general discussions about quality assurance using process criteria, there are numerous descriptions of the method as applied in the specific agency setting. Shaw (471) reports on two of thirteen demonstration projects. In one hospital dramatic changes in medical and allied therapy of stroke patients have occurred as a result of deficiencies uncovered by medical audit. In the other, less spectacular changes in documentation of treatment practices have resulted from quality assurance procedures. Holloway et al. (243) discuss using a modified nominal group/delphi technique to develop hospital levels-of-care criteria for use in utilization review. Tom et al. (503) report using head nurses to screen for deviations from utilization review criteria and paying physicians to review these. They claim that the system has decreased length of stay, provided educational benefit in that reviewers gain insight into how their colleagues are doing things, and given physicians an opportunity to discuss cases with a reviewer at no extra cost. Devit and Ironside (132), on the other hand, present two studies (mastectomy and cholecystectomy) in which no improvement in performance could be noted. They also suggest caution in interpreting changes in performance since many factors other than audit-based education may affect results of reevaluation.

Ethridge and Packard (160) describe in considerable detail an audit system which uses standard nursing care plans for one hundred diagnostic categories. Basic care flow sheets (work records) are the
primary source of documentation. Clerks count the number of times a nursing action was performed and calculate an index based on the number of times it should have been performed. If a minimally acceptable score is not met, those records not complying with the standards are extracted for review. All data can be summarized by nursing unit, by medical category, by care plan, by process item, etc. The authors claim use of this system facilitates audit, simplifies shift reports, reduces nurses notes, improves orientation for new staff, and provides a more consistent approach to nursing care.

Graydon (202) evaluated the care in five experimental and two control hospitals using five process criteria in five diagnostic categories for a year prior to instituting feedback, for a year of feedback to the experimental hospitals, and for six months thereafter. It was found that most behavior changes in both groups were transitory and the few persistent changes that occurred did not vary between the study group and control hospitals. His conclusion was that feedback in this instance did not improve performance. However, defendants of the quality assurance procedure will be quick to point out that step 4, careful analysis of deviation and use of appropriate corrective action, was largely ignored and these results were therefore quite predictable. There is support for this point of view. In discussing the application of process audit in a Veterans Administration Hospital and a rehabilitation hospital, Ashbaugh and McKean (31) conclude that the vast majority of deficiencies are associated with performance rather than knowledge problems, and that departmental decision is necessary to determine whether feedback alone or with some other form of corrective action will be necessary.

Richman and Pinsker (421) describe a simple audit procedure
which depends on manual extraction of data from records and a multi-
disciplinary discussion of the identified patterns of care. Recommendations arising from these conferences have been shown to improve the quality of care.

Agencies which provide ambulatory care have reported less activity than have hospitals. Thompkins (498) describes efforts to develop a system for auditing the activities of physician assistants using clinical algorithms (protocols which dictate a series of sequential actions, each step dependent on the outcome of the previous one). To date these have been developed for eleven easily managed primary care problems. Computer analysis of the completed protocol shows the physician assistant his errors in following the algorithm, and it has been demonstrated that performance improves after feedback. Other clinics report using these problem oriented protocols for the continuous monitoring of paramedics' performance also (5,284).

The University of Vermont Health Care Center, in a unique variation of the basic five step process quality assurance system, uses patient participation in the audit as a way of widening the data base beyond the traditional records review (57). The Harvard Community Health Plan, with a membership of four thousand, employs a totally computerized records system (40). The sixty physicians and forty-five nurse practitioners use the Computer Stored Records System to both enter and retrieve all desired patient data. Since standards of care by diagnosis and problem are also stored, the practitioner receives immediate information if his decision has deviated from agreed upon action. After installation of this immediate feedback system, practitioner behavior in management of beta hemolytic streptococcal infections changed in the desired direction.
tion. Randomized trials with controls are planned as other standards are developed. A computerized records system has also been shown to be useful for review and improvement of care in a community mental health clinic (501).

Diamond et al. (135) describe a peer review and feedback system in a decentralized community mental health center which, when linked to continuing medical education, influenced shifts from polypharmacy toward more appropriate use of specific psychotropic medications. Russo et al. (448) developed a system for rapid review of ambulatory patient records using sets of process criteria (about five criteria per condition) for twenty-six common conditions. Of three hundred thirty-seven records selected for review, 25% were not reviewable and, of the remaining 75%, virtually one-half were deficient in one or more criteria. Practitioner compliance (the target group was private physicians, house staff, medical students and nurse practitioners) could not be conclusively shown to improve during the five month study period. The only corrective action used was private feedback and discussion. Using this system it was possible for senior pediatricians and the record librarian to review one chart each minute. Skipper (447) reports that in criteria such as choice of medicine and use of specialist consultants, behavior improved when the peer group audited medical care in a family medicine clerkship.

Situation-Specific Process Evaluation: Regional Systems

Generally, regional review systems have their origins in third party payment organizations. Prior to the seventies both Medicare and Blue Cross and Blue Shield Plans pioneered claims review. These early efforts were aimed primarily at avoiding unnecessary services and not at quality control.
One exception to this early pattern of regional review was a study of physician behavior (382) which was initiated at the request of the Hawaii Medical Association in 1968 and completed in 1971. The study consisted of four parts: an episode of illness study in which physician panels developed diagnosis-specific "optimal" management criteria for pre, during, and post hospital care and weighted these for importance; an office care study in which criteria were chosen for their importance in management of ambulatory patients; a hospital organization questionnaire; and a continuing education study. The hospital and office study combined twenty-one diagnostic categories in twenty-two Hawaii hospitals and in the office practices of all involved physicians. The weighted criteria made possible an expression of scores as a physician performance index (PPI) in terms of per cent of criteria met. This data could be reported by physician, by hospital, by diagnostic category, and by criterion. Some of the more significant findings were that appropriate admissions were at a high rate, but lengths of stay were generally less appropriate; specialists caring for patients in their area of expertise scored better than other physicians, but this was not related to Board certification; numbers of years of practice did not influence scores; and after two years of educational efforts in four hospitals using eight diagnostic categories, a remeasure of PPI showed desirable, if somewhat uneven behavior change in some but not all diagnostic categories and hospitals. The investigators concluded that a regional approach to physician performance review is useful because one can take a more global approach to health problems.

It has been stated that a number of claims review organizations existed prior to the seventies. In addition to private and public
health insurers, groups of physicians formed foundations in an effort to demonstrate that physicians could offer quality medical care at a reasonable cost without intervention of non-medical parties. Many of these review organizations received financial assistance for development of evaluation systems when the United States government provided funds for Experimental Medical Care Review Organizations (EMCRO's) to develop working models for systematic and ongoing review of medical care (354). It was stipulated that each project would be based on explicit criteria and standards and would incorporate review findings in local continuing education programs. Twelve organizations were funded, including the Hawaii Medical Association and some of the existing foundations; some private insurance companies were included as cooperating organizations.

One of these foundations, the San Joaquin Foundation for Medical Care, reports that after monitoring the prescribing and dispensing patterns of physicians and pharmacists, a 12.1% drug expenditure savings was documented and changes in physician behavior were achieved (492). The primary mode of corrective action taken with deviant practitioners was personal letters and phone calls, claims denial being used with only the most recalcitrant.

Under EMCRO funding, this foundation also instituted a medical peer review back-up system for claims adjustment (74). For thirteen of fifteen procedures (six types of injections, six types of physician visits, and three lab tests) a statistically significant, though non-uniform, relationship was found between the per cent of billing claims adjusted and subsequent changes in practice patterns as measured by decreases in the monthly numbers of these services being claimed. The physician-serviced review procedure accounted for 15% of dollar amounts
of claims adjustments and more than paid for itself, but the bulk of dollars adjusted (85%) were administrative (clerical) review. The author of this report concluded that peer review may control quality more than cost. Also, since relationships between adjusted claims and changed practice behaviors were not perfect, there clearly are limits to which claims review will effect practice patterns.

Kaiser Permanente, as a comprehensive group health care foundation, is concerned with both ambulatory and hospital care. In Kaiser's review program, trained lay reviewers screen all claims against established diagnosis-specific performance criteria. Approximately 85% of all claims go through this screen and the remaining 15% are referred to a physician reviewer who may settle the claim or refer to a committee of physicians. Medical reviewers have access to a computer generated physician profile and patient history to assist in making decisions. Harrington, in discussing this procedure, states that more efficiency may be attained in letting small claims go through. (218).

The New Mexico peer review system, another EMCRO project which covers both inpatient and ambulatory care, has been described extensively (66). This study reached a number of significant conclusions about the effect of physician peer review, using process criteria, on the cost and quality of health care: 1. physician involvement in review had a major impact on quality of certain aspects of care but did not reduce costs; 2. net savings to the Medicaid program were brought about by the administrative activities of the fiscal intermediary in cooperation with the peer review organization; 3. peer review of ambulatory care based on computerized data from claims forms is feasible; 4. peer review of ambulatory care, in most cases, will increase the use of services and
thus increase costs; 5. cooperation of the fiscal intermediary and peer review organization suggested that savings from the former activity might defray the higher expenditures brought about by peer review efforts; 6. Medicaid saved money by reclassifying nursing home patients to lower levels of care, but this approach should be evaluated to see that humanness of care was not sacrificed; 7. length of stay approach to controlling hospital costs was ineffective; 8. a level of care approach might be tried for controlling hospital costs, providing it is found to be successful for nursing homes.

References to quality improvement in the New Mexico study are made primarily on the basis of appropriate use of injections, which improved dramatically during the study period. This was thought to be brought about first through educational methods and, where this failed, claims denial. Changed behavior towards more appropriate use of injections was also associated with group practice, with board certification and being a pediatrician, and with being a doctor of medicine (65). When one examines the criteria actually in use, the claims regarding improved quality of medical care in general seem somewhat overstated. This report (66) also reviews a number of other previous or concurrent cost control programs. The authors conclude that figures showing cost gains are based on tenuous reasoning and inadequate data and that therefore more work must be done before generalizations about the effects of these claims review programs can be made.

Another EMCRO project was developed by the Utah Professional Review Organization (UPRO). In fact, three separate systems (for hospital, neighborhood health center, and ambulatory care) of review are in operation. The ambulatory review program has received the most attention
in the literature (355,357,358,373). Process or desired patient management criteria are developed for procedures, for common diagnoses and for drug therapies. Specified contraindications and possible omissions are also stated. About nineteen thousand claims per week are screened with one thousand claims needing physician review. This is done within the context of a computer generated patient history. A summary sheet allows the reviewer to make one or more of a number of recommendations. Feedback to the deviant provider is graded from an inquiring letter to information to warning to denial of payment. Aggregate data about deviations from desired criteria are used by the Utah Academy for Continuing Medical Education. The project directors believe this approach permits objective and subjective judgements about physician competence. It also provides information about abuse of the payment system by patients. Although claims can be denied, the major thrust has been educational, and this has been successful. Conclusions have not been reached about cost effects. UPRO's hospital review program has also attracted some attention because it employs nurse reviewers to screen hospital admissions against pre-established criteria, decreasing the required amount of more expensive physician time (373). All final decisions which run counter to the admitting physician's wishes are made by a physician who discusses this with the admitting doctor; there is provision for appeal to a committee of peers.

Confusion as to whether claims review programs using process criteria control cost and/or quality is noticeable in the literature to date. In another article on this subject, Sayetta (454) questions on both methodological and interpretive grounds earlier findings and conclusions about the effectiveness of yet another EMCRO project, the Certified
Hospital Admission Program which was originated by the Medical Care Foundation of Sacramento, California. These reports and others challenging the findings of the EMCRO and similar projects raise serious questions as to whether a sound experiential base existed for the national PSRO legislation which followed the EMCRO projects.

In 1972 Congress amended the Social Security Act to provide for the establishment of Professional Standards Review Organizations (PSRO's). Based primarily on earlier experience with EMCRO projects, this law has influenced profoundly the rate and direction of developing performance evaluation systems in all of the professions, but none more markedly than American medicine. A number of authoritative (515,516,518) and interpretive (45,222,350,490) references exist to explain this law. Basically the law provides for the formation of regional organizations of physicians (PSRO's) which will review care paid through federally funded Medicare, Medicaid, or Maternal Child Health Programs to determine that the health service provided in short stay facilities was medically necessary, consistent with professionally recognized criteria of care, and could not have been provided more economically. PSRO's may also review ambulatory and extended care services. Each PSRO is expected to establish regional care criteria and recommend award or denial of payment. PSRO's must review health services provided by other health workers in consultation with those workers.

In addition to the concerns expressed regarding the effectiveness of the EMCRO's which served as prototypes for the PSRO legislation, the law has also been criticized on other grounds: it requires a physician-dominated review process, and therefore may perpetuate the practice of crisis-oriented health care; doctor-patient confidentiality is
threatened; physician reviewers may be subject to defamation suits; the review process encourages "cookbook" medicine; and its implementation requires an administrative structure that may be a bureaucratic nightmare. Which of these fears are imagined and which are real remain to be seen. Sullivan (490) reports considerable resistance to the law from American doctors, however it is quite clear that gradually physicians in most sections of the country have participated in the formation of PSRO's.

Many of the former EMCRO's have been designated PSRO's for their region. Frederick (170) describes in detail later operations of the New Mexico organization and the Sacramento Foundation, providing information on the roles and functions of personnel, operating costs, and estimates of savings. New organizations have also emerged. Miller et al. (340) describe the development, implementation and impact of the Ohio Utilization and Medical Care Assessment Program, which is a process-oriented review system for psychiatric services. Included are prospective, concurrent and retrospective review of quality, quantity and cost of inpatient care of all patients by diagnostic category, patient, physician, hospital district and state. Criteria underlying the computer based system were developed by consensus of practicing physicians. Effects of the program have been measurable: admission rates and lengths of stay have decreased, as has the rate of deficiencies in care identified. Quality of medical records has improved and patients' rights are now documented.

Restuccia and Holloway (413) report on a study in one hospital to determine whether the correct level of care had been prescribed by the physician. When the nurse coordinator of the study, using predetermined
criteria, judged a patient to be inappropriately located on a particular day, she also identified the appropriate location (skilled nursing facility, home health care, outpatient care, or no care) and the reason the patient remained in the hospital. Approximately 10% of almost two thousand days were judged to be inappropriate, and the most significant barriers to appropriate utilization were unavailability of appropriate health service and the attending physician's conservative management of the patient. Clearly there is a limit to the impact continuing education programs for physicians will have on the situation.

In a rather atypical regionally based review project which focused on quality of drug utilization, an inter-disciplinary team developed process criteria for ten single entity drugs commonly used in skilled nursing facilities (488). Utilization patterns in five facilities were charted, and the greatest variance was found with regard to the areas of drug monitoring and indications for use. Each facility received information about its own drug utilization pattern and deviation from the established standards. Despite this feedback, a subsequent survey showed no changes in utilization patterns. Knoben (278), in thoughtful overview of the current status of drug utilization review, points out that most systems are regionally based, concerned more with quantity than quality standards, and as yet have failed to become an integral part of the medical care review process.

In general, it seems that regionally based review systems are able to effect only moderate change in physician behavior, even when using feedback, other educational approaches, and a last resort denial of payment. It is perhaps fortunate that the PSRO legislation allows review procedures to be delegated to agencies to be carried out on an in-house
basis where the agency has the capacity to do so. Jack (252) and Nelson (355) both advocate in-house review in preference to external review systems, pointing out that relevance to individual practices, opportunity for participation in all phases of review activities, and greater speed in receiving feedback are factors which contribute to the greater success of in-house systems in changing practitioners' behaviors.

In completing this section on systematic assessment of situation-specific performance, it should be noted that as this approach to evaluation has become more widely used, source documents providing model sets of process criteria for different clinical problems have been produced. These exist for medicine (104,457), nursing (87,507), and podiatry (21). The origin of these materials has been both professional groups and private projects, and the process used to derive the criteria is not always made clear. These references would be useful starting points for any professional group wishing to develop patient care criteria, but are not intended for wholesale and unthinking adoption and use.

Unplanned Evaluation of Performance

The literature indicates that random or unplanned judgements as to the adequacy of professionals' performance tend to occur primarily within three jurisdictions: the place of work (usually a hospital), the profession itself, and the licensing authority. The regularity and consistency with which judgements are applied in these areas is such that most would agree with Derbyshire (128) who states with reference to
licensing and disciplining of physicians in the United States, "--- there is no system - there are so many variable laws and regulations ---".

The Place of Work

The most likely location for incompetent behavior to be noticed even when systematic assessment has not been undertaken is a place of work where practitioners are functioning in close contact with colleagues. Rosenberg (429), in discussing the deaths of twin physicians in New York from apparent barbituate addiction, describes the difficulties of hospitals in effectively policing inept physicians. The brothers' superiors had to keep weighing the individual's right to confidentiality and due process against the right of the public to be informed and protected. Although complaints were registered at one hospital for over a year before the men were dismissed, and subsequent to that for over two years at a second hospital before dismissal, it came as a surprise to all that the brothers were drug addicts.

Although hospital and other agencies have always been liable for the professional competence of employees under the doctrine of respondeat superior, the scrutiny of non-employee staff has traditionally been a formality and largely limited to instances where complaints have been made (341). This situation has changed rapidly since the courts have enunciated a corporate liability doctrine which makes hospitals also responsible for taking all reasonable steps to ensure that physicians who have staff privileges perform in an adequate manner (94, 191, 242, 359, 446, 483). As agencies are heavily dependent on professionals themselves to determine what behavior is and is not correct, responsibility for monitoring physicians with hospital privileges is delegated to medical staff,
which may also be held liable (430).

Ludham (312) advises that medical staff adopt a low key approach to discipline with first efforts directed at controlling and correcting undesirable behavior using the support of the physician's colleagues. Every effort should be made to avoid confrontation and frank discipline; referral to a state board for investigation should be a last resort. Mitchell (345) is perhaps somewhat more helpful. He outlines ten realistic and positive actions that any health agency should take in order to institute and to maintain an effective system for dealing with allegations that a medical staff member is giving poor care. He also advocates responding to problems with a graded series of actions beginning with education-feedback and culminating in restriction of privileges; he does not include referral to the licensing body. Other writers provide similar information about evaluation and control of medical staff, including consideration of the right of staff to due process and of the difficult area of physician-administrator relationships (44,69, 94,191,359,446).

The Profession

The literature indicates that the ability and willingness of health professionals to police themselves and their peers is at best variable. There is at least some evidence that professionals do not always act in the public interest when aware of unacceptable conduct among their peers (129). Wertheimer and Manasse (532) found that pharmacists who have records of violating drug regulations are different in no measurable way (socio-demographic variables, grades etc.) from their peers and were not in any way singled out, seen, or treated by other
pharmacists as different. They concluded that deviance in the matter of adherence to drug regulations is not only tolerated but condoned by the norm group. Jago (257), in discussing the traditional role that dental societies play in peer review, states that historically these societies have acted more as an arbitrator for fees and that peer review committees might more aptly be named "peer justification committees". Cohen (99) refers to the reluctance of doctors to create difficulties for their peers.

On the other hand, the literature of the seventies indicates that professional societies are beginning to take a more active and aggressive interest in evaluation and discipline of potentially incompetent practitioners. The California Nurses Association has since the early seventies organized Professional Performance Committees which, among other things, act on complaints of a practitioner's performance (81). Hospitals and medical society boards are now exploring ways of working together, rather than in isolation from each other (394).

American medicine is sufficiently concerned about "disabled" physicians that in 1975 the AMA sponsored a national conference on the issue (138). One novel approach directed toward helping the alcoholic, drug-addicted or mentally ill physician is reported by the New York Medical Society (177). A cadre of volunteer doctors respond to complaints about physicians by quietly investigating and offering help to the physician if he is in trouble. This does not preclude a complainant going to the state board, but as more physicians become aware of this informal and early assistance, they may use this to aid the physician informally instead of waiting for sufficient evidence to accumulate to justify formal disciplinary action. Similar projects have sprung up in other
states. Unfortunately, because of their very nature, no records exist to serve as evidence of these programs' effectiveness. Furthermore a practitioner being aided by one of these programs is not restrained in any way from continuing to practice.

Additional evidence of an awakening interest in professionals' responsibilities in effective evaluation of peers is provided by a resolution passed by the 1976 annual meeting of the American Medical Association which urged all peer review committees to refer to appropriate disciplinary bodies any physicians who do not meet "accepted professional standards" (11). Finally, it should be noted that in many Canadian provinces, a professional organization is the de facto licensing body. Activities of these and other licensing boards are discussed next.

The Licensing Authority

The public has, through the licensing system, provided at least one vehicle for evaluating a health professional whose behavior is questionable. All licensing laws have provision for revocation of license. This control is ultimately in the hands of the public, though more often than not administration of the law is delegated to professional organizations or to a government agency composed largely of appointed professionals.

The responsibility for ensuring competence of professionals is a complex issue, one with which Claude-Armand Sheppard dealt succinctly at a recent conference on regulation of the professions (472) when he stated:

"not only do they (those responsible) have to verify continuing professional physical and mental competence, but they must assist professionals in keeping up to date with developments in their fields, as well as deal
with the very real difficulties of professionals who face obsolescence due to one form or another of incompetence. Indeed, while the protection of the public must be of paramount concern, it would be unfair and unwise to subject professionals, --- to insecurity, arbitrary rejection and unemployment."

(One might add as a significant consideration the loss of investment made by the public in the training of skilled manpower). Sheppard goes on to say:

"Furthermore, inadequate knowledge, poor physical condition or mental difficulties do not necessarily incapacitate a professional or render him completely unfit. The consequences will depend on the nature of the profession, the character of an individual's practice and all the circumstances of the case."

Regulation of the health professional through licensing is concerned with professional competence, mental and physical competence, and adherence to an ethical code. Cohen (99) itemizes some of the many difficulties licensing boards incur in discharging their regulatory responsibilities: reluctance of practitioners to cause difficulties for their peers; lack of expertise of board members with regard to the process required for acceptable disciplinary action; a tendency of the courts of appeal to reverse board disciplinary decisions (often due to inept handling and violation of "due process" by boards); and licensing laws which lay down grounds for discipline but are inadequate because they are incomplete or so imprecise as to be useless. An extreme example of this is the medical practice law in Washington, D.C.; at present, only conviction of a felony or a crime or moral turpitude are grounds for loss of license in this jurisdiction (504). Golin adds to this list of difficulties the claim that state boards have inadequate budgets and are assigned to overworked and inexperienced legal counsel (192).

Many of these problems are also cited by Derbyshire (128), who probably is responsible for the most thorough research done to date on the nature and extent of professional discipline of American physicians.
Unfortunately no comparable published investigations in the other professions or in Canada were found during this search. Tables I and II summarize his findings as to types of and grounds for discipline of American physicians during the four year period.

| Types of Disciplinary Action Taken with American Physicians 1963 - 1967 |
|---------------------------------|-----|
| Probation                       | 375 |
| Revocation                       | 334 |
| Suspension                       | 161 |
| Reprimands                      | 68  |
| **TOTAL**                       | **938** |

Table I

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Narcotics</td>
<td>440</td>
</tr>
<tr>
<td>Mental Incompetence</td>
<td>94</td>
</tr>
<tr>
<td>Fraud and Deceit</td>
<td>74</td>
</tr>
<tr>
<td>Conviction of Felony</td>
<td>72</td>
</tr>
<tr>
<td>Abortions</td>
<td>71</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>41</td>
</tr>
<tr>
<td>Professional Conduct</td>
<td>68</td>
</tr>
<tr>
<td>Moral Turpitude</td>
<td>26</td>
</tr>
<tr>
<td>Gross Malpractice</td>
<td>7</td>
</tr>
<tr>
<td>Fraud and Application</td>
<td>6</td>
</tr>
<tr>
<td>Gross Immorality</td>
<td>3</td>
</tr>
<tr>
<td>Fee Splitting</td>
<td>1</td>
</tr>
<tr>
<td>Gross Misconduct</td>
<td>1</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>904</strong></td>
</tr>
</tbody>
</table>

Table II

It is difficult to know whether the situation is the same ten years later. Moreover as many licensing laws are unclear on incompetence or unprofessional conduct (which is sometimes taken to mean incompetence), it is difficult to know whether in the above instances incompetence was a co-existing factor and, if so, whether it was cause or effect.
What is known is that various actions to protect the public from deficient practitioners are now being initiated by both the professions and the public through a number of legislative reforms (241). In 1969 Florida enacted a "sick doctor" statute which permits the medical board to take action on a license for conditions which prove injurious or lead to inadequate performance (377). Specified are organic illness, mental or emotional disorders, drug and alcohol abuse, loss of motor skill, and deterioration through aging. The law has served as a model for other state laws enacted in the seventies. In 1975 this same medical practice act was amended to allow the board to discipline for immoral or unprofessional conduct, incompetence, negligence and wilful misconduct. The law includes a clause which permits acceptance of discipline by hospitals, medical staffs, PSRO's and professional societies as evidence of unprofessional conduct. A parallel law makes it mandatory for hospitals to report any disciplinary action taken within thirty days. The board has a wide range of actions at its disposal including medical and psychiatric examination, counselling, restriction of scope of practice, probation, prescription of supervision or education, and suspension, revocation or denial of license.

Virginia (203) New Mexico, and Arizona (508) are also among the states which have impaired physician statutes. Arizona's legislation requires physicians, medical societies and hospitals to report to the licensing board any information which appears to show a doctor may be incompetent, unfit for safe practice or guilty of unprofessional conduct (129,508). It would seem that legislation of this nature precludes the formation in these jurisdictions of the informal "helping networks" referred to earlier. A 1975 California statute not only directs
hospitals to report all privilege limitations and dismissals, but also provides immunity to the informant, and gives the medical licensing board expanded investigatory powers, control of its own staff and funds, and the power to suspend practice pending a disciplinary hearing (189). Illinois has established a separate (from licensing board) medical disciplinary board and given it considerable power as well (153).

New legislation in Quebec recognizes the judicial nature of the professional disciplinary process and it has attempted to increase its effectiveness by providing that a government appointed attorney will act as chairman of the disciplinary committee of each one of the licensed professions in the province (90,400). Ontario has used legislation to clarify, regularize, and strengthen the disciplinary process for five major health professions (229), and British Columbia has enacted changes in disciplinary procedures for registered nurses (408).

It is of interest that the majority of legislative activity reported is concerned with the discipline of physicians. As the central figure in the health team it is predictable that primary concern will be directed at this professional, however it is also fairly apparent that in the United States the present medical malpractice crisis figures heavily in the flurry of "crackdown" legislation (129,189).

In Canadian nursing there appears to be some activity aimed at devising regulations to enable the licensing or registering body to more effectively and equitably apply the disciplinary process. The College of Nurses of Ontario has established a set of "standards" which are rather general statements of nursing functions to be applied by the Registration, Complaints, and Disciplinary Committees of the College (102). It is assumed that the nurse is competent until found incompetent through
discipline; since one of the standards is "seek help and guidance when unable to perform ---", a registrant who does not attempt to perform beyond his ability would apparently not be at risk for discipline. The Order of Nurses of Quebec sets out a detailed description of ethical behaviors in twelve categories (370) and a conceptual plan for evaluation of nursing services and performance (371). The Registered Nurses' Association of British Columbia is undertaking to describe through a "Statement of Roles and Functions of Registered Nurses" behaviors subject to jurisdictional authority by the Association; additionally the Association has specifically defined incompetence, which is justification for discipline in that jurisdiction (33).

There were found no discussions of activity connected with professional discipline in the dental or pharmacy literature, although Pavone, a dental educator, suggests that licensing boards should be given the authority to order remedial education for deficient dentists (381). He then goes on to enumerate all of the difficulties this action would entail and concludes that this will be an unpopular task which should be shared by a consortium of boards, universities and schools and associations.

Summary and Comment

The public and health professionals collectively have dealt with the difficult issue of evaluating health practitioners' clinical behaviors by developing a variety of techniques and approaches. The literature concerned with these approaches has been discussed in this
chapter within three broad sections: systematic assessment of total performance, systematic assessment of situation-specific performance, and unplanned evaluation of performance.

Discussion of systematic assessment of total or a broad range of clinical behavior is found in the nursing literature and, to a notably lesser extent, in the literature of medicine* and pharmacy. It should be noted that nurses have largely been organized and recompensed for service as an employer group and the available literature is perhaps reflective of the long established ability and authority of an organized and knowledgeable payor to demand accountability for service. Taken as a whole, the literature indicates that evaluation approaches which look at a broad range of professional behaviors vary greatly in complexity; specificity of performance criteria and standards; and the extent to which they have been tested for reliability, validity and utility for formative evaluation. In particular this material reveals that traditional performance appraisal methods from business, industry, and education have been adapted to meet the requirements of the clinical environment. Most often a conceptual framework is devised to organize clinical performance into major categories: behavioral descriptors are attached to these; rating scales and checklists are used as measurement tools; and anecdotal notes are employed as supportive material. The paucity of information about the utility of these approaches to actually influence and improve competence may explain the more recent emergence of evaluation systems which

* Barro (41) reviews several studies which attempted to assess a wide range of physician performance prior to 1970 (Taylor, 1954; Peterson, 1956; Clute, 1963; Jungfer and Last, 1964). These were primarily research studies whose methods never found their way into common use, probably for reasons of expense, lack of authority to initiate such surveillance, and lack of acceptability to physicians and/or patients.
focus not on global conceptualizations of performance, but rather on highly specific performance criteria.

Literature dealing with evaluation of situation-specific clinical performance is much more recent and is found primarily in medical and, to a lesser extent, nursing literature. Borrowed from the field of industrial quality control, this approach involves identifying specific criteria to be met for specific clinical problems, measuring to see that they are met (most often using patient records, abstracts or payment claims as a data source), taking corrective action if desired behaviors were not present, and remeasuring to ascertain success of corrective action. Situation-specific performance evaluation is not often used in individual practices, although physician interest in this is growing. On the other hand, much activity of this nature is found in health care agencies, and a variety of corrective actions are employed: feedback, counselling, peer pressure, educational sessions, administrative changes, and removal of practice privileges. Much activity is undertaken at the regional level also, largely as a result of federal legislation in the United States, with the most common corrective actions being personal feedback, individual and group educational efforts, and denial of payment. There is evidence that situation-specific performance evaluation is particularly useful in facilitating and documenting change in professionals' clinical behaviors; that continuing education is only one of several major change strategies; that change strategies must be used selectively to be effective; and that while these review activities may increase quality of care, their contribution to cost control is debatable. There is some indication also that in-house review activities are more likely to bring about desired behavior change than external or
regional review. Two additional observations might be made: first, this approach measures highly visible technical skills and actions more likely to be documented in records than are health teaching or use of certain comfort measures; second, where criteria are established by clinical problem, this approach generally assumes that diagnostic data have been accurately collected. Thus it is apparent that a rather narrow range of clinical behaviors is being evaluated.

Finally, literature concerned with random or unplanned evaluation of professionals' clinical behaviors in response to complaints indicates that as a result of the corporate liability doctrine, agencies are increasingly systematizing evaluation of all practitioners. Additionally, the professions are using a variety of approaches to monitor and take appropriate action with questionably competent members. Finally the public has, through legislation and provision of resources, acted in support of a more effective and equitable application of the disciplinary process where professional incompetence is an issue.

It can be concluded from health sciences literature of the early and mid-seventies that there have been considerable interest in and development of methods to evaluate clinical performance of practitioners. Further work is required however, before more than tentative generalizations can be made about the usefulness of the various approaches. In addition this work must be viewed within the context of yet another issue: given that we can develop effective approaches for evaluating the process of giving health care, to what extent should we judge the practitioner by what he does rather than by the results he achieves?
Chapter IV

ASSESSING RESULTS OF CLINICAL PERFORMANCE: OUTCOME

Escalating health care costs have caused the public and professions to question the value and desirability of many health care practices, and by extension of this reasoning, to conclude that competence of practitioners should be judged by the results achieved, i.e. by whether client health improved. The assumption underlying outcomes evaluation is very simply that if desirable results were achieved, the care givers must have been competent. Some of the problems associated with this thinking, e.g. the fact that many factors not within the control of the practitioner influence health state, will be discussed in Chapter VI. Despite such problems, developments in outcomes based evaluation have made an important contribution to what is known about assessing competence of health professionals. The literature concerned with assessing results of clinical performance is organized in this chapter in two broad sections: general measures of health status and situation-specific outcomes evaluation.

General Measures of Health Status

In this search, there emerged few accounts of general or non-specific health state measures used for evaluation of health personnel. Most of those in existence have been constructed for research and designed to apply to target populations.
Bergner (48) describes development of the Sickness Impact Profile (SIP) which measures health status and is used as an indicator of quality of health care. Evidence of the validity of the SIP is found in the high correlations of the SIP scores with self-assessment of health status, clinician assessment of health status, and other measures of dysfunction. McDonald and Martini (319) report the use of a condensed version of the SIP as an outcome measure of primary care.

Zyzanski et al. (548) describe the revision of a previously developed self-administered scale of patient satisfaction with primary care given by physicians. The modification and rationale are provided, the complete scale is reproduced and scoring is explained so it can be used without further information. Using this and another article which is referenced, one could capitalize on this work to develop a similar satisfaction index for other health care personnel and other settings.

Janke (258) has developed an operational multi-dimensional model for health status measurement, the Relevant Health Status (RHS). Based on the premise that health status is a composite of bio-medical state, patient satisfaction and social function, the RHS tool provides an approach for collecting, organizing and expressing information about these variables for aggregates of patients in relation to the health care system objectives and environment. Janke claims that this approach to organization of data provides a common information base for planners, educators, administrators, providers and lay directors. Directions for application of the RHS model to several problem areas, including education and evaluation of health care participants, are provided.

Barro (41), in her review of physician performance measures, devotes a chapter to outcome approaches. She describes various influen-
tial writers' (Donabedian, Shapiro, Elinson, Williamson, Siegal) conceptualizations of outcome, all of which were arrived at non-empirically. She also discusses Sanazaro's and Williamson's 1968 effort to arrive at a classification of patient outcomes through analysis of critical incidents reported by internists; mentions three studies aimed at assessing patient dissatisfaction; overviews three studies which attempt to measure patient compliance; and summarizes attempts at measuring outcomes using diagnosis-specific physical data (Payne), general condition (Williamson), physical function (Katz), and social function (several). Much of the literature reviewed by Barro predates 1970.

**Situation-Specific Outcomes Evaluation**

Several writers provide comprehensive descriptions of assessment procedures which are based on desired outcomes for a highly specific patient population. In this context, patient outcome is variously defined as "the health status of the patient at suitable end point or stabilization point following a reasonably well-defined episode of care" (318), "an alteration in the health status of the consumer as an end result of the activities of professionals" (239), "effects or results of care" (544), and "end results to be achieved through activities performed by or under direction of the nurse" (262).

Such authorities as McClure (318), Williamson (535), Lang (288), Taylor (495), Zimmer (541, 542, 543, 544) and Hilger (239) agree on the essential steps in situation-specific outcomes assessment:
1. Select and define a patient population with a common problem. Hilger suggests that numerous frames of reference may be used: developmental stages, medical diagnoses, diagnostic or treatment procedures, or functional problems.

2. Identify a time frame, e.g. end of a critical period or inpatient phase, or specify several significant stages in the natural history of the condition.

3. Identify the most significant patient problems for this condition, e.g. flatulence, post-colostomy.

4. State expected health status in relation to this problem for each stage of the condition in the time frame, e.g. 0-5 days, flatulence x8; 5-8 days, x4; 8-10 days, 0.

5. Set an acceptable degree of goal achievement, e.g. 80% of patient population will meet expectations.

6. And as with process quality assurance approaches, measure actual health state of a sample or the whole population against the criteria and standard, investigate probable causes for unacceptable deviations from the standard, take corrective action and remeasure.

Direct measures of process or structure are not taken (535, 543), although both obviously must be considered in analysis of unacceptable outcomes. Zimmer (543,544) recommends specifying, but not measuring, the activities and resources most important, i.e. powerful and cost effective, to achievement of each desired outcome.

Health professionals wishing to undertake results-oriented evaluation of care will find some assistance in the form of directions or examples in the literature. Zimmer (541,542) gives guidelines for assessment of nursing service which include a description of the characteristics of outcome criteria, instruction for the formulation of sets of criteria, examples of application of the criteria for evaluation of delivered care, and a discussion about peer, inter-disciplinary, and consumer participation in establishing outcome criteria. Zimmer's treatment of this subject is concise, comprehensive and based on the experience of many nurses, however supplementary reading would be required.
Another relatively complete reference for the nursing profession is the American Nurses Association's *Guidelines for Review of Nursing Care at the Local Level* which discusses in some detail the processes which can be used to develop, edit and validate outcome criteria (514). Both Zimmer's manual (542) and the ANA guidelines (514) provide several sample sets of outcome criteria for common nursing problems. Other authors have proposed specific nursing outcome measures for evaluation of neurological patients (495); of respiratory status after abdominal, pelvic or thoracic surgery (262); of functional progress of patients with congestive heart failure (27,28), of health state after hysterectomy (35), and of progress of adolescents with anorexia nervosa (127).

Physicians will be especially interested in the second of a two volume report (34) which presents the results of a study to develop disease-specific short term outcome measures for assessing the quality of medical and/or surgical care delivered to patients with one of eight different conditions. For each condition there is a technical chapter which presents epidemiology or pre-surgical condition; available data on incidence and prevalence, the use of medical services and associated disability and mortality; operational definitions of the disease or indications for the surgical procedures in the patient population to be studied; an identification of factors not under the control of medical care that may influence outcomes from the disease or surgery regardless of the quality of care provided; a method for controlling the factors not affected by the quality of medical care; specific outcome standards that could be used to assess the quality of care given to patients with the condition or surgical procedure; discussion of an expert panel's judgment of relative importance and sensitivity of each outcome in assessing
quality of medical care; and estimates of the size of the sample frame necessary to identify a specific number of patients with a given condition. While the book is most useful for medical audit, all health professionals will be interested in the appended item pools for anxiety and depression, activity limitation, patient satisfaction, socio-demographic factors, belief in efficacy of care, and health perception.

Ryge and Snyder (450) surveyed dental literature and concluded that evaluation methods to date were either too complex for real world application or lacked specific criteria. They developed a rating system for assessing quality of dental restorations: "satisfactory", with sub-categories of "meets all standards" and "observe at next visit"; and "not acceptable", with sub-categories of "replace for prevention" and "replace statim". Specific criteria and corresponding key words were developed for each of the four categories with respect to three characteristics: surface and color, anatomic form, and marginal integrity. On trials, inter and intra-rater reliabilities were high. The authors believe that the combination of ratings and reason given for the ratings provides constructive information for individual corrective action, for evaluation of training programs, or for identifying specific needs for continuing education of dental personnel. The system could be expanded to other types of restorative work.

Advice on deriving practice-relevant outcome measures is offered in several articles. A number of authors (27,35,262,318) emphasize that practitioners, not academics or researchers, should play the primary role. McClure (318) mentions that a Delphi technique can be used with physicians to arrive at consensus about appropriate outcomes. Johnson (262), rather than rely on nurse judgement as to desirable respiratory
status after surgery, had nurses collect data on each of the four parameters to be measured by observing a large sample of optimally recovering patients.

Another important consideration for those undertaking situation-specific outcomes assessment is the design and maintenance of clinical records. The traditional process oriented clinical record, which tends to provide more information about what the provider did than about the health state of the client, does not readily lend itself to an efficient system of outcome assessment. Wirtshafter and Messel (536) propose a modification to inpatient records so that a "minimum care assurance data set" collects and organizes information by diagnosis-specific outcome and related patient management, thus forcing the documentation of each.

Strategies for ensuring that outcomes based evaluation will have optimum impact have received considerable attention in the literature. McGuire (321) places great emphasis on rational selection of the topic which will be studied. She advocates an approach in which conditions responsible for the majority of the population's disability are graphed in terms of degree of disability at each age in three different ways: first, as they would exist without prevention or intervention; second, as they would exist given full application of present knowledge and resources; and third, as they currently exist in the target population. Small gaps between the first and second indicate targets for biomedical research and/or resource allocation; large gaps between the second and third indicate targets for organizational change and/or educational intervention.

Hirsch (240) believes that a diagnosis-specific outcomes approach is more definitive than is necessary for achieving most desir-
able changes for least investment. He proposes an evaluation of hospital care in terms of reasons for admission and objectives of management. In the preliminary study of five hundred consecutive discharges from five hospitals, it was found that adult patients could be classified in one of twenty-five reasons — for — admission categories. Objectives for hospitalization can be similarly defined within an assigned time frame. Each admitted patient is monitored against the expectations specified and, where a deviation is apparent, the record is reviewed. As this is a concurrent procedure, poor care is immediately corrected and practitioner learning is highly relevant.

McClure (318) recommends selecting for study patient conditions or problems which are frequent in occurrence, disabling, amenable to medical treatment, and probably not being adequately treated at present. He also emphasizes that several outcome evaluation studies should be administered in an overlapping sequence and that greatest benefits in terms of improved patient care will be achieved if many groups of problems are studied in preference to continuing to refine the few already studied: "there is too much sub-standard care to allow the luxury of refinement."

Finally, there are a limited number of references which consider particularly useful applications of outcome based evaluation. Several (27,35,240) cite the use of sequential outcome measures as a guide to concurrent practice; Aydelotte (35) refers to these specifications as "daily working material and part of the intellectual equipment of each nurse". Additionally such measures are used for patient teaching tools (27), establishing contracts with patients (127), identifying learning needs (27,85) and for specifying essential content for educational programs (85).
Gonella et al. (194) suggest that outcomes-based measures are good for comparative study of population groups. His research team established a scale to measure severity of health state for each significant stage in the natural history of a condition, collected data on proportions of patients at each level on the several scales, then made inferences about the quality of care prior to that point. This concept has been applied to eighteen different conditions. In the example reviewed (appendicitis), a significantly higher proportion of patients from "other government insurance schemes" had ruptured appendices at admission than did patients from Medicare or private insurers, thus indicating a problem with prior process for the first population group.

**Summary and Comment**

Outcomes based evaluation of health professionals is predicated on the supposition that if desired results in terms of improved client health are achieved then competence can be assumed, and that if desired results are not achieved, competence must be questioned. The literature dealing with outcomes-based assessment can be subdivided into general measures of health status and situation-specific outcomes evaluation.

General measures of health status have not been developed or used to any extent as an aid in evaluating competence of health personnel. Most of those in existence have been used only for research and are designed to apply to target populations, not to individual clients. To the extent that a practitioner, or a group of practitioners, is sole contributor to the care of that population, some tentative
generalizations might be made about the adequacy of performance. Because of the many factors affecting general health status however, it may be concluded that these measurement approaches probably have more value for health care services planning than for any effective contribution to controlling competence of health care professionals.

To control for the numerous variables affecting the client's health state, health care providers attempting evaluation of services have increasingly defined outcomes in terms of a few highly significant indicators relevant to the practitioner's action. The literature which deals with situation-specific outcomes evaluation reflects preoccupation with the steps of the evaluative process itself. Additionally, it is concerned with providing criterion examples and resource information and with maximizing benefits through wise selection of study topics and innovative use of the evaluation process. Incomplete patient care records have been identified as a barrier to effective use of systematic situation-specific outcomes assessment.

It is worth noting that literature dealing with an outcomes based approach to evaluation of health professionals is small in volume. Its primary liability (failure to account for other variables influencing health state) and its primary asset (the inherently appealing logic of accepting results as the final proof of competence/incompetence) have influenced the mushrooming development of evaluation approaches which measure and judge adequacy of resources, professional behavior, and health outcomes in varying combinations. Most of these efforts have at least some implications for assessing practitioners' competence.
ASSESSING COMBINATIONS OF STRUCTURE, 
PROCESS AND OUTCOME

It is possible to evaluate three broad parameters indicative of health professionals' competence: structure (skills and knowledge possessed by the practitioner), process (behavior exhibited in practice), or outcome (health state of the patient). In response to some of the advantages and limitations inherent in each of these (this topic will be discussed in Chapter VI), the health care industry has increasingly turned to the use of evaluation studies which focus on varying combinations of these three areas. This chapter discusses literature concerned with such combined evaluation approaches under the following broad headings: concurrent evaluation of varying structure-process-outcome combinations; outcome assessment with selective evaluation of process; and random evaluation of outcome and process of care. It should be noted that this review does not include studies which attempt to establish statistical correlations between combinations of structure, process, and outcome. These are more properly the concern of research in clinical or health care administration.

Concurrent Evaluation of S-P-O Combinations.

This section will consider literature dealing with each of several possible combinations of structure, process and outcome.
evaluation. Concurrent evaluation of all three parameters is considered first, followed by a discussion of structure-process combinations. The third topic is agency based process-outcome evaluation, and the fourth topic is regionally based process-outcome evaluation. Reports of structure-outcome evaluation activities were not found in the literature.

Concurrent Evaluation of Structure, Process and Outcome

A comprehensive and lucid discussion of the theoretical advantages of concurrent evaluation of structure, process and outcome is offered by Aydelotte (36). Her frame of reference, however, is health care service, not health professionals' competence. Thus her comments on structure apply to resources, equipment etc. as well as to knowledge and skills held by practitioners. She advocates ongoing evaluation of the pattern of care (staff performance, provision of resources, etc.) against the effects care is supposed to have; such a system would predict problems of patient groups, describe desired health outcomes for these patients, prescribe necessary practitioner activities to achieve such outcomes, and specify the resources (including knowledge and skills of practitioners) required for such activities. The health agency would work to ensure the presence of all of these elements in a continuing program of measurement and action. Like other authors, Aydelotte emphasizes the importance of both defining the patient population to control for variables grossly influencing health outcome and selecting outcome indicators which relate to the problem under study, are influenced by health care, and reflect patient health status at critical points of care including exit from the system. A manual (409) published by the Registered Nurses Association of British Columbia promotes structure-process-
outcome evaluation in an agency setting by providing directions for establishing an evaluation system, a rationale for these directions, sample criteria and forms, and further references on the subject.

There were found no accounts in the literature of concurrent evaluation of structure, process and outcome where the focus within the structure or resource area is that of practitioners' knowledge and skills. In fact, literature dealing with comprehensive evaluation of all three areas is surprisingly sparse even when the more common structural elements of equipment and facilities are considered. Ramey (403) in 1973 reported the development of such a tool for nursing service, however there is little attention to outcomes. Glasson (187), four years later, described a more situation-specific evaluation tool for nursing outcome and process and indicated that simultaneously the hospital's "quality control survey" measures for such universally desirable structural elements as clean environment, complete charts, etc. It is apparent that this is not an integrated evaluation system in which the structural elements chosen for measurement are deemed specifically necessary for nurse performance in relation to a given patient population. Selvaggi (466) described the experiences of a nursing service which conducted structure-process evaluation and process-outcome evaluation simultaneously. Again, it does not appear that the two activities were linked in any effective way.

Schonfeld (456), in discussing evaluation of dental care systems, proposed a framework which would encompass structure, process and outcome parameters. This article suggests evaluators must address four areas of concern: specific procedure, oral cavity, client, and group; and four dimensions of resources: technical, logistical, organizational and financial.
The Medicaid Vendor System (MVS) (434), a computer-based claims review system for the New York City Medicaid Program, is an imaginative use of combined structure-process-outcome evaluation on a regional basis. Described as an "eclectic approach to quality control," it clearly is aimed at locating the greatest numbers of, the most significant to health, and the most correctable deficiencies in medical care for least effort. Basically this system is oriented to process; the computer prints out exceptions to patterns of practice for the provider group or its specialty, and these are followed up by closer analysis of deviant individual provider claim profiles. The statistical profiles for provider groups are used also to identify aggregate deviations from acceptable standards. The latter indicates need for educational programs or for manipulation by changing incentives or the qualifications of providers able to claim reimbursement. This monitoring process is supplemented by assessment of structure (office inspections of deviant practitioners) and outcome (routine follow-up sampling of patient health). Priorities for corrective action are deviations from outcome. Within this group those practitioners with exceptional provider profiles are investigated first. The MVS is seen by the authors of this article as an optimally pragmatic approach to evaluation. An annual saving of almost $2 million is claimed, but this is not a net figure.

Structure-Process Combinations

The literature search yielded few accounts of structure-process evaluation. A study team under contract to the American Heart Association produced process criteria and related structural requirements for each stage of common conditions connected with heart disease, cancer and
stroke (171). An occupational health group has produced structure and process criteria for planning, administering, and implementing care by nurses in occupational health settings (484). Finally, an extensive project to develop a method for evaluating the structure and process of nursing care in acute general hospitals has been underway for some time and is discussed below in some detail.

The Methodology for Monitoring the Quality of Care, developed by the Medicus Corporation under a contract with the United States government, is an instrument which attempts to evaluate a wide range of nursing resources and behaviors (225,226,232,260). It is still under development. The project staff synthesized and expanded on criteria from existing process instruments such as QUALPACS, correcting omissions and eliminating redundancies and items without discriminatory power. Items were placed in a classification structure of six major categories (four related to nursing process, while the others assessed unit management and support services) and several sub-categories. The entire instrument was subjected at successive stages to rigorous statistical analysis to insure reliability, sensitivity, and internal consistency of the construct as a whole. The instrument structure is capable of expansion to incorporate additional items and/or factors unique to a given clinical situation. All criteria are designated by applicability to type of patient, i.e. self care, partial care, complete care, and intensive care (it was considered that items differed most in applicability by intensity of illness rather than by age, disease etc.). The items are also coded by type of unit, i.e. general medical-surgical, recovery room, newborn nursery; and by source of information, i.e. chart, patient interview, nurse observation etc.
The assessment procedure is initiated by generating an evaluation work sheet, from the master list of over two hundred twenty criteria, which contains a sample of items relevant to the type of patient and unit to be assessed. This selection process is such that any one work sheet covers the entire spectrum of possible items, but selects randomly from the major categories and sub-categories. This is thought to make it impossible for the nurses being evaluated to know which performance criteria will be assessed. With 10% of the unit's census being reviewed at random times, the process yields a computer calculated performance index (total score and by major category) each month.

The Medicus Quality Monitoring Methodology has been successfully applied in nineteen pilot hospitals of varying sizes and types and has required considerable effort to achieve an acceptable level of rater reliability. The method has been shown to reflect results of educational programs and changes in unit leadership (226) and is regarded as useful for inservice planning (172). Extensive use of the instrument as a tool for actually identifying performance problem areas and providing sufficient direction for corrective action has not been reported. Thus, although regarded by many as the ultimate in summative evaluation of group performance at the nursing unit level, the methodology has yet to prove its worth as a tool for group formative evaluation. The principal investigators indicate that it may be possible, through modification of the methodology, to provide performance profiles of individual practitioners suitable to peer review procedures (226). Attempts to validate certain items on this instrument against outcomes in patients with chronic heart failure and abdominal hysterectomy have been moderately successful (226,230,231).
Agency Based Process-Outcome Evaluation

A number of groups and individuals appear to have recognized the desirability of a combined process-outcome evaluative approach and to have developed a standard procedure, often supported by texts, forms, and workshops, to implement their chosen approach in an agency setting.

Davidson (119) offers an entire textbook of guidance for nurses and others responsible for in-house review activities which will meet PSRO expectations. Basic information about the characteristics and purpose of each type of review (admission certification, continued stay review, retrospective medical care evaluation studies and profile analysis) is given. Presentation of all subjects is supported with references, tables, charts, and sample forms. Similar references are available for establishing review systems in long term care facilities (251, 311).

The Commission on Professional and Hospital Activities (CPHA), a Michigan based data processing service for hospitals, has long provided information in a form which can be monitored to indicate adequacy of service (520). In 1972 the Medical Audit Study Worksheet was disseminated to assist hospitals in assessing their service patterns against regional norms for the most common diagnostic and treatment categories of patients; for example, normal appendices removed, antibiotics prescribed in out-patient departments, etc. were easily noted if deviant from regional practices. In 1974 the Commission incorporated emerging evaluation practices into a new system called Quality Assurance Monitor (QAM). The QAM is a computerized data retrieval system which allows a participating hospital to receive retrospective information about patterns of care for diagnostic and treatment groups accounting for about 50% of all
admissions against either regional norms or its own self-imposed standards. The information is organized for review as total hospital data and also by diagnosis, by surgical procedure and by medical service. The Commission provides advice, forms, and direction on how agencies can establish their own process and outcome criteria, systematically review actual performance, plan to take corrective action, and document the entire process for externally imposed or internal accountability requirements. The Commission emphasizes the importance both of developing only those criteria where corrective action will be taken (this avoids wandering into areas where research is needed) and of remeasuring after corrective action (this avoids pursuing fruitless corrective action).

Martin (329) studied the use of the CPHA system in one hospital for audit of care of six diagnostic groups. His report analyzes costs and technical problems, discusses potential extensions of the system and concludes that the chief problems lie not with data manipulation and display, but rather with data collection, validity and interpretation. Holloway et al. (244) compared the CPHA system to a manual system for collecting data not contained on the face sheets of medical records. The results indicated the CPHA system was less costly if more than 41% of hospitalized patients were reviewed; was as timely as the manual system; provided fewer data elements than physicians requested; and was less protective of human error. After evaluation of the relative importance of these results, the manual system was recommended for implementation. Experience with this system is also described by Chapin et al. (93). This two hundred and ninety-six bed hospital used the review program to direct continuing medical education activities, which usually took the form of informal group discussion of results rather than structured
education sessions. While the hospital was sufficiently satisfied with the system to expand its use to other areas, whether or not change in the desired direction actually occurred is not discussed.

The California Hospital and Medical Associations, supported by the California Nurses Association, have cooperated to develop their own approach to process-outcome evaluation of care (474). While designed to meet requirements of external groups such as the Joint Commission of Hospital Accreditation and PSRO, this California review system emphasizes the educational aspects of audit. The usual evaluation-action process is rather fully described in this article which usefully includes such pragmatic comments as "be sure you can get data on your criteria," and "group ratification of criteria (as opposed to mailing or posting for ratification) is valuable because it promotes learning, achieves better critique, allows a public commitment to be made". The author also considers the difficult step of problem analysis more fully than most, emphasizes the importance of fixing accountability and deadlines for corrective action, and reminds readers of the legal importance of keeping audit records separate from medical records and free of data which breach confidentiality. Hanna (217) describes the application of the California system in the nursing service of a two hundred bed community hospital. Again the approach to problem analysis is more fully discussed in this article than is the usual case.

In addition to the specifically developed and structured versions of process-outcome evaluation described above, a number of authors share experiences with this general approach. Gold (190) reports an unstructured approach to peer review by nurse clinicians using implicit
criteria and standards and concludes that establishment of explicit process and outcome criteria by practitioners in the specialty area is a critical requirement for successful peer review activity. Medicus Systems (79) provides direction to nurses for establishing process-outcome criteria, conducting a patient care evaluation study, and planning corrective action. Relationships with nursing care plans and problem oriented records are also illustrated.

Ethridge and Packard (160) describe in considerable detail the integration of three administrative tools of nursing service (audit, care plans, patient records) into a system for ongoing situation-specific evaluation of care. In some instances process and outcome are combined in one study, and in other cases the audit may concern itself with only process or outcome. Priority in analysis and correction is given to studies showing unacceptable outcomes and, within this group, those charts showing unacceptable process are attended first; thus the most clearly unacceptable situations receive attention even when resources are scarce. Shanahan (469) reports the use of process and outcome criteria to assess the care given to one patient group (acute myocardial infarction) by several different nursing units in the same hospital at different stages of illness. While the initial goal of this project, which was to determine whether nursing performance in one area effected patient outcomes in another area, was not realized, the committee concluded that conducting audits which measure nursing care provided to one group of patients at different stages of hospitalization has great potential, and that evaluation by unit or department alone places unnecessary limitations on efforts to improve care.

Pharmacists' services were assessed by Keys et al. (272) in an interesting study in which independent reviewers using implicit standards
considered adequacy of drug communications in terms of potential for benefitting patient and of actual impact on patient health. The method showed a high degree of inter-reviewer reliability and moderate correlations between process and outcome. The high inter-rater reliability was likely achieved because reviewers discussed the information abstracted from the chart before proceeding to an independent judgment of adequacy.

Medical care evaluation using combined process-outcome in a rural fifty bed hospital is described by Fitzgerald (165). This article would be of interest to small operations with limited resources; the hospital's system meets the requirements of external bodies as well as that of its board of directors for an accountability mechanism. Format for data display sheets and summary of audit results are illustrated. In a more complex approach to medical care evaluation, Rosenberg (432,433) proposes a format for criteria which overcomes the problems associated both with long lists of process criteria to cover any eventuality and with short term outcomes-only criteria which ignore long term results of care. His detailed format for disease-specific criteria includes critical diagnostic considerations, possible associated problems, patterns of treatment, expected response to treatment (short term outcome, complications, long term outcome), disposition management, and certain utilization criteria (indications for admission, discharge, and length of stay). The author acknowledges the method's drawbacks: the need for substantial resources (e.g. twelve minutes/chart is used to abstract necessary data) and the possible unacceptability of close regulation of process. The latter might not be a problem if criteria were set within the agency, but externally imposed criteria of this nature are not likely to be tolerated.
When one considers that most outcomes are affected by actions of many health care workers, it is surprising that the literature contains relatively few references to multi-disciplinary process-outcome audit. Davidson (119) refers to the need for multi-disciplinary involvement in audit, but the context for this reference is justification for nurse participation in medical audit. Dunham (147) also supports the concept and reports that one hospital, using the California system described earlier, employs multi-disciplinary review of outcome and provides for uni-disciplinary process audit either concurrently or subsequently. In another hospital of two hundred and twenty beds, medical, nursing and sometimes other staff separately established their own process and outcome criteria (communication about developments is provided for through cross-representation on committees), but then analysis of all results of all audits is conducted in a multi-disciplinary conference which yields joint staff recommendations (283). The author notes that this conference, which is video-taped, serves as an effective educational resource. Unfortunately the efficacy of this approach in terms of changed service patterns is not discussed.

In a rare comparison of two patient care (joint medical-nursing) audits for diabetic ketoacidosis, one in a teaching center and the other in a small community hospital, reviewers concluded that multi-disciplinary audit was effective and clinical input did not vary substantially with hospital difference, but modifications of criteria needed to be made according to local norms and institutional policy (101). Equally important is the authors' observation that while audit serves to identify deficiencies, it cannot dictate corrective responses; health professionals must correct problems by pursuing methods that reflect local needs
and available resources. Finally, the North Dakota State Hospital (185) reports multi-disciplinary audits of charts using a check list of process and outcome variables. Criteria and standards are implicit; review and follow-up action are handled using discussion procedures which the author believes to be an effective form of continuing education. This conclusion is not supported by evidence.

Regionally Based Process-Outcome Evaluation

Claims review mechanisms account for most regionally based process-outcome evaluation developments to date. The Utah Professional Review Organization (UPRO) has developed a system which provides diagnosis-specific concurrent review to ascertain propriety of hospital admission and to assign a norm-based length of stay (356). At the same time therapeutic management objectives, stated in the form of health outcomes, are established. Outcomes data are collected throughout the hospital stay and on discharge; identified deviations point to areas for analysis and corrective action. Preliminary experience shows a half-day reduction in length of stay and a slight decrease in admission rates at a cost of $9 per admission reviewed. Where analysis of audit results indicates educational needs, the information is referred to a state-wide medical continuing education foundation. The author notes, however, that the effectiveness of educational intervention has yet to demonstrated.

A California based PROSO prototype (204) has developed "criteria mapping" as a process-outcome evaluation technique which minimizes the problem of trying to apply a general list of diagnosis-related criteria to a specific patient. This method uses sequential judgments based on specific clinical data about the individual patient to assess quality,
thus leaving room for alternate medical decisions. It is expected that this method will more accurately reflect physician intent and that process, when measured by this approach, may correlate better with outcome.

Dental claims review has found combined process-outcome evaluation at the regional level an especially useful approach. DeJong and Dunning (121) report that twenty years experience in the United Kingdom has culminated in a system wherein practice patterns of dentists are monitored through comparison with norms for the practitioner group. Significant deviations are analyzed more carefully, and treated patients recalled so that completed treatment may be examined.

This general approach has been extended to prospective review of proposed dental therapy in many American dental insurance schemes (50,175,494). These mechanisms are characterized by:

1. Prior authorization of the plan of treatment based on submitted x-rays and examination findings.

2. Routine monitoring of quality of treatment (on all claims, all claims over a given dollar figure, or a sample of all claims) through examination of post-treatment x-rays, supplemented by direct examination of a sample of patients and/or computer analysis of practice profiles.

3. Follow-up on all complaints.

4. A graded system of decision making wherein the majority of review decisions are made by dental auxiliaries with provision for appeal first to staff dentists, then to dental consultants, and finally to a peer review committee. This is usually operated by the regional or state dental society.

There exists in the dental literature a number of proposals and general guidelines for the establishment and operation of these peer review bodies (39,61,372,449). There has been an increasing effort to specify firm and locally relevant treatment criteria (449,494), however most review programs have relied on already developed tools to evaluate
technical quality of care (174, 450).

Of particular interest as a reference in the evaluation of dental care is Friedman's *A Guide for Evaluation of Dental Care* (173), which prescribes evaluative strategies in considerable detail. Included are discussions of the important features to be considered in evaluating dental practice, process and outcome criteria for assessing technical quality of care, forms and a scoring system for direct and indirect measurement, and tables which classify the evaluative information one might seek in assessing overall dental programs. Friedman has also developed a dental care index (174) which establishes process and outcome criteria for evaluation of adequacy of dental services to a target population; as such this tool assesses the competence of the delivery system (one component of which is competence of a group of dentists).

Baillet et al. (38) report the development and testing of a system for dental assessment which considers history, physical exam, diagnosis, treatment plan, and treatment given. They used recognized dental practitioners to establish normative criteria. Both direct and indirect data collection by trained reviewers showed that high reliability was achieved for most items. Content validity was established and concurrent validity was satisfactory as evidenced by low, but significant, correlations with a subjective evaluation method. The authors give details on development as well as training and evaluation time, and conclude that this method is reasonable as an approach to large scale evaluation of dental care and to needs identification in continuing education.

Another approach to evaluation of dental care is that used by the United States Indian Health Service (2). Assessment of technical quality of service, quality of care (interpersonal relations, teaching
etc.) and quality of service to community is included. Within each area variables are chosen and criteria stated (e.g. contour approximates normal anatomy) and standards are specified (e.g. 85% of all restorations must meet criteria). The criteria tend to be a mixture of process and outcome and appear to have been selected on the basis of the ease with which data can be collected. Evaluation and follow-up counselling are carried out primarily in a supervisor-subordinate relationship. The Indian Health Service scheme is most notable for its very comprehensive approach to evaluating all aspects and levels of the delivery of dental service.

Outcome Assessment With Selective Evaluation of Process

While recognizing the importance of information about process if one is to correct inadequate outcomes of health care, there are a growing number of people who believe that the most cost effective procedure is to then evaluate first for outcome, and, if deficits are found, to assess the same sample against process criteria in order to locate the likely cause. This section will discuss experiences in outcome assessment with selective evaluation of process.

The originator of this general approach seems to be Williamson (534), who advocates a system of priorities for four types of evaluative action. Williamson believes that greatest impact on health for the population receiving care will be achieved by promoting correct and early diagnosis so as to increase the proportions of preventable or correctable conditions. Thus the first priority would be to evaluate for correct
diagnosis, both missed cases (or false negatives) and cases diagnosed for
the condition incorrectly (false positives); this of course means re-
diagnosing the population receiving care---a feat apparently not
attempted by health agencies despite Williamson's thesis. The second
priority would be to evaluate for inadequate outcomes of care; again,
concern for finding cost-effective ways to impact health state has
prompted Williamson to suggest that outcomes should be assessed using a
scale of functional health rather than diagnosis-specific outcome meas­
ures. His suggested scale is six levels: no impairment; measurable
impairment, asymptomatic; symptomatic, working; not working, ambulatory;
bedridden; dead. Third and fourth priorities would be evaluation of
process when degree of accuracy in diagnosis or degree of patient func­
tion in outcome is unacceptable.

Schroeder (458) describes one attempt to use Williamson's
approach in a health maintainance organization. Three patient popula­
tions were selected for study: depression, requirements for contracep­
tion, and hypertension. The investigation revealed widespread under-
diagnosis (44 - 74%) and inadequate therapeutic outcomes in two patient
groups. There was considerable difficulty with data collection. The
investigators chose contraception to follow through the analysis -
correction - reevaluation cycle. Despite several corrective activities
including education, program and personnel changes, no improvement was
documented on reassessment. It was concluded that further refinement of
this approach is required before its widespread use is feasible for
improving or controlling quality.

Williamson's strategy of evaluating outcomes then process has
had considerable influence in process-outcome evaluation approaches. The
MVS system (434) and one nursing service evaluation system (160), both described earlier, used this priority system for analysis of simultaneously collected data. More common is collection of outcomes data, and, when deviations occur, subsequent collection of process data. Nichols (363) discusses this as applied to nursing. She points out that the care plan with patient oriented objectives (outcomes) and nurse oriented plans (process) provides the evaluative framework. Outcomes should be assessed and, if criteria are not sufficiently met, process criteria should be measured to provide direction for corrective action. If the process criteria in this situation are found to be met, this is evidence that the plan of care itself is inadequate or other influencing variables are not controlled.

Gonella (193) builds on the concept of outcome then process evaluation by proposing a way of identifying the many process variables which may have contributed to inadequate outcomes. His twelve cell framework for outcome-process studies considers outcomes in terms of three dimensions (degree of prevention, degree of patient dysfunction, and degree of family and community dysfunction) and four categories of contributing process variables (physician performance, hospital performance, and influence of physical and social environment).

The Medical Audit (MA) portion of the American Hospital Association's Quality Assurance Program also uses Williamson's approach. As described by Brown and Sale (72) MA is characterized by an organizational structure which is controlled by medical staff reporting through the administration to trustees. The program establishes criteria for confirmation of diagnosis, management of therapy, and outcome. Outcomes screening is done first and deviations are submitted to process screen-
All deviations are studied and corrective action of an administrative, educational, or disciplinary nature may be taken. Ainsworth (6, 8), in discussing this program, emphasizes the importance of selecting study topics of high incidence or where outcomes presently show high cost or mortality.

With the possible exception of PSRO legislation, the greatest influencing factor in evaluation of health services appears to be the accreditation requirements of the Joint Commission on Accreditation of Hospitals (JCAH). Hartman (221) overviews the historical development of evaluative approaches prior to present day requirements. Porterfield (397) states that present JCAH standards require an audit system that is flexible, efficient, clinically sound, capable of documentation, and action oriented. The system must be based on measurement of explicit criteria, and it is recommended that the audit be a retrospective outcome audit (223). Rosenberg (431) emphasizes that an institution which complies with the JCAH audit requirements will likely also satisfy requirements for PSRO review. To assist agencies to meet these standards, the JCAH has developed a model performance evaluation procedure (PEP).

The PEP Primer (265) is a detailed handbook which describes actions to be taken to implement an effective patient care audit program. In addition to directions and supporting forms for an audit procedure, such related topics as data retrieval, civil liability, and records laws are discussed. The PEP methodology is essentially an outcomes based audit with built-in provision for selective evaluation of process in that patient care is scrutinized when complications occur. If unjustifiable variations are identified, analysis of cause is urged and the process audit is used at this point as needed.
Discussions of the PEP methodology as applied in hospitals throughout the country are numerous and yield little information not discussed in previous chapters. Knowlton (279) urges initiating PEP by choosing non-controversial topics, where it is unlikely that deficiencies will be identified, as a way of calming suspicious physicians. Ashbaugh and McKean (32) applaud PEP as a sound and efficient approach to quality control, but point out it should not be exclusively relied on to identify continuing education needs. Such education would be remedial without promoting new advances in medicine sufficiently. Their hospital supplements audit based continuing medical education with bi-weekly literature review meetings and/or guest speakers. Analytical discussions of criteria sets developed according to the PEP format are the focus of additional articles concerned with review of services to patients with acute and chronic otitis media (505); acute myocardial infarction (154); tonsillectomy with adenoidectomy (4); and pulmonary tuberculosis (3).

While the PEP Primer appears to have been written primarily for medical audit, nurses find the procedure readily adaptable to evaluation of nursing service (223,465). Diddie (137) supports nursing’s use of the disease focus in selecting a patient population because charts are easily retrieved and because this focus lays a base for future multi-disciplinary collaboration. She also emphasizes that criteria should originate with practitioners but be supported by recent literature. Reproduction of forms used by her hospital in the various audit procedure steps makes Diddie’s article a useful reference. Rinaldi and Rubin (422) state that analysis of deviations identified by PEP should involve the practitioners concerned and can use such readily available process and structure information as saved assignment sheets, daily census reports, and process
audits; follow-up action should include positive feedback and praise for excellence. The nursing literature also contains analysis of criteria sets based on the PEP format for acute myocardial infarction (468) and essential hypertension (470).

A pharmacy audit committee has applied the PEP methodology to a drug-specific audit (273). The group chose use of Warfarin in pulmonary embolus because of its potential for adverse effects, frequency of use and application to future disease-specific audits. They found on analysis that deviations were due to inadequate criteria, unacceptable professional practice of doctors and nurses, and poor documentation. Pharmacy shares with nursing the problem of retrieval of records focusing on a study topic other than disease or surgical procedure: it was found impossible to extract charts of all patients receiving warfarin.

In completing the review of literature concerned with evaluating combinations of structure, process and outcome, it should be noted that the medical profession in the United States has produced a number of model criteria sets which can be used as references by those planning medical care evaluation studies (10,13,14,15,530).

Random Evaluation of Outcome and Process of Care

Unplanned evaluation of process and outcome of care will occur when it is known or suspected that outcomes of care are poorer than they should have been. Under tort law, professionals may be held liable for such damages as injury, unnecessary suffering, or death resulting from professional negligence. For civil liability to be established, three
conditions must exist: damage, negligence, and a causal relationship (472). In this instance, the courts become evaluator and the standard for practitioner behavior is that of a reasonable prudent practitioner with the same experience and training (201,399). While formerly common practice in the community was accepted as the behavior of a reasonably prudent practitioner, this is no longer so. In 1968 a Massachusetts court ruled that a practitioner is required to follow a general standard of care applicable to all specialists in his field (94,242).

Tort law exists primarily to provide compensation for wrongs done to a person (95). It is difficult to know to what extent fear of litigation promotes careful competent practice, and the literature rarely considers this as a mechanism for evaluating or controlling competence of health care workers. Pritchard (399) outlines several problems associated with using tort law to ensure competence. The most important of these are high costs of defensive medicine (the certain result of a situation which combines the absence of competitive market pressures and the presence of increased threat of law suits); the freedom that the incompetent practitioner has to continue in practice until damage has been done and proved; and the necessity for the victim to recognize his injury, know it was the result of negligence, and depend on the testimony of professionals (whose malpractice insurance premiums will rise with every successful suit) to prove negligence. Coroner laws, which provide for investigation of unexpected and suspicious deaths, are limited compensation at best for this last defect (292).

If there are few who expect the civil liability system to control competence, there are a growing number who advocate aggressive monitoring of professionals' competence as the only way to control sky-
rocketing malpractice settlements and insurance premiums (22,182). There has been some hope expressed that the PSRO legislation, which monitors and influences certain practitioner behaviors while simultaneously providing immunity from liability to practitioners adhering to PSRO standards, will effectively decrease the size and number of malpractice settlements (545). Most agree, however, that while PSRO standards will supersede or supplement those traditionally applied in malpractice proceedings, other cause and effect relationships between the PSRO legislation and malpractice activity are less clear (64,115,277, 517,545).

The potential for information from malpractice settlements to contribute to more regular and systematic evaluation has been recognized by medicine and dentistry. A committee of orthopedic specialists and lawyers derived a list of avoidable adverse events by analyzing three hundred and fifty malpractice cases involving orthopedic surgery that had reached the appellate court between 1910 and 1972 (493). The events were classified and then used as a starting point from which to develop a taxonomy of avoidable adverse medical events. The author points out that such a list could be developed for any group of practitioners and from sources other than malpractice litigation (e.g. disciplinary hearings, hospital incident reports). Once the system is developed and weights are assigned to items, it could be used as a functional tool for monitoring and analyzing performance. Milgrom (339), who performed a similar procedure on all dental malpractice suits between 1900 and 1974, suggests that this approach allows the identification of practitioner behaviors that are clearly and causally tied to outcome and thus is a more valid evaluation approach. The fact that ineffective and wasteful practitioner behavior would go unnoticed is not recognized.
Summary and Comment

Theoretically, one can evaluate professionals' competence by looking at structure, process, or outcome elements, or at combinations of these.

The health care literature concerned with combined approaches recognizes the potential advantages of an integrated evaluation system which links structure, process, and outcome studies, but provides limited reports of its application. The most notable of these is the Medicaid Vendor System, a regional medical services monitoring scheme (434). Even the MVS, however, looks at structural elements in terms of the delivery system rather than in terms of the knowledge and skills possessed by professionals.

Reports of structure-process evaluation activity are equally limited, although the Methodology for Monitoring Quality of Nursing Care (225,226,232,260) is notable both for its comprehensiveness and the care with which the system to date has been developed.

Agency based process-outcome evaluation reports are numerous. A number of influential organizations have developed complete process-outcome evaluation systems which they promote through workshops, consultation services, and supporting materials replete with special forms and words. Perhaps because process-outcome evaluation is more recently popular and major measurement problems were worked through earlier, this area of literature treats analysis of deficiencies and choice of corrective action somewhat more fully than does the literature discussed in previous chapters. The literature describing experience with agency based process-outcome evaluative procedures is particularly rich in examples of
variations to the basic approach; these range from simple to complex systems employing small to large dollar resources. Reports of the several professions jointly monitoring their services appear for the first time with medical-nursing combinations accounting for the majority.

Literature concerned with regionally based process-outcome evaluation yields information about one medical care evaluation program with a corresponding regional continuing education delivery system which could not be shown to be effective. Another regional medical care review program has developed a measurement system, "criteria mapping", which allows a choice of process and/or outcome criteria depending on developments in the patient's condition during the course of his illness. Descriptions of regional process-outcome evaluation of dental care are particularly impressive in terms of scope, precise measurement practices and involvement of the many parties concerned (patients, insurors, professionals).

Another literature area considered in this chapter is outcome assessment with selective evaluation of process. In this approach process evaluation usually occurs after outcomes have been found deviant from the established standard. This approach is widely used, apparently as a result of promotion of the PEP evaluation methodology developed by the powerful Joint Commission on Accreditation of Hospitals. Williamson's work (534) on priorities for evaluation is an even earlier influencing factor, however, and six years later this remains a useful and important item for study by those concerned with process-outcome evaluation.

Random evaluation of alleged negligence resulting in adverse outcomes occurs in connection with malpractice litigation. The extent to
which tort law influences or controls health professionals' competence is unknown. Because malpractice action involves few practitioners and is after the fact, it has little direct impact in terms of evaluating health professionals' practice. The threat of malpractice action, however, is thought to act as a deterrent to incompetent practice.

As structure, process, outcome, and combined evaluation approaches have, in succession, come into widespread use, the experience has clarified a number of conceptual issues of a technical and contextual nature which are related to monitoring and/or promoting professional's competence. In some cases research has been part of this process of growth and understanding. The next two chapters will discuss the more prominent of these issues.
Chapter VI

SELECTION OF EVALUATION AND ACTION STRATEGIES

Recent experience has clarified a number of conceptual issues related to evaluation and action strategies in health care. Reports and discussions of early efforts are beginning to permit both identification of the specific decisions which must be made and generalizations about the implications that at least some of these decisions have. This chapter considers this information in the following sections: General Considerations in Choosing Evaluative Strategies, Deciding What Will Be Evaluated, Data Management, and Evaluation Based Interventions.

General Considerations in Choosing Evaluative Strategies

Writers agree that there are some universal attributes that one will attempt to achieve in any approach to evaluation in health care. These include validity (125,435), reliability (125,435), sufficient sensitivity to allow analysis which will lead to action (125,435,506), capacity for comprehensiveness (12,125), and practicality (12,125,435, 506). Such approaches will also provide for full documentation (12,125), for ratification of criteria and standards by providers whose care is being evaluated (12,125), and for standards which are sufficiently flexible to permit innovation and clinical advancement (125). In addition to
being mindful of these generally desirable characteristics, evaluation planners concerned with competence of health professionals will wish to have a well developed understanding of the evaluation-action process as applied in health care, possible frames of reference for evaluation, and the role of values in evaluation.

The Evaluation-Action Process as Applied in Health Care

Fundamentally evaluation is no more than establishing what will be measured, measuring and recording results, and making a judgement as to whether these are adequate or acceptable. Sometimes the judgement about adequacy is built into pre-established standards. In health care, evaluation most often is undertaken to provide information on which to base further action. Hence most discussions concerned with competence of health professionals begin with the evaluative process but then extend to further steps allowing for analysis of evaluative results, appropriate follow-up action, and re-evaluation to determine effectiveness of action. This evaluation-action sequence has been the organizing influence throughout this paper.

Several multi-step evaluation-action sequences have been described in previous chapters and will not be repeated here. Planners wishing to review the evaluation-action process in greater detail have several references from which to draw. Brown's bi-cycle concept (70,71) is remarkable not only because it is one of the first conceptualizations of the process applied to competence of professionals, but also because its twelve steps leave little margin for oversight. Measuring the Quality of Patient Care (255), the Ambulatory Care Evaluation Primer (335), the Quality Assurance Program for Medical Care in the Hospital (16), Quality Assurance in Long Term Care Facilities (311), the PEP
Primer (for hospitals) (265), Guidelines for Review of Nursing Care at the Local Level (514), and "Measuring Quality of Nursing Care: Parts I and II" (306) all are comprehensive references on evaluation in health care that provide the reader with a well developed explanation of steps in the evaluation-action sequence.

As planners identify the discreet steps they will undertake, the probable costs begin to assume more concrete form. The literature offers little information on this important matter. Zimmer (543) estimates that 1 - 2% agency manpower effort should be expended on quality assurance activity, and McClure (318) believes that 1 - 2% of a hospital's gross income can profitably be spent on such activity. Williamson (535) estimates that a full-time medical records analyst, 10% of a full-time equivalent physician position, and volunteered physician time are required for medical outcomes evaluation studies. Specific information on costs of medical and nursing audit programs and utilization review is provided by three hospitals with limited explanations of the methods used to arrive at these figures (207, 376, 521). In 1976, Medicus Systems estimated costs of its Methodology for Monitoring Quality of Nursing Care at $1,000 initial and $150/month operating cost. This assumed training of observers could be absorbed and only minor modifications to the existing computer program were required (226). Experience has shown this estimate to be low (197). A number of existing evaluative strategies are discussed by Costanzo and Vertinski (113) in terms of whether they require limited or extensive resources for implementation. Costs associated with the large scale (national) development of situation-specific criteria have been carefully documented by the consortium of professional organizations which pioneered a methodology for specifying criteria for
evaluation in ambulatory child health care (374). It is clear that with so little rule-of-thumb information on costs, the planner will wish to project anticipated costs carefully when selecting evaluation-action strategies.

Possible Frames of Reference for Evaluation

The frame of reference for this paper has been the continuing competence of health professional in practice. No attempt has been made to distinguish evaluative approaches which make judgements about individual professionals from those which make judgements about groups of professionals; the former will be desirable in some instances. Given this frame of reference, the process of care (clinical performance) is taken to be a direct measure of competence with structure (knowledge and skills) and outcome (results of care) remaining as indirect measures. This interpretation is supported by Hegyvary and Haussmann (232) and Donabedian (141).

Given other frames of reference, however, structure or outcome, and individuals, groups or whole populations might be more important. In practice, evaluation programs may attempt to serve more than one purpose, therefore it is important to have a way of conceptualizing many different frames of reference. Donabedian (141,143) has suggested that a useful view can be obtained by recognizing that one will usually wish to focus on a part of at least three dimensions:

1. individual practitioner - groups of providers - care systems
2. individual patient - caseload - target population
3. physical function - psychological function - social function
Each of these three-part dimensions can be thought of as the dimensions of a cube with the possibility of intersection at anyone of a number of points. If one is concerned with the physical health of the individual client as affected by an individual practitioner, one will choose different criteria than would be the case if the focus of concern were psychosocial health of the caseload of a group of providers. As the frame of reference changes, criteria identified as desirable may conflict. For example, the nurse may elect to give superior care to an individual patient at the expense of her caseload; the physical needs of a target population will be served by careful reporting of venereal disease contacts at the risk of the individual's psychological or social functioning. In selecting evaluation strategies, it is important to clarify the frame of reference and purpose; if there is more than one frame of reference to be used or purpose to be served, one must identify possible areas of conflict and make anticipatory decisions.

The subject "competence of health professionals" basically is concerned with the health worker as he acts to support and increase the physical, psychological and sometimes social function of patients for whom he is responsible. This frame of reference can and should be expanded to include groups of professionals if action strategies can be effectively planned for groups. It is therefore important to establish exactly what behaviors might be included for evaluation. The literature indicates several approaches have been used:

1. description and classification of behaviors that occur in practice, through use of experience (144), goal analysis, (103) or critical incident technique (166, 291, 455).

2. multidimensional conceptual schemes. One (76, 106) used the dimensions of content (developed through a list of most important health
problems), abilities, and tasks; a second (42) chose the problem solving process, clinical discipline (pediatrics, psychiatry, etc.), and context of care (acute, rehabilitation etc.).

3. health needs and problems which give rise to required behaviors (524).

4. literature review (219,232).

The chief value of a conceptual framework for evaluation of performance is that it allows one to identify the universe which can be evaluated. Once criteria are set, such a framework also provides a reminder about areas of performance not chosen for evaluation. In view of the current preoccupation with evaluating technical decisions about physical care, this is an important point.

The Role of Values in Evaluation

De Geyndt (120) states that "to define quality is to limit its meaning; --- (this) reflects a value judgement". Although Donabedian acknowledged values as a major influence in selecting evaluation strategies better than ten years ago (142), it is only since 1975 that the issue has been fully acknowledged in the literature.

Lang (288,289) explicitly states that values clarification is a distinct step in the process of evaluation in nursing care. Truby (506) alludes to the importance of values when she advocates consideration of the agency's philosophy when making evaluation plans. She extends this further by saying that evaluation planners should be aware of exactly where they are willing to change and should be guided by possible decisions that will result and a sense of data needed. Moore (346) explores four common value systems which have an important influence in North American health care systems: survival, rationality, hedonism, and the Judeo-Christian ethic. Somers (482) identifies expectations which
the public holds for the health care system: technical competence; a good doctor-patient relationship; institutions and procedures which will monitor competence, and corrective action which will be taken as necessary.

Methods for values clarification and specification are beyond the scope of this paper. Nevertheless it is apparent that thoughtful writers support the notion that explicit statements of values help avoid conflict and give direction to planners as they make decisions about evaluative strategies.

Deciding What Will be Evaluated

Given an established frame of reference, a way of identifying the possible range of clinical behaviors, and a clear sense of values, the planner will need to make further decisions about what to evaluate, whether to establish explicit criteria and standards or rely on the considered judgement of reviewers (implicit standards), and about how to select criteria and standards if these are to be specified. Literature concerned with these questions is reviewed in this section.

The Structure - Process - Outcome Debate

As conceptualized by Donabedian, quality of care assessment relies on one or more basic types of information: structure, process or outcome. Earlier chapters have shown that the bulk of evaluative activity in the first half of the century was directed at measuring professional's knowledge and skills (structure); that the fifties and sixties
saw an increased emphasis on evaluation of performance (process), and that in the seventies, after a brief flirtation with outcomes evaluation, there emerged a decided preference for combined process–outcome studies.

In selecting one or more of these basic types of information for evaluation, the advantages and disadvantages associated with each must be considered. These have been discussed extensively in the literature and are summarized and referenced in Tables III, IV, and V.

### Advantages and Disadvantages of Measurement of Structure

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>generally reliable (335)</td>
<td>questionable validity re impact on patient health (62, 67, 142, 335)</td>
</tr>
<tr>
<td>easy, relatively inexpensive, to carry out (62)</td>
<td></td>
</tr>
</tbody>
</table>

Table III

### Advantages and Disadvantages of Measurement of Process

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>when a process criterion is not met, planners have direct information about what requires correction (54, 120, 335, 452)</td>
<td>questionable causal validity re improved patient health state (7, 67, 133, 145, 163, 233, 435)</td>
</tr>
<tr>
<td>corrective action can focus on individuals or on groups of providers, whichever is efficient and appropriate (233, 335)</td>
<td>tends to perpetuate a sometimes indefensible status quo, with such results as rising health costs etc (145, 318, 335)</td>
</tr>
<tr>
<td>familiar, ie, providers find it easy to establish criteria (62, 335)</td>
<td>may increase malpractice litigation rate (7)</td>
</tr>
<tr>
<td>relatively easy to carryout (54)</td>
<td></td>
</tr>
<tr>
<td>permits concurrent review and feedback of results at a time when it is most meaningful (7, 62)</td>
<td></td>
</tr>
<tr>
<td>able to consider problem areas of missed and erroneous diagnoses (534)</td>
<td></td>
</tr>
</tbody>
</table>

Table IV
Advantages and Disadvantages of Measurement of Outcome

Advantages -
- high degree of validity (318), provided there is adequate control for other influencing variables (226, 332).
- allows for consideration for all possible causative factors, thus will lead to most improvement for least cost (163, 318, 543).
- capable of focussing on a target population, not just a caseload (335).
- stimulates innovation in care (335).
- aids in decisions about health policy (120, 67).

Disadvantages -
- identifies problem but does not aid in defining cause (54, 62, 142, 145, 233, 335, 452).
- difficult control for variables other than case or to identify provider(s) responsible (54, 233, 435).
- little available information or agreement on what constitutes reasonable outcome measures (62, 67, 142, 145, 335).
- difficult to develop reliable outcome measures for some problems (54, 233, 335, 452).
- difficult to collect data, especially on long term outcomes (67, 142, 145, 335, 452).
- when applied within a caseload or facility, assumes patient assessment and diagnosis has been correct (159).

Table V

Most evaluation studies in health care share the common assumption that adequate resources contribute to adequate diagnosis and treatment which in turn result in a better health status than would otherwise be the case. In selecting structure and process variables and criteria, the planner will attempt to choose those which have a high positive correlation with favorable outcomes. One immediate problem, as indicated by the tables, is that few such correlations are known or can be predicted with certainty. Establishing relationships between elements of structure, process and outcome is the business of clinical research (62, 233, 452). The demands in terms of precise measurement are greater in such
Clinical research in the technical aspects of medical care has a long tradition and a growing body of literature. The other professions are beginning to develop such a knowledge base as well. Less well developed are studies which attempt to establish correlations between structure, process and outcome for non-technical areas or for care delivery systems. Several items retrieved as reports of evaluation activity proved to be reports of clinical trials with the major purpose being that of establishing correlations. Most (149, 205, 231, 249, 337, 382, 415, 416, 459, 489, 490) were structure-process studies whose results offer little comfort to those wishing to use structure variables as proxy measures of the adequacy of care. One study (149) particularly germane to competence evaluation showed no relationship between nurses' scores on a knowledge-theory test and the observed use of related principles in the performance of nursing practice. Process-outcome studies (186, 253, 364) and two studies incorporating variables from all three areas (226, 428) were also identified. Brook et al. (67), in a monograph on outcome assessment, summarize forty-seven evaluation studies, some of which establish correlations as well. It should be noted that a pre-requisite of research to establish correlations is the availability of reliable measurement methods. A further difficulty in establishing process-outcome correlations in health care is that once a process becomes conventional professional practice it becomes practically and ethically difficult to submit to clinical trial (62).

Given the current problems of evaluation studies using structure, process or outcome measures in isolation, the tendency to take
concurrent measures in at least process and outcome areas is understandable. Pre-eminence may be given to one of these areas, depending on the frame of reference. Because the object of evaluation is an understanding of whether desirable results can be encouraged and undesirable results prevented, it is likely that such combined measures will always occupy the attention of evaluators seeking improvements in the delivery of health care generally or the performance of health professionals specifically (141).

Implicit Versus Explicit Standards

Having made the decision to evaluate structure, process, outcome or combinations of these, the evaluator can specify the standard by which care will be judged adequate (explicit standards) or he can simply make a judgement based on some internalized standard (implicit standards).

Krumme (286) states categorically that measurement and evaluation approaches using implicit standards are highly unreliable. She claims that scoring systems which rate on a comparative basis (poor, average, good,) against performance of a reference group, or according to degree of guidance required, permit an imprecise measure which depends on each evaluator's intrinsic standards. These will vary between reviewers (inter-reviewer reliability) and in the same reviewer over time (intra-reviewer reliability). Richardson's study (420) of judgements of the quality of clinical care in retrospective analysis of charts in obstetrics, pediatrics, and surgery shows that inter-reviewer reliability using implicit criteria and standards was unacceptably low. Decisions were influenced by clinical bias, missed items, geographic differences in
standards, and reluctance to criticize. He concluded that the number of
independent judges required to reach a stable judgement of care quality
exceeded the number logistically available to meet review demands, and
that pre-established criteria and standards should be used. In a subse­
quent study (419), he demonstrated that where explicit process criteria
for a given condition become too numerous for efficient review, it is
possible to reduce these to manageable numbers. His strategy involves
rating a sample of charts using explicit criteria (in this demonstration
there were sixty-one), rating the same charts through a subjective review
process, then subjecting this data to three consecutive analytical
methods. In this case it was shown that evaluation of twenty-six cri­
teria would produce virtually the same results.

Other researchers are not so willing to abandon implicit stand­
ards as undesirable or impractical. Peters (388) showed that judgements
undirected by instruction and unmodified by consultation are sufficiently
accurate if a selective review strategy is used. By submitting only
records previously rated as unsatisfactory to subsequent review, and
including them with previously unrated records so that reviewers won't
know they have been identified as unsatisfactory, efficient use of
reviewer time is achieved. The proportion of unsatisfactory records in
any sample should approximate 50%. Peters showed that, assuming 15% of
the original sample is unsatisfactory if a record is identified as
inadequate three times it has a 99% probability of being unsatisfactory.
The net result is that some unsatisfactory, records escape but the worst
are identified at a low investment of time.

McClain (317) has explored the use of confidence scales to
increase reliability where implicit standards are used. He found the
scale of little use in increasing inter-reviewer reliability and concluded that this is best improved through discussion and instruction. Intra-reviewer reliability can be achieved by using reviewers who have full confidence in their judgements and who use the confidence scale as a function of their trust in the data presented for review. Techniques for identifying reviewers who use a confidence scale in this manner are discussed.

Brooks (63) has shown that even when reliability of method is not at issue, acceptability of care given two hundred ninety-six patients with four different conditions varied from 1.4% to 63.2%, depending on whether process or outcome and explicit or implicit criteria and standards were used. Judgements about the process of care using explicit standards yielded the lowest percentage of acceptable records. Another study, however, found little variation in quality judgements between explicit and implicit standards review (365). Variations likely depend as much on the nature of the explicit criteria (most of the above studies were conducted using rather long lists of explicit criteria of doubtful validity) as on the fact that criteria and standards are explicit or implicit. Research using shorter lists of highly critical explicit criteria might well show higher agreement with an implicit standards approach to evaluation.

The above studies suggest that explicit standards are advantageous because they are generally more reliable and can be applied by less expensive personnel, but that implicit standards may be more valid. This view is supported by some authors (62,144). It may also explain the growing practice of using explicit standards as a screening device only. In this approach all instances of non-compliance are further evaluated by
professionals using implicit standards before any problem analysis is attempted. It may be tentatively concluded that at present the use of explicit criteria and standards does not obviate the need for eventually applying implicit ones.

Derivation of Criteria and Standards

Donabedian (144) pointed out ten years ago that criteria and standards can be selected from data about existing practices (empirical criteria) or from what is considered to be good practice (normative criteria). While recognizing that use of empirical criteria may perpetuate an unfortunate and undesirable status quo, Donabedian has also cautioned that such survey data will give us valuable information about what has been achievable and practical to date.

Sanazaro (451,452) has further developed the empirical and normative criteria concept. He postulates three basic types of criteria and standards: statistical (empirical); normative (derived from consensus about good practice); and scientific (validated by clinical trial). Sanazaro sees empirical criteria as endorsing the status quo and rejects these as inappropriate for evaluation of quality. He subdivides normative criteria into general consensus criteria which have been identified by practitioners as representative of good care, and essential criteria which have been identified by experienced and recognized experts as critical to a desired outcome for a specified problem. Sanazaro places in the general consensus group the innumerable criteria sets, often referred to as laundry lists and cookbooks, which list all commonly accepted process measures for any given condition. He states that using this type of criteria set will not only increase costs of care as the criteria are
adhered to, but also will lead to ineffective corrective measures. Other authors point to documented difficulties in relating popular but otherwise unjustified process criteria to desired outcomes (133,163,428). Sanazaro favors the essential criteria approach to evaluation despite the risk of non-acceptance by practitioners who do not "own" them. He goes on to emphasize that the third type of criteria, those which have been scientifically validated by clinical trial, are the criteria of choice when available.

While Sanazaro's conceptualization is helpful and his preferences are theoretically attractive, there are additional factors to consider. The first is that practitioners and experts may not vary greatly in what they consider to be important. Whereas Wagner (519) found that education and experience influenced the selection of process criteria for medical care, the nation-wide pediatric criteria demonstration project showed that there was a high level of agreement between practitioner and expert groups (327,499). The second is that even if experts and practitioners do not agree on criteria specification, a number of authors believe practitioners should set criteria anyway. Reasons commonly given are that the process is thought to be educational (125,159,387) or is thought to increase likelihood of practitioner support of the review process (125,355,460). Experience in Utah (355) suggests that this latter reason may be a valid one. Some sources suggest that practitioners modify or ratify pre-established criteria as a compromise to the dilemma of expert versus practitioner originated criteria (12,125, 265,474,514).

Although it may be premature to generalize about the greater desirability of expert generated criteria, the choice of essential
criteria in preference to "laundry lists" seems supportable in terms of practicality as well as validity (125). (It has already been suggested that judgements using essential criteria are more likely to show agreement with judgements using implicit standards). Thompson and Osborne (499) emphasize that where criteria are not supported by clinical research, it is important to accept only those criteria on which there is a high level of agreement as to their criticality.

Before leaving the discussion of standards, the use of indices should be mentioned. Payne (315) suggests that measurement of performance can be carried out by at least two general approaches: using single distinct measures or combining several measures into one number as an index which is usually expressed as per cent of criteria met. The first choice is indicated when a few behaviors will define the component under study, when few measures are available or can be assumed to be representative of others, when few analytic comparisons are planned, or when the purpose is to improve performance by feedback about discreet behaviors. The second choice is justified when performance is conceptualized as a composite of many imperfectly related items, when statistical reliability of the construct is increased with more items, when complex analysis is planned, when purpose is summative evaluation only, or when purpose is formative evaluation but summated indices are used as a first screen to direct attention to areas of greatest need. His research on the use of item importance weighting in arriving at index scores suggest that weighting can reduce the number of criteria needed, and that differential and equal weighting will produce the same results, but too short lists of criteria may decrease reliability of results.

The most common practice in health care evaluation to date has been to regard each criterion chosen for measure as its own standard,i.e.
100% of all criteria must have been met if care is regarded as acceptable. It has been suggested that compulsive pursuit of occasional situations of non-compliance is not cost-effective. Increasingly authors are encouraging evaluators to set other standards of acceptability, e.g. 80% of the criteria met, based on the threshold at which action appears worth taking (125). Sanazaro (451) writes that where criteria sets are based on general consensus criteria the limits of acceptability are necessarily low; where they are based on the essential criteria approach, a 100% level of compliance must be expected.

**Data Management**

Decisions about how information will be collected, classified, stored, retrieved and reported are crucial to effective evaluation. While all possible aspects of data management are not discussed in the literature searched, at least two topics are the object of considerable attention. These are patient care records as a data source and systems which facilitate retrieval of data.

**Patient Care Records as a Data Source**

Rosenfeld (435) states that methods of data collection include direct observation, review of records, and review of record abstracts. He suggests that patient records and abstracts are usually chosen because these methods are less costly than direct observation and possess greater reliability. There are decided disadvantages to records review however. Comprehensiveness is usually sacrificed (62,286,435) because records are
inadequate sources of information about relationships, communication, psychological and social aspects of management. Koran (285) believes the most significant failure of records review is that it cannot judge accuracy of physical findings or technical quality of treatment procedures. Brook suggests that even though records theoretically could be useful in judging quality of decision making and documenting the fact of treatment, they are currently an inadequate source of data about even these variables (62).

This last disadvantage has stimulated a number of investigative studies. A major project to establish national process criteria for child health supervision and three diseases (327, 374, 500) showed that overall documentation of the criteria was about 50% despite the fact that practitioners had previously agreed that these criteria could be found in records. Measurements and lab data were recorded frequently; counseling items infrequently. In another study of ambulatory clinic records (448), 25% of the sample could not be reviewed and of the remaining 75%, virtually one-half proved to be lacking documentation about one or more of the established criteria. In still another study which assessed ambulatory care for two chronic, five acute and three preventive categories (307), investigators concluded review of office records is not feasible.

The above studies do not answer the question "was care given and simply not recorded, or was care not given and therefore not recorded?" At least three studies in ambulatory care settings attempt to answer this. Long and Rogers (309) used direct unobtrusive observation of care of patients with streptococcal throat infections. They showed that practitioners adhered to 12 - 100% of the criteria they agreed afterward were important and actually recorded 0 - 94% of criteria. Six
out of fourteen criteria were recorded 40% (or more) less frequently than performed. Bentson (47) reports that observers of fifty-nine patient-resident encounters in a family medical care unit at the University of Western Ontario counted 2.45 problems per encounter actually dealt with, while residents recorded only 1.51 problems dealt with. When problems were classified as somatic, emotional and social, the emotional and social problems were less often recorded. Another study (547) showed that when information in fifty-one tape recorded patient encounters was compared with the corresponding medical records, the diagnosis, chief complaint, scheduled appointments, non-drug treatments, diagnostic studies and names of medicines were well recorded. Conversely, medication dosages, levels of function, compliance, and reasons for follow-up were poorly recorded. The investigators believe that the practitioners studied in this university clinic were not typical. For example, other studies show diagnosis recorded 60% of the time as compared with 97% in this study.

An important point is that all of the above studies were done in ambulatory care settings. There is some evidence that as organizational structure of care becomes more formalized, documentation of care given becomes more complete (38,307,500). Thus the problem of records being an inaccurate reflection of reality may increase with independence of practice. A related finding which has implications for overall consideration of competence is that there appears to be a positive relationship between the quality of medical records and the quality of care (314,547).

Although in-patient records apparently have not been scrutinized for their ability to reflect reality, there has been some concern
about in-patient records abstracts. Donabedian (142) states that patient records are a poor enough reflection of reality and that records abstracts are even further removed from the real world. One hospital reports difficulty getting sufficient data from abstracts to meet its audit requirements (244). Experience in other hospitals has shown that training of records abstractors and analysts is essential to prevent unwarranted dilution of data (269,293). In one rather extensive study designed to investigate this problem area (352), a trained team reabstracted a national sample of patient records of Medicare and Medicaid patients. Comparison with the abstracting service records showed high agreement with patient records for data on length of stay and patient characteristics but lower levels of agreement for primary diagnoses (65.2%) and procedures (73.2%). The findings suggest that broader diagnostic coding is more reliable, that it is often difficult to determine primary diagnosis, that individual diagnoses affect reliability, and that the quality of the hospital's abstracting service is significant and dependent on training procedures.

Systems Facilitating Retrieval of Data

If problem-specific process and outcome criteria are monitored using patient records, the records of patients with specific problems must be readily retrievable and information relevant to each criterion must be available and easily located by clerical personnel.

The literature reveals some attempts to correct the present inadequacy of documentation in patient records. Several writers have recognized the importance of the problem oriented record (POR) as a promising record format which not only demands documentation of informa-
tion relevant to evaluation, but also organizes it in a form which is easily scanned (54,70,80,452). This record system includes the collection of data, the formulation of problems based on data, the development of plans and treatment for each problem, and follow-up through the use of notes specific to each problem. Without specifying the POR format, other authors detail minimum data requirements as those data elements which describe the achievement of patient-specific, problem-specific objectives (536). Sanazaro (452) and White (533) describe early efforts to establish a national Uniform Minimum Basic Data Set with sub-sets for ambulatory care and long term care. Keller (269) believes that by standardizing records formats it would be possible to monitor quality of care not only for caseloads and facility clientele, but also for entire target populations.

Problem classification systems are an additional challenge to those professions wishing to undertake problem-specific evaluation. Without a way to classify problems, it is difficult to move beyond evaluation of the individual case to evaluation of a sample of like cases. It is equally difficult to compare research findings which establish process-outcome correlations for specific problems. Both of the above are crucial to effective and efficient quality of care evaluation.

The medical profession has done considerable work in this area over a period of years and is able to draw on a variety of classifications including the 9th generation of the International Classification of Diseases, the hospital adaptation of ICDA and more recent efforts which modify medical language for computer manipulation. These include the early SNDO (Standard Nomenclature of Diseases and Operations) which uses the dimensions of anatomical topography and etiology, SNOP (Systematized
Nomenclature of Pathology) which uses the dimensions of typography, morphology, etiology, and function, SNOMED (Systematized Nomenclature of Medicine) which expands and refines the three files of topography, morphology and etiology and adds procedures, functions and utility. This last includes syntax usages, adjectives, quantifiers, signs and symptoms etc. (43,178,183,531). Recent work has also established the ICHPPC (International Classification of Health Problems in Primary Care) which can be combined with a classification system for patients problem from the patient's perception and language (528,533).

The nursing literature shows a beginning awareness of the importance of classification of nursing diagnosis for research and evaluation in nursing. While the nursing literature is largely speculative and theoretical (53,439), some beginning attempts at a classification system are being made (183).

There were no references in the literature to highly systematized classification systems for dentistry or pharmacy. The question of a multi-use system suitable to the evaluation and research needs of all health professionals was not discussed either; it is probable that SNOMED has such a capacity or could be adapted for multidisciplinary use (183).

**Evaluation Based Interventions**

A planner interested in instituting an effective evaluation-action sequence will need to move beyond the technical questions of evaluative strategy to those that anticipate using evaluation information to achieve desired results. While the literature reporting successful/
unsuccessful evaluation-action sequences can be looked to as a rich source of ideas, few patterns have emerged from which generalizations can be made or decision guidelines formulated.

Table VI summarizes a sample of these evaluation-action sequences in terms of the variable(s) chosen as an indicator of competence (Competence Measure), type of interventive or corrective action used to upgrade current practice (Intervention), and whether there was any evidence of change in the desired direction (Improvement).

The contents of the table permit a limited number of observations to be made. First, the interventions that have been successful in promoting desired change are notable for their variety. Educational programs; administrative action in the form of new policies, procedures and services; adoption of protocols or written procedures; denial of insurance claims; personal and individualized feedback; and additions of new staff and equipment all have been successfully used to correct what initially appeared to be a performance problem. Clearly the evaluation-action sequence made popular in health care through Brown's bi-cycle concept of continuing education is applicable to the diagnosis of more than educational problems and of more than problems related to competence of health professionals. One hospital (308) reports that of twenty-one studies (seventeen clinical problems evaluated by process techniques and four by outcome assessment), only five yielded recommendations for action which would change physician behaviors. The remaining sixteen involved changes in criteria, administrative action, special studies, or no action at all. The table also indicates that it is not uncommon to undertake more than one corrective action for an identified problem.

A significant proportion of these projects either were not carried through to re-evaluation or this last step was simply not
Table VI

EVIDENCE OF DESIRED CHANGE IN REPORTED EVALUATION - ACTION PROJECTS

<table>
<thead>
<tr>
<th>REFERENCE</th>
<th>QUALITY MEASURE</th>
<th>INTERVENTION</th>
<th>IMPROVEMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barnett (40)</td>
<td>treatment of beta hemolytic streptoccal throat infections</td>
<td>computer generated immediate feedback for deviant behavior</td>
<td>yes</td>
</tr>
<tr>
<td>(multidisciplinary)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brook (66)</td>
<td>antibiotic prescribing patterns</td>
<td>education, feedback, claims denial</td>
<td>yes</td>
</tr>
<tr>
<td>(medicine)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condon (107)</td>
<td>head nurse perception of improvement in staff performance re 4 steps nursing</td>
<td>workshops on how to carry out inservice education</td>
<td>yes, but gains</td>
</tr>
<tr>
<td>(nursing)</td>
<td>procedure</td>
<td></td>
<td>variable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Christoffel (97)</td>
<td>death rate - myocardial infarction</td>
<td>new equipment purchased</td>
<td>unknown</td>
</tr>
<tr>
<td>(medicine)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(medicine)</td>
<td>diagnostic techniques and antibiotic choice, resolution of pneumonia</td>
<td>education, new lab procedure for specimen collection</td>
<td>unknown</td>
</tr>
<tr>
<td>(medicine)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(medicine)</td>
<td>incidence of neonatal pneumothoraces</td>
<td>increased chart documentation then further investigation</td>
<td>unknown</td>
</tr>
<tr>
<td>(medicine)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(medicine)</td>
<td>pathology of removed uterus</td>
<td>feedback and discussion</td>
<td>yes</td>
</tr>
</tbody>
</table>

...continued
<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention</th>
<th>Effect</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(medicine/ nursing)</td>
<td>insulin prescribing patterns... diabetic teaching practices... incidence of hypoglycemia... incidence of urinary tract infection in diabetics...</td>
<td>education preparation of diabetic teaching material new treatment protocol new diagnostic protocol</td>
<td>yes, and incidentally length of stay was decreased</td>
</tr>
<tr>
<td>(medicine)</td>
<td>rate of justifiable surgery for duodenal ulcer</td>
<td>counselling, provision of consultation to offending staff members</td>
<td>unknown</td>
</tr>
<tr>
<td>(medicine/ nursing)</td>
<td>diagnosis and care of patients with schizophrenia</td>
<td>policy changes, increased staff creation of new unit, protocol for suicide precautions</td>
<td>unknown</td>
</tr>
<tr>
<td>Devitt (132) (medicine)</td>
<td>degree of pathology found after mastectomy and cholecystectomy</td>
<td>education</td>
<td>no</td>
</tr>
<tr>
<td>Diamond (40) (medicine)</td>
<td>use of psychotrophic medications</td>
<td>education</td>
<td>yes</td>
</tr>
<tr>
<td>Dyer (151) (nursing)</td>
<td>patient care score derived from several instruments</td>
<td>head nurse modelling, assistance with goal setting</td>
<td>yes intially but most gains lost at 12 months</td>
</tr>
<tr>
<td>Escovitz (159) (medicine)</td>
<td>diagnostic accuracy diagnosis specific process criteria</td>
<td>change in consulting program education</td>
<td>yes yes but not to acceptable level</td>
</tr>
</tbody>
</table>

...continued
<table>
<thead>
<tr>
<th>Author</th>
<th>Field</th>
<th>Process/Measure</th>
<th>Outcome/Change</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glasson</td>
<td>Nursing</td>
<td>Discharge planning, all patients</td>
<td>Education, changes in records forms</td>
<td>yes</td>
</tr>
<tr>
<td>Graydon</td>
<td>Medicine</td>
<td>Process criteria for 5 common diagnoses</td>
<td>Feedback of results from source external to hospital</td>
<td>no</td>
</tr>
<tr>
<td>Haussmann</td>
<td>Nursing</td>
<td>Composite of several process measures</td>
<td>Inservice education, reorganization of unit</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Return of previously sick head nurse</td>
<td>yes</td>
</tr>
<tr>
<td>Lewis</td>
<td>Medicine</td>
<td>Perinatal death rate, incidence of specific surgical procedures, completeness of physical exams</td>
<td>Education</td>
<td>no, except physical exams were more complete</td>
</tr>
<tr>
<td>Miller</td>
<td>Medicine</td>
<td>Several problem-specific criteria, patients' rights, admission rates, length of stay</td>
<td>Type of intervention not reported</td>
<td>yes</td>
</tr>
<tr>
<td>Mitchell</td>
<td>Medicine</td>
<td>Length of stay for uncomplicated cholecystectomy</td>
<td>Education</td>
<td>yes</td>
</tr>
<tr>
<td>Mosley</td>
<td>Medicine</td>
<td>Attitude scale, work performance</td>
<td>Experimental education</td>
<td>yes for work, no for attitude</td>
</tr>
<tr>
<td>Study (Year)</td>
<td>Topic</td>
<td>Methodology</td>
<td>Result</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>-------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Ramirez (404) (nursing)</td>
<td>Quantity and quality of nursing care plans</td>
<td>Feedback and education</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Rubenstein (442) (medicine)</td>
<td>Several process measures (e.g., use of IV heparin, packed cells, etc.)</td>
<td>Education</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Russo (448) (medicine)</td>
<td>Diagnostic and treatment criteria for 26 common conditions</td>
<td>Private discussion, feedback</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Sanzaro (451) (medicine)</td>
<td>Complication rate post cholecystectomy</td>
<td>Education</td>
<td>Yes, from 25% to 13%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pathology of removed appendices</td>
<td>Education</td>
<td>Yes, from 45% to 19%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inappropriate use of antibiotics</td>
<td>Education</td>
<td>Yes, 70% to 40%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Composite performance indices using process measures for 8 conditions</td>
<td>Education</td>
<td>Yes, but slight and variable among conditions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Follow-up of abnormal lab results</td>
<td>Education</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pathology of removed appendices</td>
<td>Administrative requirement to provide relevant diagnostic information</td>
<td>Yes, did not persist when requirement removed</td>
<td></td>
</tr>
</tbody>
</table>

...continued
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Protocol</th>
<th>Yes/No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process criteria for diagnosis and treatment of acute sore throat</td>
<td>Treatment protocol</td>
<td>Yes</td>
</tr>
<tr>
<td>Process criteria for several conditions</td>
<td>Computer generated feedback for deviant behavior</td>
<td>Yes, no for control group</td>
</tr>
<tr>
<td>Rate of justifiable pelvic surgery</td>
<td>Personal discussion with offending surgeons, provision of criteria to staff</td>
<td>Yes</td>
</tr>
<tr>
<td>Rate of abortion and delivery in population desiring information about contraception</td>
<td>Staff discussion, reorganization services, revision of records format, sex counselling services</td>
<td>No</td>
</tr>
<tr>
<td>Treatment of stroke victims</td>
<td>Education, additional (expert) staff</td>
<td>Yes</td>
</tr>
<tr>
<td>Use of specialty consultants, type of intervention not reported</td>
<td>Feedback, claims denial</td>
<td>Yes</td>
</tr>
<tr>
<td>Prescribing and dispensing patterns</td>
<td>Programmed learning unit</td>
<td>Yes for performance and knowledge no for attitude</td>
</tr>
<tr>
<td>Performance in problem situations, test of knowledge and attitudes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...continued
<table>
<thead>
<tr>
<th>Williamson</th>
<th>several diagnosis-specific process and outcome studies</th>
<th>education</th>
<th>possibly</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>education</td>
<td>possibly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>education</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>education</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>administrative action</td>
<td>possibly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>administrative action</td>
<td>possibly</td>
</tr>
</tbody>
</table>

Table VI
reported. There is at least some evidence to suggest the former is the case (184,351). This hampers both the investigator, as he cannot know whether action has been effective, and the development of sound principles or decision rules which can govern the conduct of evaluation-action sequences. Finally, it is apparent from the table that few investigators have addressed the question of whether improvement is transient or lasting. Put another way, there is as yet little concern as to whether strategies for maintaining changed behavior are needed.

It should be noted that Table VI represents a small sample of the reports of evaluation-action sequences available and provides only the most basic information. This is a potentially fruitful area for investigation and it is likely that an in-depth study of a larger sample might well indicate patterns typical of successful and unsuccessful projects. Questions worth asking are: is success influenced by practitioners' perception of whether there is a problem? by severity of problem? by existence of an on-site administrative structure for evaluation activities? by degree of practitioner participation in the evaluation-action project? While these are only a few of many possible questions they are listed because current literature suggests these may be influencing factors and also because many of the present reports contain information relative to these issues.

Summary And Comment

Experience with evaluation-action approaches to improving care is sufficiently well reported in the literature to provide planners with
some guidelines. There is general agreement that an effective system will be planned with attention to validity, reliability, practicality, comprehensiveness, and potential for giving direction to and evaluating action. It will be capable of documentation. At the same time it will allow for participation by practitioners and will not stifle innovation. Planners should proceed from a base which is characterized by a clearly defined frame of reference, a sense of the universe of professional behaviors which collectively are an expression of competence, and an unambiguous statement of values.

The decision as to which category of variables (structure, process, outcome) will be measured should be made with consideration for the decided advantages and disadvantages associated with each. A chief problem area is the low or non-existant correlations between structure and process, process and outcome, and structure and outcome measures. This is a priority area of research for all the professions. Such research demands a greater degree of scientific rigor than do evaluative studies undertaken to improve care; the two should not be confused. In an effort to minimize disadvantages and maximize advantages, investigators are increasingly turning to collection of process and outcome data with respect to any given problem.

There has also been a marked tendency to use explicit criteria and standards in evaluation, rather than leave these totally to the judgement of the reviewer. This approach is seen as less costly and more reliable, but its primary usefulness may be as a screening device in mass evaluation. As yet the need for validity has tended to lead evaluators to use implicit standards in a final review of identified deviations.

Explicit criteria may be established using a variety of approaches including the use of statistics which reflect existing prac-
tice, popular opinion about what represents good practice, expert opinion about practice essential to desired outcome, or research derived evidence of what constitutes good practice. In the absence of the latter, essential criteria are seen as preferred even though they risk non-acceptance by the practitioner group.

The literature addresses some but not all aspects of data management. Patient records appear to be a relatively cheap and reliable source of data in institutional settings, but are not so useful in less structured situations. They are obviously limited in the kind of information they provide. Some attempts have been made to increase availability of data by proposing minimum basic data sets. The problem oriented record is seen as a useful data base for evaluation purposes. Retrieval of aggregate data about care of specific patient populations and of information about established structure-process-outcome correlations is hampered by the lack of problem-based classification systems in all professions except medicine.

Finally, reports of evaluation-action projects which have been undertaken in the past several years make it clear that improvements in care can be effected; that change strategies are of many types, of which continuing education is only one; that many studies fail to remeasure to determine success of actions taken; and that almost no attempts have been made to investigate or ensure the permanence of successful change.

One can conclude that a beginning has been made in that tentative guidelines have been developed for the evaluation aspects of evaluation-action projects. Understanding of the problem analysis and action phases as they exist in health care is not so well established. As reports become more numerous and complete, it is likely that some
helpful generalizations can be made. Other areas needing research and development are problem-based classification systems relevant to the professional practice of all health workers; further testing of the reliability of structure, process and outcome measures; work which establishes correlations between these; and organization and provision of information on cost and impact of evaluation-action systems.

The need for this last item may be viewed as a pragmatic reflection of the concerns of a society which is increasingly aggressive in questioning the competence of the entire health care delivery system. Activities concerned with evaluation and maintainance of professional competence must be viewed within the larger social context; this is the focus of the next chapter.
Chapter VII

THE NORTH AMERICAN SOCIAL CONTEXT

Maintaining competence of health professionals implies that there will be both evaluation and appropriate action in light of that evaluation. Consequently, the bulk of this paper has dealt with literature relating to these two issues. Evaluating and maintaining competence are not solely mechanistic endeavors which employ sterile procedures however. These activities are undertaken by and for human beings individually and collectively. The literature searched deals with some of the psychological considerations and socio-political developments which both influence and are influenced by evaluating and maintaining competence of health professionals. This literature is described in Chapter VII.

The first section, The Social Ecology of Maintaining Competence, is a general discussion of literature items related to psychosocial, economic, and political issues which seem to be most directly related to evaluating and maintaining competence. The second section, Regulatory Mechanisms and Trends, deals with literature concerned with society's attempts to regulate professionals and certify their competence. The last section is Mandatory Continuing Education. This specific credentialing trend is singled out for special discussion because of its special interest for continuing educators.
Developments in monitoring competence which are discussed in this paper have taken place in a decade characterized by a growing understanding about the needs and rights of individuals, the use of the group (collectivism) as a powerful tool to ensure protection of these rights, and a growing awareness that the expenditure of limited resources will be decided in the political arena. While not wishing to report fully the health care literature dealing with these issues, it is instructive to note their influence on practices for monitoring competence of health professionals.

Needs and Rights of Professionals

Self regulation is one of the traditional hallmarks of a profession. To suggest that peer review is a new concept would be misleading. It is apparent, however, that evaluative practices in the four professions considered have recently become more intensive, more specific and in some instances more direct. In such a situation, both economic and personal security are threatened. When a professional is accused or suspected of inadequate practice, he has a right to due process - the fair hearing of his case. The increased awareness of the importance of due process is reflected in the literature and has been referred to in earlier chapters. Personal security in the sense of good reputation or favourable self-concept is more difficult to protect however. Concern for economic security and need for the esteem of self and others influences attitudes about evaluation, especially evaluation by a peer group.

There seems to be little literature dealing with professionals'
attitudes about peer review. At Duke University's Department of Psychiatry, staff members' attitudes about peer review were measured with a thirty-six item questionnaire (198). While two-thirds of the respondents conceded that peer review was needed, most tended to believe that review panels should consist of senior staff psychiatrists, thus showing both a preference for interpreting the concept "peer" rather broadly and an inclination to accept the authority of greater knowledge and experience. Among all respondents, having a degree of resentment about peer review was significantly correlated with seeing a low need for it. Dekker (123) reports that in a 1976 nationwide survey of American physicians asking whether intensified peer review activities would improve patient care, 45% said "yes", 50% said "no" and 5% were "undecided". While the nature of the responses indicated little outright opposition, skepticism was widespread. Among the questions raised: How do you agree on a definition for quality care? Is non-standard care sub-standard? Is poor care worse than no care at all? Despite what appears to be fairly neutral attitudes on the part of the majority of the medical profession, the tendency to keep evaluative activities at arms length is noticeable: "We must face surgical committees of our peers, but these committees examine the records --- never does a committee insist on scrubbing in, taking notes as we operate, or testing our knots --- our offices are sacrosanct, and we write down what we please --- rarely are our descriptions of findings, diagnosis or treatment challenged" (228).

Dentistry has not concerned itself greatly with attitudes about peer review. Jago (257) writes that historically dental societies have acted primarily as an arbitrator for fees and that peer review committees
might more aptly have been named "peer justification committees". If Jago is correct, dentists' experiences with peer review have been non-threatening; thus one might expect present attitudes about peer review to be relatively neutral.

Direct and purposeful performance review has a long history within the nursing profession. The nurse's status as an employee has meant continuing supervision and evaluation by an administrative superior from within the profession. Recent nursing literature employs a far narrower interpretation of peer review than simply evaluation by a fellow professional however. This literature infers that a peer is of the same education, rank, and qualification within the profession. Pardee (379) clarifies the issue of "what is a peer," by suggesting that the definition of peer will be flexible in relation to the setting, purpose and goals of review; i.e. one must know what is to be evaluated before one can identify the peer. For example, a head nurse will have other head nurses as peers when reviewing administrative performance and will have staff and clinical coordinator as peers when reviewing clinical skills. The nursing literature gives guidance to nurses wishing to institute peer review systems (19, 81, 190, 316, 379). The advice offered could apply to any evaluative system, however, and is not related specifically to review by peers.

There are no attitude surveys reported in the nursing literature, but two "personal experience" reports (190, 224) show that nurses find peer review both threatening and rewarding. Both situations involved face-to-face interaction and used direct observation and records as a source of data. Both also reported the review to be more useful
when characterized by explicit standards and a structured review procedure.

Despite recent reports of neutrality or support of intensified peer review activities by the various professions, a 1974 Wisconsin Medical Journal editorial (228) may be a more honest and useful assessment of professional attitudes about peer review: "At one end of the psychological continuum are the lusty exhibitionists who loudly declare that they are not afraid of peer review, that everybody makes mistakes, and that mistakes must be talked about, are the very stuff of teaching (or staff) programs, and those who cannot take the heat should get out of the kitchen --- Another position would be that of the delicate doctor who is in abject terror of anyone at all looking at his work. He will go to any length, devise whatever intellectual strategem, to prevent the abrogation of his right to privacy, his God-given right to practice medicine as he sees fit -- all human beings fall somewhere on the scale --- no matter what the beauty of the logic of a review system, most of us will be rattled by it. For these reasons I believe peer review passions as well as peer review anxieties should be taken seriously". The writer's conclusion is a sensible and sensitive one. There is considerable evidence to date that attention has been lavished far more on review procedures than on review passions.

Rights and Needs of Consumers

If professionals have rights and needs which must be respected, so do the consumers whom they serve. The right to privacy, a contentious issue throughout contemporary North American society, has profound
implications for review mechanisms, whether they are based on patient records or direct observation of care. To what extent can a non-involved professional delve into records or observe the giving or receiving of care before he has seriously violated the patient's right to privacy or perhaps even adversely affected that care? This question does not appear to have been seriously explored in evaluation literature, although it is likely that much of the literature concerned with patients' rights in medical research would be applicable.

Elaborate confidentiality regulations have been developed in connection with PSRO review (361,427). Despite earlier statements that PSRO data should be made available to other government organizations for health care planning and disciplinary purposes, such information is carefully guarded in the name of confidentiality. This arrangement appears to have been designed to protect the professional as much as the consumer, but perhaps realistically reflects the degree of threat the professions will tolerate before their members tend to subvert instead of support evaluation systems.

It is thought that the presence and participation of consumers in all areas of health care, including evaluation of professionals, will ensure that the interests of the consumer are protected. Cunningham (116) reports that the Joint Commission on Accreditation of Hospitals will provide for hearings for consumers in connection with hospital accreditation surveys, and states that the American principle of lay interference in politics and education is spreading to medicine. Consumer participation in health care is largely a motherhood issue however,
and most of the literature is rather vague on exactly how consumers can productively participate.

Angevine (29) is typical in this regard when he says that consumers must have input into standards but does not suggest how this can be effected. Mitchell (343) is more specific, suggesting that consumers can contribute to determining the amenities to be provided in the health care system, can examine results of evaluation, and can decide non-technical issues of patient care. He states that there is evidence that fears of consumers trying to dictate technical decisions are unfounded. Maloney (325) suggests that the ethical standards of a profession should reflect the values of society at large, and that professional behavior standards which have their origin in a professional code of ethics, e.g. a ban on advertising, should be subject to review by consumers. Jago (257) discusses consumerism in health care and reviews two "consumer participation" models which attempt to quantify the degree of consumer influence in a system. He also mentions that Trantow has searched and failed to find any evidence that consumer participation has increased quality. There is a noticeable trend toward legislation in both Canada and the United States that mandates consumer participation on licensing boards; this will be discussed in the section Regulatory Mechanisms and Trends.

Collective Action

The use of group to ensure that the needs and rights of the individual are met is not a new concept in North American society, the
labour movement being the most notable example. If patients' rights or consumer groups have banded together outside government to assure or promote health professionals' competence, this is not reported in the body of literature searched.

On the other hand, the several health professions have organized to ensure that they are specifically and meaningfully included in the activities aimed at promoting or evaluating competence. A most notable example is the dominance of the organized professional groups in the development of the many standard criteria sets referred to earlier. Blumstein (56) refers to the influence of organized medicine in gaining a position of almost unilateral control of PSRO activity. Other authors report considerable activity on the parts of dentistry (367) and nursing (25) organizations to gain power in the PSRO decision making structure.

The nursing literature deals specifically with collective bargaining as a tool to influence matters related to promoting and monitoring competence. Jacox (256), Ramphel (405), Kellog (270) all suggest collective bargaining can be used by nurses to increase the influence of practitioners in practice-related decisions. Ramphel believes that a strong staff nurse organization strengthens the bureaucratic head of nursing service by increasing expertise and providing support in the face of interference from the medical staff and hospital administration. Both Ramphel and Jacox emphasize that to avoid a conflict of interest, the activities of peer review must be kept separate from those of the bargaining unit even if the latter brought the former into being. Ramphel believes the main avenue open to the employed practitioner who wishes to assume responsibility for quality is collective action through the professional association. Jacox shares this view but adds that
political activity and participation in setting criteria and standards and in evaluation programs are powerful tools as well. The California Nurses Association has published specific guidelines for peer review groups which operate in conjunction with bargaining units (81).

Nurses have used job action, including strikes, to gain power in decisions about evaluation of care (316, 402, 540). Maas et al. (316) provide a particularly detailed account of the development of a nursing organization in one institution brought into being via a contractual agreement which made nurses collectively responsible for the delivery of nursing services, with the nursing administration responsible for executing policies decided by the organization. As other professionals increasingly assume an employee status, the possibility of collective bargaining as a tool to gain influence in competence monitoring activities may be of greater interest to these groups.

The Rising Costs of Health Care

In the seventies concern about health care has expanded from concern about the individual to concern about the community. As health care consumes an ever larger proportion of available resources, questions about the efficient delivery of that care and its value to the health of the community are asked with increasing frequency. Blum (55) believes that if our goal is better health for the total population, maximum benefit will be derived if we shift our evaluative focus from behaviors of professionals to behaviors of patients, i.e. we should measure health status, practices and behaviors at entry to care and health status on exit. This, of course, is the philosophy of Marc La Londe's *A New Perspective on Health Care for Canadians* (82). Mitchell (343) states,
"Obviously we must begin to think about the overall health system and sub-systems --- while not neglecting the necessity for testing the excellence of individual acts of medical care".

Assuming that the competence of professionals will continue to be examined, the issue of cost cannot be ignored. In discussing the relationship of quality to cost, Lewis (298) postulates that there are at least three general sets of cost-quality function curves:

1. linear and positive (i.e. the better the quality, the higher the cost); these seem to be associated with the more complex medical phenomena, e.g. end stage renal disease

2. a scatter diagram (i.e. there is no relationship between quality and cost); one would see this with such self-limiting diseases as the common cold

3. quadratic in form (i.e. 80-90% of excellence to be achieved will be gained with a relatively small investment, and a continued investment will yield only marginal gains); a familiar example would be the information derived from a thirty minute physical as opposed to the exhaustive in-house check-ups which corporations arrange for their top executives.

Lewis recommends research to develop problem-specific cost effectiveness curves. Such an accomplishment could have profound effects on both formulation of review criteria and the focus of review activities. Rubenstein et al. (443) report using process and outcome criteria to evaluate emergency room care given women with urinary tract infection in two hospitals. In both instances process and outcome scores were positively correlated. The data suggest that correlation is curvilinear and that there is a threshold process score below which a poor outcome is extremely likely. This study indicates that the greatest improvements in outcome of urinary tract infection may result from raising the quality of care from a poor to an adequate level, rather than from an adequate to an optimal level.
The information provided by Lewis and Rubenstein does not support the beliefs of those who claim that PSRO review can effectively and simultaneously control cost and improve quality. Blumstein (56) contends that the PSRO legislation, which encourages wholesale adherence to optimum process or outcome criteria without regard for their economic impact or import, is a potentially effective and unopposed instrument for increasing health expenditures. While perhaps increasing the technical quality of services provided the individual patient, the effectiveness of the health care sector in maintaining or improving the health of a defined population may suffer. Several other writers (51,161,196) have also predicted that the effect of PSRO legislation will be to increase direct health care expenditures. Initial evidence indicates that PSRO-type organizations achieve, at best, marginal cost control effects and some gains in quality (389,401,424).

It may be concluded that enthusiastic and vigorous implementation of any review program without concern for its direct and indirect costs and benefits could well yield a net loss rather than a net gain in the health state of a given population.

Malpractice

Malpractice litigation as a mechanism for the occasional evaluation of the process and outcome of care has been discussed in Chapter V. The economic effects of malpractice litigation on the health care system, particularly in the United States, have influenced other developments related to evaluation of professionals. PSRO activity, for instance, has been opposed in the belief that it will increase malpractice litigation by exposing poor care; it has also been supported with the prediction
that it will decrease claims by preventing shoddy practice (169,438). At least one liability insuror has provided incentive for systematic evaluation by reducing the agency's corporate primary coverage malpractice premium on the assumption that the agency's review program will not only control undesirable practices, but also document proper care (227).

Liability-conscious administrators are careful to keep review records separate from patient records (390). Some jurisdictions support this precaution by ruling that such records are not discoverable as evidence. Such is the concern about malpractice or corporate negligence that a new term, "risk management program", has appeared in health care literature (212,245). A typical program involves an aggressive quality assessment system supported by a large budget, a mechanism for handling consumer complaints, management planned campaigns aimed at promoting good relations between staff and patients, and tightly structured mechanisms for evaluating competence and extending certain practice privileges only to specified personnel.

Agency-based limitation of practice as a strategy to avoid lawsuits has been discussed to some extent in Chapter III. Several authors question whether limitation of hospital privileges is justifiable, as it puts the hospital in the position of being a "super licensing board" (88,100,396). The Joint Commission on Accreditation of Hospitals suggests that such procedures need to be applied flexibly, with consideration for available manpower, the needs of the population, and availability of alternative resources (396). Hospitals may also use external mechanisms, such as specialty certification or special licenses, as a part of their litigation defense program (310,412). Whether the desired effect is ever obtained is not documented, but Egelston (155) predicts an
increasing use of both in-house and externally awarded credentials, earned through a variety of evaluative mechanisms, for trustees, administrators, physicians, nurses, and allied health personnel. Thus it appears that existing evaluative devices leading to a credential will continue to be used and new ones, primarily agency based, increasingly will be superimposed on this credentialing non-system.

Health Manpower

Rapid technological developments coupled with concern for rising costs has resulted in a proliferation of categories of health workers (236). Some of these, like the physician assistant, are new workers. Dekker (122) reports that in the eleven years since Duke University opened the first formal training program for physician assistants, some forty states have enacted legislation governing their use. Canada has not used this worker. In both countries nurses' roles have expanded to include a physician associate worker most often referred to as the nurse practitioner. Bullough (75) states that nursing has moved through two major phases in licensure (basic registration and mandatory registration) and is currently in a third (role expansion). In addition, a variety of technical and auxiliary workers are being trained and employed to support newer developments in social and medical services. There arises the question, how will these workers be evaluated and otherwise controlled?

Some writers question using traditional licensing systems (176, 325,336) for these groups. They point out that new worker categories were developed so that highly paid and extensively educated professionals would not be doing technical work. Licensure would restrict entry to the
occupation, thus limiting availability of services and artificially inflating wages. Additionally, it is believed that licensing promotes rigid categories of workers, further interfering with efficient use of manpower in a time of rapid change (134,336).

A prominent lawyer in the American health care field has suggested a legislative alternative, institutional licensure (208,235,236,497). Under the Hershey proposal, institutions would be responsible for establishing task lists, job descriptions and for monitoring the competence of the personnel filling such positions. This would be done under the surveillance of the state licensing agency. The proposal has sparked similar alternative proposals (167). Despite considerable discussion, there does not appear to be any serious moves toward institutional licensure, perhaps because as Guy (208) points out, such an approach would remove all the checks and balances.

The control, evaluation, and efficient use of health manpower continues to be a major issue, complicated by difficulty of reaching national standards (128,188); movement of workers, especially women, in and out of the work force (68); the presence of health professionals trained for another health care system; and the "foreign graduate problem" (128,271,511). This issue, and many of those discussed earlier in this section, have promoted considerable activity with respect to formal credentialing mechanisms for health care personnel throughout North America.

**Regulatory Mechanisms and Trends**

The credential is the currency of competence in North American
society. The credential is not competence any more that the dollar bill
is gold. The utility of the credential is greatly diminished if one fails
to understand the purpose for or conditions under which it was granted.
Historical development of regulatory mechanisms, licensure and certifica-
tion today, and current trends in credentialing practices will be discus-
sed in this section.

Historical Development of Regulatory Mechanisms

The idea that society has the right and responsibility to
control competence of health professionals is not new. The Code of
Hammurabi, circa 1955 B.C., specified that those physicians with records
showing a high degree of mortality amongst their patients lost their own
lives (331).

Derbyshire (128), provides the following account of the devel-
opment of regulation in medicine up to the twentieth century. The Code
of Frederick II, 13th century Sicily, laid down educational requirements,
set fees, regulated ethical conduct and other practice behaviors, and
allowed the medical faculty to license physicians. Surgeons were appren-
ticed and eventually examined by masters in the trade prior to joining a
guild. When England in 1518 granted a charter to the Royal College of
Physicians and Surgeons, the principle of self-regulation of the profes-
sion was well established. Licensure by school and regulation by peers
is the system still predominant in Europe today.

The relatively unstructured society in the developing New World
colonies provided few of the traditional controls. At first, training of
physicians was by apprenticeship. In the early nineteenth century, exam-
ination by the local medical society was required by law; later medical
degrees from educational institutions were accepted in lieu of an examination. By the end of the Civil War, however, "diploma mills" were common and a degree carried little meaning. Moreover, local medical societies were often little more than groups of cultists. For these reasons the concept of direct examination and licensure by the state became a reality. It is worth noting that both the United States Constitution and the British North America Act (Canada) assigned matters of health to the responsibility of the several states and provinces, rather than the federal governments.

The decisions which North American society had made regarding control of physicians set a pattern which was to influence the regulation of all health professionals for the next half century. Roemer (426) notes that the contents of licensing laws for all health occupations are similar because they are modelled after medical practice acts. Generally they establish the scope of practice and prescribe the conditions and methods by which licensure can be obtained, maintained, and revoked. In relation to the latter, incompetence is a frequently mentioned condition for some form of discipline, the ultimate sanction being revocation of license (128).

As the health care delivery system continued to take shape in the early twentieth century, additional attempts to assure competence were initiated. The most significant were accreditation of professional training programs, accreditation of health care agencies, and certification of specialty qualifications.

Historically, accreditation of educational institutions and programs has been a voluntary non-governmental process of evaluation which aims to assure that education is of high quality (362). Approval
by the state or province also exists in some occupations and generally is aimed at ensuring minimum standards. Between 1900 and 1930 accreditation was begun in the fields of medicine, nursing, dentistry, occupational therapy, physical therapy, podiatry and osteopathy; by 1972, accreditation mechanisms existed for a total of forty-two occupations (426). The number and types of accrediting services for professional education programs today are extensive and the situation is complicated by shared and overlapping jurisdictions. In the United States, there are even two accrediting agencies to accredit the accreditors. As these services developed in the first half of the century, licensing laws began to include graduation from an accredited school as a requirement for licensure (105,128,426,511).

Accreditation of health care agencies also began around sixty years ago in the United States when the American College of Surgeons initiated a Hospital Standardization Program (83). Eventually, cooperative agreements among Canadian and American medical and hospital associations led to these groups sponsoring the establishment of the Joint Commission on Accreditation of Hospitals (JCAH) in 1951. The Canadian Council on Hospital Accreditation (CCHA) was established as a separate entity in 1958, and the JCAH then devoted exclusive attention to American institutions. Accreditation services are gradually being broadened to include extended care institutions as well as ambulatory, personal, and intermediate care settings. Many of CCHA and JCAH standards relate to evaluation of personnel and services (83). In addition to these voluntary accreditation programs, most health care institutions are affected by a variety of government licensing arrangements and financing controls (426).
Specialty certification is generally a voluntary process and the function of professional associations (426). Usually post graduate educational requirements and successful completion of an examination are required. This form of credential is rarely regulated by law, but is often an important requirement for receiving staff privileges. In addition an agency with specialty certified staff may be in a more favorable position to receive accreditation or garner private or government funds for special projects (426).

In summary, as recently as the fifties, the public and the professions in the United States and Canada were apparently satisfied with a system for assuring competent care which rested upon three broad foundations:

1. licensure of the health professional after he had graduated from an approved or accredited educational program; successfully completed an examination, and met various other requirements. (In most Canadian provinces government control is indirect with these functions being delegated to professional associations);

2. certification, usually by the profession, of additional specialized qualifications for those professionals wishing to be so recognized; and

3. a combination of voluntary self-regulation and government regulation of health care agencies where at least some professionals practice at least some of the time.

Of these, only the first was universal, and there is evidence to suggest that this has not been effectively implemented (99, 128, 336, 369, 427). For a detailed and comprehensive review of past and present regulatory mechanisms, several summaries provide more extensive information (60, 128, 362, 426, 511, 512, 513).
Licensure and Certification Today

The license is the credential common to the four professions considered in this paper. Writers agree that its purpose is protection of the public from incompetent practitioners (95,100,136,167,200,275,472). Despite this apparent agreement, Cohen and Miike write that there is confusion on this point and that some believe licensure is for admission of new members to a profession (100). Kinkela and Kinkela's (178) more detailed explanation of the protection of the public argument can be summarized as follows:

1. The right to pursue a lawful calling, business or profession is a fundamental, but not an absolute right; i.e. it is one which is subordinate to the public good.

2. The state protects the public good through the exercise of its police power; this power extends to all matters pertaining to the public health, welfare and safety.

3. Licensure of health care professionals is one example of the proper exercise of the state's police power.

4. The state legislature is the only body empowered to determine what licensure laws shall be.

5. Administration of the law may be delegated.

Although written about American law, each of the statements above applies under Canadian law to the several provinces. With reference to the last item, delegation in the United States is most often to a state-appointed board or commission; in Canada, to a named professional organization. In both countries this rationale is used to license, or give permission to, designated individuals to practice and to bar all others from practice (mandatory licensure). Sheppard (472) and Forni (167) note, however, that licensure may also simply confer exclusive right to use a specified title (voluntary licensure). Licensed groups maintain a register of
those licensed to practice or to use a protected title, and they often use the term "registration". This is a confusing point as any organization can issue a certificate of membership registration to its members, with or without permission of the state. To establish that a health worker is licensed, i.e. controlled in some way by the state, one must know that the law protects his right to practice and/or hold title. If one considers the marketing equivalent of these two strategies (mandatory licensure would be like removing from the shelf all goods thought to be defective; voluntary licensure would be similar to leaving all goods on the shelf, but labelling those thought to be safe), the economic implications become obvious. Clearly a key point in deciding that mandatory licensure is in the public interest is in predicting how much damage a potentially deficient practitioner is likely to do and how successfully his performance can be monitored to prevent damage.

Certification is used by the professions usually to recognize superior performance, often in a narrow practice area or specialty (148, 306). Lloyd (306) states that certification is important in assuring high quality because it establishes standards above the minimum licensing level. Dunkley (148) notes, however, the historical evidence that certified excellence in one generation becomes a minimum standard for specialty practice in the next. Child and Zuidema (95) write that as professional organizations develop standards in response to self-regulatory privileges, these have been substantively borrowed and written into law. An example would be the use of the National Board of Medical Examinations tests for licensure purposes.

It can be concluded from the foregoing that licensure is associated with state or provincial control and the concept of a minimum
standard of performance below which the health, safety, or well-being of the public is jeopardized. Certification, on the other hand, is more often associated with standards above minimum level and with private control, usually from within the profession. Both of these mechanisms, together with registration practices of groups not recognized by law and with certification practices of educational institutions which award diplomas, certificates, and degrees, comprise a complex and formal credentialing network for health professionals.

Rationalization of Licensing Procedures - Trend One

Licensing practices are as varied as the states and provinces regulating them; Derbyshire bluntly describes the situation as "chaos" (128). Perhaps the most notable credentialing trend in the seventies has been considerable study, drafting and passage of licensing laws which attempt to achieve a system which is consistent, logical and effective. This being attempted in a number of ways:

1. The licensing of new categories of health workers is being discouraged (90,155,306,325,414,510,511,512). The usual reasons given are that licensing contributes to rising health care costs and has not been shown to be effective (325). Licensing is considered justifiable only where consumers will be substantially harmed by lack of control and there are no alternative methods of protection. Several sources (90,414,440,510,511,512) establish fairly extensive criteria which should be met before an occupation is licensed.

2. Where licensing is deemed necessary, voluntary licensure (protection of title) is favored if this will afford adequate protection. Reserved title is considered sufficient for professionals who practice in a structured environment subject to ongoing control (90).
3. Where licensing is used, there is an increasing tendency to favor defining the scope of practice (33,90,199,237,441,512). Rozovsky (441), in discussing definition of nursing practice, cautions that attempts to define nursing in law ignores the desirability of change. He suggests that who does what, and when, is often a matter of professional judgement.

4. National licensing exams are favored, with the standard being left to the state or province (136,414,510,511,512). National standards for certification are also advocated (306,414,510,511,512). The trend towards national standards is supported by the increasing sophistication and wide-spread use of equivalency and proficiency exams (156, 271,414,510,512).

5. The feasibility of limited licensure (restriction of practice to a specialty area within the discipline) is explored and usually rejected. While this has been suggested as an approach to basic licensure (58,88,491), the more frequent suggestion is an unlimited basic license, with specialty certification used to designate those able to give adequate care in a specialty practice area (26,306,497,538). Opinions as to whether this should be controlled by law vary.

6. Present mechanisms are being strengthened by providing administrative agencies and boards with adequate staff and budget to effectively implement existing licensing laws, especially the provisions for discipline (90,414,510,512).

7. There is a strong tendency to establish a single licensing authority within each jurisdiction which would oversee all licensing activities, thus increasing the likelihood of compliance with established policy (60,90,237,440,462).
Broader Base of Licensing Control – Trend Two

Kelly states that the background, qualifications and philosophy of licensing board members are of prime importance to the licensing process, as these people evaluate qualifications of applicants for licensure (271). It is therefore not surprising that the composition of licensing boards has been questioned. This has resulted in a move to extend the membership base of the regulatory body beyond the profession being regulated (306).

Most frequently an attempt is made to include consumers on the regulating board. Although Derbyshire (130) states that historically physicians have opposed lay people on boards, this is a noticeable trend in licensure practices and is supported by many groups including medical ones (89,100,164,237,235,414,462,510,511). Just why this is believed to be important is not so clear. Advocates of lay membership suggest that, at the very least, this is a mechanism for public accountability. R.E. Alley, Chairman of the Board of the Consumers Association of Canada, believes lay participation will be more effective if the lay representative must report back to a group. “In my experience, the internal dynamic of working committees or boards is such that all participants get caught up in the common rules of logic and schemes of perception. This is not always or even usual culpable co-option of the lay person, but rather a phenomenon more like acculturation to the milieu” (368). Arguing against lay membership on regulatory bodies is akin to arguing against motherhood, but at least one author flatly states that lay persons add to the cost and not the effectiveness of licensing boards (267).

Less common than consumer participation in regulation of health professionals is the suggestion that allied health groups be involved.
This has been favored by some groups in the United States (164,414,462, 510). Additionally, the Professions Code in Quebec (90,440) provides for an interdisciplinary board and interprofessional advisory council; this board oversees the regulatory activities of each professional group.

Multiple Assessment Procedures - Trend Three

Another prominent credentialing trend which is increasingly discussed in the literature is the use of different types of assessment procedures prior to the granting of a credential. Until recently the almost universal practice has been to award a credential after successful completion of an accredited educational program and a written or oral exam (301). The seventies has seen the emergence of many different requirements and assessment techniques used alone and in varying combinations. Most of the more unorthodox assessment activity has been undertaken by the professions in connection with certification procedures.

In presenting guidelines for physician recertification, Brewer (59) states that assessment should be based on a point system which considers training and present professional standing, post-graduate study, re-examination of basic knowledge in the specialty, and periodic review of surgery performed or patients treated. Reals (406) reports that many specialty societies are making use of a variety of techniques, including examinations, required participation in continuing education, and peer review of practice. The American College of Radiology has adopted the principle of on-site evaluation of physician performance, while the American College of Family Practice is examining the feasibility of records review (486). The American Board of Internal Medicine has adopted a three part certification process which consists of training
in an approved program, on-going evaluation of clinical competence during that program against specified performance criteria, and cognitive testing (46).

The specialty certification program sponsored by the American Nurses Association also depends on a variety of assessment procedures. Basic to the certification process in any of the specialties are the requirements that nurses must have at least two years' experience and be currently practicing in the specialty, must pass a written examination, and must submit, for panel review against pre-established standards, both documented evidence of having given excellent care and references from professional colleagues (23,24,91,148). In a more recent development, the ANA has committed itself to a two-level certification program which will include certifying for competence (adequacy) in a specialty as well as for excellence. Although a masters degree in nursing is pre-requisite to certification at the level of excellence, it is not clear exactly how these two levels of performance will be established or evaluated (26).

The Washington State Board of Nursing has established licensure requirements applicable to two specially licensed categories of Registered Nurse: Advanced and Specialized; these requirements include evidence of participation in continuing education related to the specialty, a resumé of current clinical practice, and a minimum of three references documenting continued clinical competence (538).

Re-Credentialing - Trend Four

There is general support for the idea that any credential which certifies a degree of competence should be subject to periodic renewal based on evidence of continuing competence (1,164,306,333,407,414,480,
The American Board of Medical Specialties, comprised of some twenty-two specialty groups, has adopted a position which supports voluntary recertification of medical specialists; all of the member boards have individually endorsed this. In discussing the ABMS position however, Reals (406) raises a number of potentially problematic issues: older physicians may experience considerable personal difficulty if they are not grandfathered, certifications would have to be dated and the public and professions informed as to the meaning of such dating, there may be a decrease in available medical manpower as physicians prepare to meet recertification requirements, and, finally, there is the as yet unanswered question of whether recertification is worth the costs in the presence of more direct performance monitoring procedures.

Recognizing the problems inherent in any large scale recredentialing scheme, government groups responsible for licensing are inclined to move rather slowly, leaving exploratory work to be done by the more elite and smaller professional groups. A federal United States committee on credentialing of health manpower has recommended adoption of mechanisms which assure continued competence and intensive study of the effectiveness of these mechanisms as an item of high priority (414,510,512). A Michigan Task Force final report on the same subject recommends that each of the professional licensing boards in the state be given six years to experiment with various approaches to assessing competency and to develop a plan for competency-based relicensure (182). The report has not been acted upon.

The literature offers little in the way of definitive proposals or answers on the "how to" of competency-based recredentialing. Child and Zuidema (95) advocate a national registry in which data about
surgical operations would be entered after peer evaluation; cooperation with the registry would be a condition of recertification. In July 1976, the Georgetown University Health Policy Center issued a "model medical practice act", recommending that state medical boards require relicensing every six years, based on evaluation of the quality of care provided by the applicant. While vague on exactly how this might occur, it is suggested that recommendation of a specialty board could be accepted as evidence of competence (131).

Richards (417) has offered a more developed proposal on how to effect competency-based relicensure. He suggests a five step process:

1. for each practitioner, develop a practice profile and select the relevant standards;
2. review his records against standards;
3. examine relevant knowledge or accept attendance at relevant continuing education offerings;
4. test critical skills or accept documented evidence of these skills;
5. review references, complaints, and peer statements for evidence of ethical fitness.

Richards estimates that a three year relicensure cycle would be possible. He believes that in such a system 95% of practitioners should relicense without difficulty, that 1 - 2% would lose licenses, and that the remainder would require remedial work. There is no information as to how these estimates are reached. It does not appear that Richards has considered costs except to say if we believe competency based relicensure is important we must be willing to finance the process and bear the cost of results.
There exists an enormous array of literature which focuses on the idea that requiring professionals to participate in continuing education will ensure their continued competence. The discussion which follows is based on a sample of relevant items; it is not an exhaustive review. The material is organized under the broad headings of background, pros and cons of mandatory continuing education, mandatory continuing education for special purposes, and accreditation of continuing education programs.

Historical Background

When consumerism and human rights became contentious issues in the sixties, the application of these concepts to health care became inevitable. Protection of the public from shoddy health care practices was a challenge to government and the professions. Howard (248) reports that a 1967 United States National Advisory Commission on Health Manpower noted the need for continued learning and concluded that making education opportunities available to professionals would not assure their utilization without incentives such as requiring examination or education for relicensure. While not necessarily agreeing with the conclusion, the American Medical Association did recognize the need for continued learning and in 1968 established the voluntary Physician's Recognition Award as a way of encouraging and commending physicians who carried out a personal program of continuing medical education (326).

The pressures in the seventies to make education opportunities available, or even to make their utilization mandatory, have been enumera-
ated by a number of authors. Among these pressures are demand by consum-
ers for the better care (338,349); a desire on the part of the
professions to impress politicians and pressure groups with their good
intentions to protect the public (92,112,418,463); the need to find ways
of coping with new developments in medical science and health service
delivery patterns (112,338,463); a shortage of resources to implement any
more effective (e.g. competency-based) control system (418); and a belief
that participation in ongoing professional education will decrease
practice errors, decrease the incidence of successful malpractice suits,
and keep insurance premiums down (182,333). Corbally (112) adds to this
list a need to increase education institution enrollments in order to
bolster sagging revenue and a desire to increase professional elitism.
It is instructive to note that concern about cost is as evident as
concern about quality in the pressures which have been identified.
Whether or not these forces individually are legitimate reasons for
requiring participation in education, they collectively have resulted in
considerable legislative activity.

It is difficult to know the extent of such activity in the
North American jurisdictions at any one time. Occasionally articles have
summarized the current state of affairs pertaining to mandatory continu-
ing education in medicine (233,234), nursing (485,537), dentistry (92),
and pharmacy (164). With the exception of medicine however, these appear
irregularly and offer an incomplete picture. The year end "Report on
Medical Education in the United States" in the Journal of the American
Medical Association is helpful for learning of recent developments in
connection with mandatory continuing medical education. The above liter-
ature reveals a recent slow-down in the trend toward requiring continuing
education as a condition of relicensure. Legislation may be discretionary, i.e. allow licensing boards to require education, or directive, i.e. demand such control measures (333,337).

As jurisdictions have sought ways of implementing mandatory continuing education, there has been discussion about appropriate roles for some of the involved groups. Knox (282) states that professional associations should define areas of specialization and related practice standards so as to give direction to educational planning. He also states that various types of sponsors should continue to offer educational programs, as this will provide the practitioner with learning opportunities rich in different perspectives and will keep continuing education flexible and responsive to the needs of the individual. In another article, Knox (280) emphasizes the responsibility of professional associations for supporting members who wish to engage in self-directed lifelong learning. Weir (529) states emphatically that institutions of higher education do not have a role in the policing of professionals, but can make their best contributions by identifying the pros and cons of mandatory continuing education; assisting in the identification of continuing education needs; participating in educational planning; bringing new developments to the practice arena; and contributing to evaluation, research and development in continuing education.

Pros and Cons of Universal Mandatory Continuing Education

The bulk of the literature is concerned with making education mandatory for all health professionals as a condition of relicensure. (A more selective application of compulsory education exists in both theory and practice and is discussed in the next section.) The reasons which
have been advanced for and against mandatory continuing education (MCE) are summarized in tables VII and VIII. It should be remembered that the context for these arguments is the proposal of universal continuing education as a requirement for relicensure.

<table>
<thead>
<tr>
<th>Arguments in Favor of MCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MCE will make education more available and accessible to professionals (385).</td>
</tr>
<tr>
<td>2. MCE will result in better quality education; i.e. improved instructional design (385).</td>
</tr>
<tr>
<td>3. MCE will, at the very least, ensure that practitioners have access to needed knowledge (511).</td>
</tr>
<tr>
<td>4. MCE must be useful, otherwise its popularity would not continue to increase (246).</td>
</tr>
<tr>
<td>5. MCE will satisfy the public demand for accountability; i.e. it will make the professions &quot;look good&quot; (112).</td>
</tr>
<tr>
<td>6. MCE is the only positive action for competence that society can presently afford and achieve (418); in particular it is relatively inexpensive in terms of bookkeeping (130) and continuing availability of manpower.</td>
</tr>
<tr>
<td>7. MCE will make practitioners conscious of the importance of continued competence, thus will act as a motivating factor for improved practice (511).</td>
</tr>
<tr>
<td>8. Participation in MCE is positively correlated with early adoption of innovations which is correlated with a reputation of competence among peers (246).</td>
</tr>
<tr>
<td>9. Although MCE cannot guarantee competence, not having to participate in educational events guarantees even less (111).</td>
</tr>
<tr>
<td>10. MCE has an effect on competence which is analogous to drops in a pond; i.e. our measures aren't discriminating enough to substantiate impact of participation in one offering, but effect is cumulative (98).</td>
</tr>
</tbody>
</table>

Table VII
It can be seen that supporting arguments for MCE range from the reasonable to the illogical and the irrelevant. Most of the arguments are based on opinion rather than evidence. Only the last four try to establish a relationship between MCE and competence, which is the raison d'être for licensure.

Arguments against MCE which appear in the literature are more numerous than those which support it. They can be roughly classified into three broad groups of statements: those which claim that MCE has no effect on competence or is so very limited in its effect as to be unjustifiable; those which claim MCE cannot be effectively implemented, and those which predict that MCE will have undesirable effects on health care:

<table>
<thead>
<tr>
<th>Arguments in Opposition to MCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(those questioning the effect of MCE on competence)</td>
</tr>
<tr>
<td>1. There is little evidence that continuing education makes a difference to competence under ideal let alone mandatory conditions (92,110,164,282,302,486) and absolutely no evidence at all that it will guarantee competence (463,537).</td>
</tr>
<tr>
<td>2. Education is only one variable affecting competence and there are more powerful variables. Continuing education cannot enforce ethical behavior nor protect against inadequate performance not related to knowledge (112).</td>
</tr>
<tr>
<td>3. While continued learning is necessary to competence, that learning may occur in a number of ways of which random education is only one (423); evaluation, feedback and selective corrective action are more effective factors which influence competence (71,247,423,486,511).</td>
</tr>
</tbody>
</table>

...continued
4. While participation in an education program can be legislated, learning cannot. It is an internal process, the result of which can be measured only by testing (37,110,259,302,385).

5. Teachers have had MCE since the early part of the century, and there are many incompetents in the teaching profession (246).

(args arguments claiming that MCE cannot be effectively implemented)

6. It is not clear what constitutes desirable continuing education (108,248), either in terms of quantity (108) or content (259), i.e. relevance to practice.

7. There are not sufficient educational resources to provide continuing education for all professionals (110,248,385,423).

8. Even if demands for education could be met in terms of quantity, to provide education relevant to each practitioner's needs would be impossible (37,108,282,423).

9. To create a system which prevents wholesale cheating would be a bureaucratic nightmare (168).

(args arguments which predict MCE will have undesirable effects).

10. MCE will foster the growth of poor ("mass education" type or substandard local) programs produced to meet heavy demand (182,338,385,423); this will be associated with the rapid development of vested interest groups who will place extra demands on resources (282,385,463).

11. Innovative and potentially effective non-traditional programs will be discouraged (71,282,302,385).

12. More rational approaches to maintenance of competence will not be developed because of false complacency about what MCE can achieve; ritual attendance, "a little like painting the operative site with watercolors" (37) will lend false status and credibility while diverting attention from the real issue (71,92,248,423,463,496).

13. Practitioners may be forced to participate in irrelevant education; rigid requirements will discourage professionals from being individually responsible for their learning and performance (110,164,302).

...continued
14. MCE will increase the paranoia of the majority of professionals who are competent (283).

15. MCE will fail, and in the process, discredit continuing education for the contribution it can make (282,423).

16. MCE will unjustifiably result in the significant loss of medical manpower (182).

17. Education institutions may be held liable when their learners do not perform and insurance premiums will rise (385).

18. The bookkeeping generated by MCE (points, personnel, accrediting systems) will consume too many resources (37,385).

Table VIII

As with the arguments for MCE, the above statements also are primarily opinion and range from the clearly hysterical to the plausible. However, given what is known about learning, education, factors influencing competence, and sheer numbers of professionals involved, the two sets of arguments taken together favor the conclusion that universal MCE cannot assure the competence of health professionals and may not even be the best choice of methods to promote competence. In the vast majority of cases, MCE will affect practitioners already meeting minimum standards and moving toward optimum ones; this is not consistent with the purpose of licensing. "Continuing education we should have --- but making continuing education mandatory for licensure without assurance of its validity for continued competence may be like following Alice into Wonderland where things are not what they seem" (463). Before abandoning the notion of MCE as altogether inappropriate, it is important to consider that MCE has applications other than that of a universal requirement for licensure.
Selective Mandatory Continuing Education.

It appears that on some occasions, MCE can and is being used to some effect for special situations within the licensure system. Sheppard (472) and Castonguay (90) both refer to provision within the Professions Code of Quebec for a professional inspection committee, within each professional corporation, with power to investigate performance and to prescribe such education as it sees fit. Several of the Canadian provinces require refresher courses for nurses who have been inactive, as does Oregon (485,537). New Hampshire and Washington both have special license categories for nurses with advanced training, and retention of these licenses is dependent, among other things, on evidence of participation in relevant continuing education (537,538). Sutherland has suggested that veterinarians be licensed in specialty areas only, and be required to take a transition course before moving to a new specialty area (491);Bradshaw has made a similar recommendation for nursing (58).

The use of mandatory continuing education for non-compulsory credentials is widespread. The AMA's Physicians Recognition Award (PRA) already referred to is the prototype for almost all others (326,333). The PRA requires one hundred fifty hours of participation in continuing education over a period of three years. Of these hours, sixty must be "approved" continuing education, while the other hours may be from a variety of categories including non-approved continuing education programs, medical teaching, papers, books, exhibits, independent study and "other meritorious learning experiences." While the PRA itself is voluntary, organizations such as county medical societies use the PRA mechanism as a requirement for continued membership (309); states requiring continuing education for relicensure accept the PRA as proof of educa-
tional participation (333). The ultimate result is that the PRA is often less than voluntary.

Similar voluntary recognition systems are operated by both medical (268) and nursing (18,485) organizations. Other professional organizations make continuing education a condition of recertification. The most notable examples are some of the medical specialty groups in the United States (406) and the ANA nursing specialty certification programs (91). Technically speaking, continuing education in this situation is voluntary in the sense that practitioners do not require these credentials for practice. The extent to which certifications and other awards are required for many positions, for receipt of research grants, for institutional accreditation etc. is considerable however. In this context, so-called voluntary continuing education becomes de facto mandatory continuing education.

Whether mandatory continuing education is justifiable on a selective basis is a debatable point. One might argue that where direct performance appraisal is a practical impossibility, and the education is relevant to practice and has built-in testing procedures, MCE is a justifiable alternative. On the other hand, once developed, testing procedures could be used without reference to the process by which knowledge and skills were acquired. It might be argued: MCE for recertification in the absence of universal MCE for relicensure could concievably result in a situation described by Knox as a highly favourable one. He would concentrate all available education resources on selected target groups, taking advantage of the diffusion process and what is known about influence of opinion leaders (282).
Accreditation of Continuing Education Programs

Along with the various systems for recording professionals' participation in continuing education, there have developed parallel systems for reviewing and approving continuing education programs. There is little discussion in the literature as to why this has happened. Some reasons given include a desire to protect the professional from inadequate programs (410,447), a belief that education must be of high quality if it is to have a positive effect on the quality of care (17,410), and an expectation that a review process can improve the quality of programs through formative evaluation (410,447). Also, tradition cannot be discounted; accreditation/approval of pre-professional programs has a long history.

Brown (71) distinguishes between accreditation of institutions providing continuing education, which he refers to as fourth level evaluation, and evaluation of programs themselves, which he calls third level evaluation. (Learner outcomes evaluation is second level; patient outcomes, first level). He believes that correlation with impact on care decreases in moving from level one to level four. In the continuing education field the term "accreditation" has increasingly been applied to fourth level or institutional evaluation, while "approval" is reserved for third level or program evaluation. This is a departure from customary usage. As program review systems attempt to cope with the increasing volume of continuing education programs, they tend to develop institutional accreditation procedures (17,326,333).

The AMA's extensive continuing education review system, developed to support the Physician's Recognition Award, has many of the elements of later systems. At the program evaluation level, an approved
program is "---one having sufficient scope and depth of coverage to form an educational unit which is planned, coordinated, administered, and evaluated in terms of a specific educational objective, such as a defined level of knowledge or a specific performance skill, to be attained by a physician participating in the program" (326). At the institutional accreditation level, administrative educational methods and evaluation practices are scrutinized (114). At the end of 1976, over eight hundred institutions were AMA accredited; this included medical schools, community hospitals, specialty societies, clinics and federal government facilities. Cost to the institution was $500 for initial survey and $250 for resurvey (333).

The nursing profession in the United States has developed a similar two level continuing education review system wherein the American Nurses Association approves programs directly in some instances, but more often accredits other organizations such as state nurses associations, specialty organizations, federal nursing services and universities. These organizations are then authorized to offer programs said to be ANA approved. With the exception of universities, the accredited organization may also review and approve offerings of other sponsors (410). Institutional accreditation criteria are similar to those used in the AMA's PRA; program approval criteria are detailed and would require well developed program plans. In Canada, at least three provincial nursing organizations operate continuing education evaluation systems at the program review level (410,411,453). All use detailed program planning criteria and two of these (410,411) ask reviewers to make a terminal judgement as to whether nursing care will be improved as a result of the program.
Dentistry (300), pharmacy (164) and medical technology (77) are other professions which have instituted continuing education approval procedures. The program developed by the American Society of Medical Technologists (77) is notable for its emphasis on helping program sponsors: "It is not the policy of the committee to refuse accreditation of programs deficient in any of (these) areas, but rather to offer constructive suggestions so that sponsors may upgrade the program offering to enable it to meet (our) standards of quality and effectiveness."

Formative program evaluation is also heavily emphasized in the review system sponsored by the Registered Nurses and Registered Psychiatric Nurses' Associations in British Columbia (410).

Like all systems of credit, continuing education approval systems require a currency or unit of value. While the learner-educational event contact hour is commonly used, the continuing education unit is employed also. Developed outside the health system to meet demands for a common language to communicate the value of a continuing education event, the C.E.U. is defined as "ten hours of participation in an organized continuing education experience under responsible sponsorship, capable direction and qualified instruction" (109, 209, 266). A further elaboration of just what this means is provided by Bruno (73), who states that organizations awarding C.E.U.'s should require that the event be offered in response to identified need; that there be a statement of objectives, purpose, and rationale; that content be organized and presented in a logical sequence; that clientele be included in planning; that appropriate evaluation procedures be used, etc. These criteria are quite similar to those used in program level review in most of the continuing education approval systems in the health care field.
If there has been any attempt to evaluate the effect of program review systems on quality of care or even in terms of reliability and improved educational programs, this did not emerge in the literature searched. Information on the costs of such review programs is not readily available either. In view of fears that such review systems may inhibit effective programming, this failure to evaluate is quite surprising.

Summary and Comment

The central issue in maintaining competence of health professionals is definition and measurement of competence, followed by action calculated to restore and/or maintain it. Surrounding this issue is the North American social context. Developing evaluation systems have particularly influenced and been influenced by professionals' human needs for success and respect and their civil rights to due process; consumers' rights to privacy and to participation in decisions about personal health care and about public health policy; growing sophistication on the part of all groups about the application of power through collective action; the rising costs of health care; the growing volume of malpractice litigation; and the proliferation of categories of health care workers.

The above factors have generated a prodigious amount of activity in the various health manpower credentialing mechanisms which attempt to make some statement about the professional's competence. The major credentialing methods applicable to individuals are licensure (to assure a minimum standard of competence with a view to protecting the public from harm) and certification (usually used to indicate achievement
of a level of excellence within a specialty). Certification of program completion by educational authorities is a supplementary credentialing mechanism. The more prominent credentialing trends include action to bring order to a jurisdiction's licensing policies and procedures; to extend control of licensing administration beyond the profession concerned; to use a variety of evaluation techniques prior to awarding or renewing a credential; and to limit the time for which a credential is valid, imposing conditions for renewal which are thought to relate to competence.

Mandatory continuing education (MCE), which is one of the more common conditions for relicensure, has been particularly contentious. Most arguments for or against MCE have been based on opinion rather than research or experience, however the weight of argument at this time indicates that the costs of universal mandatory continuing education for relicensure will outweigh its effectiveness in assuring a minimum level of competence. More research on the impact of continuing education, on the influence of voluntary vs mandatory systems, and on the effect of approval mechanisms is needed.

In reviewing the social context within which health professionals' competence must be maintained, it becomes apparent that the technical implementation of evaluation and action programs is heavily influenced by surrounding issues. A developing program must recognize and work with such issues if it is to be maximally effective.

Finally, it should be reiterated that the literature reviewed in this chapter represents a very cursory overview of material tangentially related to maintaining health professionals' competence. Most of the subjects discussed could well be the subject of extensive literature survey alone. They are presented here to provide necessary perspective.
Chapter VIII

OBSERVATIONS AND RECOMMENDATIONS

This concluding chapter will not summarize the findings reported in the thesis. The reader is referred to the final "Comment and Summary" sections of Chapters II - VII for that information. Chapter VIII will first offer some general comments and then discuss the influence of the North American political reality, the state-of-the-art of evaluation-action strategies, and the contributions that continuing educators can most effectively make to the evaluation and maintainance of health professionals' competence.

First, a brief comment should be made about the basic assumptions that published literature describes recent developments and current practices in the field, and that the four health professions studied are the richest information sources. With respect to the first assumption, it became obvious during this study that the literature in this rapidly developing area has lagged behind current developments to a significant extent. This is reflective of Garvey's findings that time lapse from work initiated to journal publication is thirty to thirty-seven months (181). Those wishing to know about current work in the field of evaluating and maintaining competence of health professionals should make extensive use of personal contacts by telephone and at conferences. They should also request and be prepared to pay for technical reports and pre-publication drafts of final reports of work underway. While the second assumption proved to be an acceptable one, in retrospect, the bulk
of useful information was found in medical and nursing literature. Developments in these two disciplines should be followed carefully.

One final general observation: the language used in the literature is confusing and unhelpful, perhaps because the field is still in its formative stages. To date there is no real consistency or identifiable pattern to the terminology employed by authors even within the same profession. Particularly vulnerable to both conceptual and semantic confusion are the following sets of terms: structure, process, outcome; variable, indicator, criterion, standard, index; measurement, evaluation; minimum, optimum (standards); licensure, registration. Until this difficulty is resolved, it will remain important to be alert to the concepts being discussed in any given report, whether or not definitions have been presented. Similarly, writers should be meticulous in both definition and use of terms.

The competence of professionals is fundamentally a social issue. This demands that the goal be protection of the public and the means to that goal be acceptable to general society on psychological, social, political and economic grounds. Recent history indicates that an increasingly sophisticated public is accepting responsibility for personal health state and public health policy. The general social climate emphasizes civil rights of both patients and practitioners, while political and economic forces demand cost control and accountability. Jago (257) reminds readers that society now seriously questions the professional dominance model in health care.

These influencing factors have converged so that traditional structures such as tort law and credentialing mechanisms have been applied with increased vigor; in many instances they have been revised to
enhance effectiveness. This has resulted in salutory changes, such as the more efficient use of disciplinary boards; in changes where the effect is as yet unknown, such as the addition of consumer members to regulatory bodies; and in changes where desired impact is highly questionable and indeed may be counterproductive, as is the case with universal mandatory continuing education laws.

Concurrent with traditional mechanisms being strengthened, more innovative approaches, primarily practice-based evaluation-action programs, are being developed. While these hold considerable promise for the effective maintainance of competence and are therefore of considerable interest to continuing educators, they will be useful only to the extent that current technical and developmental problems are resolved in a manner consistent with existing values held by the general society. When these requirements are considered together, a number of important recommendations for the continued development of evaluation-action systems emerge:

1. Clearly worded and agreed upon standards are prerequisite to a sound competence assessment system (1,247). The professions must pursue clinical research aimed at establishing definitive correlations between structure, process, and outcome elements. Where correlations exist, evaluation of health practitioners' performance against explicit standards is possible. Where these are in doubt, the use of essential, expert identified criteria is probably the best interim strategy. Evaluators should be aware of the relative advantages and disadvantages of structure, process, outcome and combined approaches to evaluation; decisions about approach should be made with respect
to the evaluation problem at hand, not based on dogmatic generalizations about a preferred approach. Where outcomes evaluation appears feasible, efforts should be made to effect cooperative multi-disciplinary patient care review.

2. Both research and more vigorous implementation of the latter stages (problem analysis, corrective action and re-evaluation) of the evaluation-action cycle are needed. So far, there has been more quality assessment than quality assurance. Tippett and Smith (502) remind us that committees all too often fall into the trap of over-spending on the review process and underspending on follow-up action. Failure to re-evaluate to ascertain impact of corrective action or permanence of achieved changes are related problems.

3. More attention must be given to evaluation-action programs which accommodate to both the unstated values of the larger society and the stated objectives of the agency/practitioner. It is particularly important to avoid being so absorbed in evaluating the technical aspects of care that we fail to attend to such other aspects as interpersonal relations, patient comfort and health teaching behaviors. Menninger (334) reminds us that the subjective aspects of caring, the patient's emotional responses to his care, are an important part of quality. Work on methods for evaluating these components of performance should continue.

4. The current oversights in evaluation of technical competence must be corrected. Among the more significant of these are the failure to
identify missed diagnoses (e.g. the sick patient treated as healthy) and the excessive reliance on inadequate records without ever asking whether the procedure/observation recorded was, in fact, performed safely/accurately etc. Pellegrino (383) points out that to date we have tended to monitor correctness of decisions more than correctness of performance.

5. Development of information systems should be accorded a high priority. This includes provision for efficient collection of data relevant to established criteria, as is possible for instance with use of the problem oriented record, and for effective storage, retrieval and sharing of problem-based criteria sets.

6. More attention should be given to testing the reliability of criteria chosen for use in evaluation systems. Unless this is attended to, we run the risk of engaging in change for the sake of change.

7. Increased emphasis must be placed on cost-related implications of evaluation-action systems. Cost and quality are inseparable. Despite the attempts in this paper to address quality in isolation, the reality is that inappropriate, unnecessary and inefficient activity in the health care system consumes irreplaceable resources. Costs of evaluation-action programs must be carefully documented. Cost-related aspects of performance, including the seldom measured quantity of work, should be considered a legitimate focus for evaluation. Effort must be made to select evaluation topics where there are known problems; to date much of the recorded evaluation
activity appears to be a great deal of expensive self-justification. Without question, competence or adequacy of health professionals' performance can and eventually must be viewed from the cost/benefit perspective.

In the last analysis, whether traditional or more innovative mechanisms are employed, the challenge of monitoring and maintaining competence of health professionals has been and will continue to be influenced by social context. In discussing earlier failures of the health professions to face evaluation issues squarely, Christoffel (96) suggests that rigorous evaluation of and action on competence is possible only if the expectations of society are respected. He paraphrases historian Henry Sigerist: "It is a sheer waste of effort to oppose powerful societal trends --- we can influence developments --- but we can do so only in certain directions. And history tells us what these directions are."

Given the volume and variety of approaches today emerging as mechanisms which have potential for contributing to maintaining and monitoring health professionals' competence, it is important to appreciate the relationships between these: effective and efficient evaluation—action programs can render the credentialing process more valid, and these taken together can render the disciplinary aspects of licensing and the malpractice litigation system less frequently necessary. Continuing educators can make some important contributions to an emerging and interrelated social system for monitoring and maintaining competence:

1. In testing of knowledge and skills (structure), promotion and careful design of combined self-assessment and follow-up study programs can assist the professionally motivated individual to continue profes-
sional development in essential knowledge areas. Additionally, structural assessment can be used as one factor in a screening program for certification and placement of personnel.

2. In assessing process and outcomes of care, whether or not the approach is general or situation-specific, continuing educators can play a supportive role in various stages of an evaluation-action program. Dennis (125) has identified at least four categories of change agent behavior which contribute to effective programs: catalyst, resource linker, process helper and solution giver. His thoughtful conceptualization, reproduced as figure 2, can provide guidance to continuing educators wishing to facilitate the evaluation-action process.

![Diagram](image)

**Fig. 2**

3. The limitations of continuing education, both as a corrective action within the evaluation-action process and as a guarantor of competence
within the credentialing process, require continuing interpretation by educators familiar with research relevant to these issues.

4. Competence, which is the focus of this paper, is concerned with adequacy in practice. Continuing education in this instance will often be remedial or at least primarily concerned with performance deficits. Ellis (158) reminds us that education programs which arise from identification of performance deficits should not be considered a replacement for programs designed to diffuse new knowledge or to enable people to assume higher levels of responsibility. In these latter instances continuing education planning which proceeds from an adoption model or traditional training approach may be preferred.

5. It is imperative that the costs of evaluation-action or traditional continuing education programs be considered as dollars and professional time irrevocably lost to direct health care services. A decision to proceed with programming, i.e. to consume available resources, can be justified only to the extent that probable benefits are commensurate with estimated costs. Cost and benefit of implemented programs must be documented. Increasingly society will accept no less than this.

It can be concluded that continuing education has an important, but certainly not exclusive or even dominant, contribution to make to evaluating and maintaining competence of health professionals. Cyril Houle (246) summarized this well when he wrote,"

"All available evidence seems to show that lifelong education improves practice (this of course is a broader conclusion than that reached by the various studies showing no learning from one event or program). As yet it is only an article of faith that the addition of compulsion does much
to change the situation. The center of our concerns, therefore, must be
the creation of forms of assessment which measure, or are provably cor­
related with, what both professionals and general society demand to know:
how well does every licensed or certified practitioner maintain an alert
continuing performance of his or her duties in as effective a fashion as
the current knowledge base permits? To the extent that there is any
deficiency in that performance, how can it be remedied?"

To this question can be added a final one: To the extent that the remedy
is an educational one, how can continuing education be delivered to
support effective and efficient learning? This last question has been
explored vigorously by the discipline of adult education. But the first
two questions must be addressed also. It is hoped that this thesis has
contributed information which will assist in answering them.
Literature Cited


44. Bellin, L.E. "If We Know What Quality Care Is, Why Don't We Get It?" Modern Hospital 119 (December 1972). Pp. 87-88.


304. Lindeman, Carol A. "Measuring Quality of Nursing Care: Parts I and II." Journal of Nursing Administration 6 (June 1976), pp. 7-9 and (September 1976), pp. 16-19.


400. Professional Code, Statutes of Quebec 1973, c.43; as amended by 1974, c.6 and c.65; 1975, c.80 and c.81.


