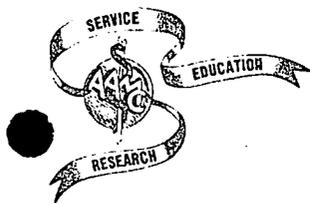


Ms. Littlemeyer



ASSOCIATION OF AMERICAN MEDICAL COLLEGES
SUITE 200, ONE DUPONT CIRCLE, N.W., WASHINGTON, D.C. 20036

June 25, 1975

M E M O R A N D U M

TO : The Council of Deans

FROM: Joseph A. Keyes

Attached are two documents related to the Council of Deans Spring Meeting in Key Biscayne. The first is a compilation of the six papers presented on behalf of the Council to the President's Biomedical Research Panel. The Panel members have repeatedly expressed their appreciation for the lucid and insightful presentations prepared for them by the Council. We trust you will find this document of interest and utility.

The second document is a draft of the Council's business meeting minutes which chronicles the deliberations on the NBME GAP Committee report and the response of the AAMC Task Force.

JAK/jsp

ASSOCIATION OF AMERICAN MEDICAL COLLEGES

COUNCIL OF DEANS BUSINESS MEETING

Minutes

April 30, 1975

8:30 a.m. - 12:00 Noon

Biscayno Room, Sonesta Beach Hotel
Key Biscayne, Florida

DRAFT

I. Call to Order

The Council of Deans Business Meeting was called to order by its Chairman, Ivan L. Bennett, M.D., at 8:30 a.m. The presence of a quorum was noted.

II. Approval of Minutes

The minutes of the November 12, 1974 Business Meeting were approved without change.

III. Chairman's Report

Dr. Bennett reported that the Council of Deans would continue to receive its Administrative Board's draft minutes and welcomed agenda item suggestions from the Council for the Board's consideration at their quarterly meetings. He indicated that correspondence regarding questions on draft Board minutes or possible discussion topics for Board meetings be conveyed to Dr. Marjorie P. Wilson, Director, Department of Institutional Development, AAMC.

IV. Action Item

Consideration of the AAMC Task Force Report on the recommendations of the NBME GAP Committee

The Council of Deans examined each of the GAP Committee's major recommendations in light of the Task Force's response and the subsequent reaction of the CAS and OSR.

Council discussion of each GAP Committee recommendation is as follows:

Recommendation #1: The NBME should abandon its three-part system of examination for certification for licensure.

Discussion of this recommendation by the Council reflected varying points of view regarding the continued usefulness of the three-part NBME examination for licensure.

Those who believed it should not be immediately abandoned cited the need for a single national standard for licensure and the importance of a nationally accepted standard of quality that medical schools can point to when defending medical education to the public, courts, and legislature. The acceptability of the exam as a standard, one council member suggested, has not been eroded, as many critics claim, as evidenced by the increase in the number of medical schools requiring the National Boards for graduation from 22 to 33.

Part I of the exam was praised for its practical use as an evaluative tool, both for use in "weeding out" undesirable students and for use as an indicator of acceptability for transfer after 2 years for students from foreign medical colleges to U.S. colleges.

Supporters of the National Board exam admit that it may have deficiencies but indicate that mechanisms exist for revision and that if modified, it can continue to perform its function as a criteria for licensure.

Proponents for abandonment of the National Board three-part exam believe that the exam has outlived its usefulness and no longer fulfills the function of being the sole standard for licensure. They point to the fact that the FLEX exam has become accepted in forty-eight states as an authoritative examination for licensure.

Part I was criticized for its tendency to require conformity to a standard kind of basic science curriculum. It thus discourages experimentation and innovation with basic science curricula. Additionally, it reinforces an attitude among students that basic sciences can be put aside and "forgotten" after 2 years of study. It was suggested that a test which examined a student's knowledge of basic medical science given at the time of awarding the academic degree would be an advance toward solving these problems.

Dr. Janeway, a member of the Advisory Committee for Undergraduate Education for the National Board, described the advisory committee's position regarding the GAP Task Force report. The committee concluded and recommended to the National Board that the three-part examination continue to

be made available as is suggested in the Task Force Minority Report by Carmine Clemente. The Advisory Committee also considered the feasibility of the formation of a criterion-referenced evaluative qualifying examination designed to assess clinical competency and related basic science knowledge for entrance into graduate medical education. Although the exam would not be related to the licensure process, Dr. Janeway admitted that, if the new exam proved effective and became generally accepted, the three-part exam might be in effect "abandoned". It was Dr. Janeway's opinion that the uniform adoption of a single set of pathways related to licensure, whether it be FLEX or another exam, would be the best way to come to grips with assessing quality in the educational process.

It was the consensus of those deans present that the maintenance of a national standard for quality and licensure was important and therefore whatever its defects the three-part system should not immediately be abandoned.

ACTION: On motion, seconded and passed, the Council of Deans voted to concur with the CAS substitute recommendation which reads, with a COD wording change (see underlining), as follows:

The Task Force believes that the three-part system should not be abandoned until a suitable examination has been developed to take its place and has been assessed for its usefulness in examining medical school students and graduates in both the basic and clinical science aspects of medical education.

Recommendation #2: The NBME should continue to make available norm-referenced exams in the disciplines of medicine now covered in Parts I and II of the National Board.

The CAS recommended that if one agrees with the substitute recommendation in #1, then by reason of logic, #2 should be deleted.

ACTION: On motion, seconded and passed, the Council of Deans voted to delete GAP Committee Recommendation #2.

Recommendation #3: The AAMC, NBME and other interested agencies should assist the schools to develop more effective student evaluation methodologies.

Discussion centered on whether the Council should adopt the Task Force recommendation which concurs with and extends the Committee recommendation by emphasizing the role of the LCME in examining methods of student evaluation in the accreditation process or adopt the CAS substitute recommendation which also emphasizes the role of the LCME but which would require schools to provide evidence to the accrediting body of the schools utilization of external evaluation in the assessment of the educational achievement of their students.

It was the CAS phrase "external evaluation data" that concerned many deans.

Dr. D. Kay Clawson, who was a member of the CAS Administrative Board when this recommendation was formulated, described the underlying rationale for the inclusion of an "external" check on medical schools.

The CAS concern was not with the well established medical school with a history of careful review of student performance by its faculty but with what appears to be the development of new medical schools whose origins have a "political" base and not a firm university base. In these schools the CAS felt that an external check would encourage and set criterial for appropriate quality assessment of both faculty and student performance.

Although a minority of deans expressed agreement with the CAS recommendation and many approved the sentiment behind it, a majority of deans believed that the recommendation was misdirected. It was the feeling of the Council that the AAMC would in reality be approving the establishment of an external standard for medical school assessment and open the door for increased political interference in the evaluation process.

ACTION: On motion, seconded and passed, the Council of Deans voted to accept the Task Force response which reads:

The Task Force concurs [with the GAP Committee recommendation] and recommends that the LCME place a specific emphasis on investigating schools' student evaluation methods in its accreditation surveys.

Recommendation #4: The NBME should develop an exam to be taken by students at their transition from undergraduate to graduate education for the purpose of determining students' readiness to assume responsibility for patient care in a supervised setting.

The Council of Deans in discussion of recommendation #4 addressed itself to two basic questions. The first, whether there should be created a qualifying examination for determining entrance into graduate medical education was discussed and acted upon at the 1974 Spring Meeting in Phoenix in the narrower context of the FMG Report which had as one recommendation that a standard qualifying examination be created and required as a prerequisite to entrance into intern or residency programs in the U.S.

At that time, the Council acted in favor of this recommendation. Dr. Bennett suggested that the Deans carefully consider the idea of requiring a qualifying exam both in light of the FMG and the GAP Report so that the Council could formulate a consistent position on this much debated question.

In the discussion which followed some important questions surfaced which were of major concern to the Council and for which no ready answers were apparent:

1. Since the qualifying exam would not be linked to the licensure process, what are the alternatives for an American graduate who fails the qualifying exam and goes directly into practice without additional education in those states not requiring an internship for licensing? What impact will this have on the health care system?

It was suggested that the examination be given early enough so as to permit adequate time for remediation for those not passing the exam.

2. Who bears the burden of remediation? If the schools were to bear the burden and set up special programs then they would have to be notified of the scores. Yet the OSR and others urge that the school not be informed of the results. Is it realistic to expect the student to bear the burden? As a practical matter, it was suggested that it would fall to the schools to look after their own graduates until they had passed.
3. What about the FMG's who do not pass? Should there be a Fifth Pathway? Is it a responsibility of American medical schools to offer remediation to FMGs? Do we let them practice without the needed experience gained from a graduate program?
4. Should passing the qualifying exam be made mandatory for only FMGs or also a prerequisite for American students? It was suggested that in the interests of fairness and a desire for a national standard of quality the exam should be given to all students.
5. If mandatory for all then what will be the fate of Part I and Part II of the NBME exam which is required in many schools? Will students be required to take both?
6. If allowed the option of substituting one for the other then what kind of legal problems surface when one substitutes a norm-referenced exam for a criterion-referenced exam?
7. What effect will a qualifying exam have on the mechanics involved in applying for entrance into graduate medical education programs and subsequent acceptance? What effect will it have on the matching program?
8. Does one pass or fail the test or will it be purely evaluative--similar to a "super" MCAT?
9. What will be the effect of the qualifying exam on the present movement toward emphasizing continuing education?

After substantial discussion of these questions, not all of which appeared resolvable, Dr. Bennett framed a series of questions for a vote.

1. Should such a qualifying exam be developed?

ACTION: Unanimous approval

2. Should this examination be a "necessary but not necessarily sufficient" condition for entry into graduate medical education programs?

ACTION: Unanimous approval

3. Should this examination when developed be interchangeable with the National Board Parts I and II?

ACTION: Unanimous approval

After these actions, the question was raised whether the Council had intended that a passing grade be required, or only that the exam be taken, with the score being one criteria upon which admission to graduate programs would be based. Discussion disclosed disagreement and a vote was taken.

4. Should a passing score be required?

ACTION: Yes, by a margin of 2.5 to 1.

Thus, the action on this matter can be summarized:

The Council of Deans voted to approve the formation of a qualifying examination, passing of which, will be a necessary, but not necessarily sufficient qualification for entrance into graduate medical education program. Passage of Parts I and II of the National Boards may be accepted as an equivalent qualification for passage of such an exam when it is developed.

N.B. The requirement that a passing grade on such an exam be achieved as a prerequisite to entrance into graduate medical education was the most vigorously contested element in the COD recommendation.

Recommendation #5: The Federation of State Medical Boards and their members should establish a category of licensure limited to caring for patients in a supervised graduate medical education setting.

Morning Session - April 29

8:30 - KEYNOTE ADDRESS:
9:00 a.m. "ACADEMIC PROGRAM CHOICE
IN A CHANGING SOCIETY"

Steven Muller
President
Johns Hopkins University
and Hospital

9:00 - General Discussion
9:15 a.m.

9:15 - FACULTY ASSESSMENT
10:00 a.m.

Hilliard Jason
Director, Division of
Faculty Development AAMC

General Discussion

10:00 - COFFEE
10:20 a.m.

10:20 - PROGRAM EVALUATION
12:00 Noon "EDUCATIONAL OUTCOMES
IN PERSPECTIVE"

John W. Williamson
Professor of Health Care Organization
School of Hygiene and Public Health
Johns Hopkins University

General Discussion

12:00 Noon- PANEL DISCUSSION
12:30 p.m.

Christine McGuire
Lee S. Shulman
James V. Grieson
Anthony Voytovich
Hilliard Jason
John W. Williamson

Evening Sessions - April 29

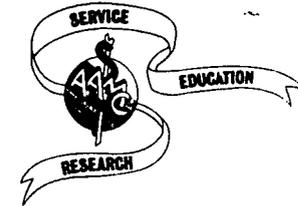
8:30 - A DISCUSSION WITH
10:30 p.m. THE AAMC PRESIDENT

John A. D. Cooper

Morning Session - April 30

8:30 - BUSINESS MEETING
12:00 Noon

12:00 Noon Adjournment



**ASSOCIATION OF
AMERICAN MEDICAL COLLEGES
COUNCIL OF DEANS
SPRING MEETING**

**ACADEMIC
DECISION-MAKING:
ISSUES
AND
EVIDENCE**

April 27-30, 1975
The Sonesta Beach Hotel
Key Biscayne, Florida

**COUNCIL OF DEANS
SPRING MEETING
April 27-30, 1975
The Sonesta Beach Hotel
Key Biscayne, Florida
Biscayno Room - North**

PROGRAM

**"ACADEMIC DECISION-MAKING:
ISSUES AND EVIDENCE"**

Evening Session - April 27

12:00 Noon- ARRIVAL AND REGISTRATION
7:30 p.m. Hotel Lobby

7:30 - GENERAL RECEPTION
9:00 p.m. Ballroom Patio

Morning Session - April 28

8:30 a.m. WELCOME AND OVERVIEW OF
THE MEETING
Ivan L. Bennett, Jr.
Chairman, Council of Deans

8:45 - KEYNOTE ADDRESS:
9:15 a.m. "EDUCATING PHYSICIANS
AND SCIENTISTS —
CHALLENGES AND
OPPORTUNITIES"

William D. McElroy
Chancellor, University of
California, San Diego

9:15 - General Discussion
9:30 a.m.

9:30 -
9:50 a.m.

**"EVALUATION FOR DECISION-
MAKING — A CONCEPTUAL
FRAMEWORK"**

Christine McGuire
Chief, Evaluation and Research
Center for Educational Development
University of Illinois
College of Medicine

9:50 -
10:00 a.m.

General Discussion

10:00 -
10:20 a.m.

COFFEE

10:20 -
12:00 Noon

STUDENT ASSESSMENT

**"INFORMATION SYSTEMS FOR
MONITORING STUDENT PERFORMANCE"**

James V. Griesen
Director, Office of Educational
Resources and Research
University of Michigan
Medical Center

**"EVALUATING PROBLEM
SOLVING SKILLS"**

Lee S. Shulman
Professor of Medical Education
Office of Medical Education,
Research and Development
Michigan State University

**"CLINICAL PERFORMANCE
ASSESSMENT THROUGH
RECORD AUDIT"**

Anthony Voytovich
Assistant Professor of Medicine
School of Medicine
The University of Connecticut
Health Center

General Discussion

12:00 -
1:00 p.m.

FACULTY ASSESSMENT

Hilliard Jason
Director, Division of
Faculty Development AAMC

General Discussion

Evening Session - April 28

8:00 -
10:30 p.m.

BUSINESS MEETING



THE ASSOCIATION OF AMERICAN MEDICAL COLLEGES
MEMBER COLLEGES
1954-1955

Meeting, April 29, 1955
11:30 a.m. - 1:30 p.m.
Royal Macomber Hotel
Ray City, Florida

ASSOCIATION OF AMERICAN MEDICAL COLLEGES

One Dupont Circle, N. W.
Washington, D. C.

MEETING OF THE COUNCIL OF DEANS
with
THE PRESIDENT'S BIOMEDICAL RESEARCH PANEL

April 29, 1975

Cape Florida Room
Royal Biscayne Hotel
Key Biscayne, Florida

<u>Title</u>	<u>Speaker</u>	<u>Page</u>
Biomedical Research and Medical Education: The Institutional Setting	Julius R. Krevans, M.D.	1
Establishing a Biomedical Research Base in Developing Institutions	Andrew D. Hunt, M.D.	5
Overhead: Rationale and Reality in 1975	Richard L. Seggel	9
The Impact of Centers and Targeted Research on Academic Medical Centers: Institutional Issues	Stuart Bondurant, M.D.	13
Biomedical Research Resources: General Support, Clinical Research Centers, Research Manpower Training	Chandler A. Stetson, M.D.	21
Organization of the Federal Research Enterprise: Policies and Leadership	Ivan L. Bennett, Jr., M.D.	27

BIOMEDICAL RESEARCH AND MEDICAL EDUCATION:
THE INSTITUTIONAL SETTING

Julius R. Krevans, M.D.
Dean, U. of California - San Francisco
School of Medicine

I have been asked to address the problem of the institutional setting for biomedical research. I will not address the question of why do biomedical research (which I suspect would be carrying a bucket of coals to the Town Council of Newcastle from looking at the members of the Panel) but will address the question "why do research in an academic medical center?" I have prepared a very simple list of reasons why I think this should be done, illustrating these (without trying to be immodest) from my own personal experience. I will not beseege you with data, which I am certain you will accumulate or already have accumulated in enormous volume.

The first reason which I believe is critical, is that the fact that research is conducted in an academic medical center affects the nature of the questions asked. The unresolved problems (questions) of health and disease occur to the physicians who are responsible for the care of patients. They are shared with the participating scientists who are examining those questions. As a perfect example of this, one of the earliest involvements I had in biomedical research was in caring for patients who developed an abnormal tendency to bleed when they received huge volumes of transfused blood.¹ This clinical observation led me and some of my colleagues to begin to question the dogma that platelets were inert bits of protoplasm and these questions, in turn, led not only to an improved capacity for the care of patients, but also led to an increased understanding of the fundamental nature of the blood platelet and eventually opened up a substantial degree of research carried on by both our colleagues in Baltimore and many others.

A second major reason that it is an advantage to carry on this research in an academic medical center is that it puts in juxtaposition the physician-scientists who are doing research at the interface of the clinical problem with basic science and basic scientists who are examining very fundamental issues. Perhaps the best example of this is the interchange that probably took place between William Castle and Linus Pauling in which the clinical problem of the inherited blood disorders, sickle cell disease in particular, led Pauling to ask some fundamental questions

¹Krevans, J.R. and Jackson, D.P., (1955). A hemorrhagic disorder following massive whole blood transfusions. Journal of the American Medical Association, 150

about the structure and synthesis of protein. These observations have increased our understanding about the molecular basis of disease.

A third reason for conducting biomedical research in an academic medical center in which the total range of questions are being asked is the effect of the research environment on physicians who are responsible for patient care. Again, to draw an example from my own experience. There was a young internist working in Liverpool named Ronald Finn who was asked to develop clinical research projects in the setting of an interest in genetics and some of the fundamental problems of genetics. He undertook to examine why women become immunized to the RH factor causing hemolytic disease in the newborn. The result of his questioning this process was principally responsible for what will be the elimination of the disease in one generation.^{2/}

The fourth reason that I think that it is very important to conduct biomedical research in an academic medical center has to do with one of the accusations which has been levelled at the whole process in the past; namely, that there is too long a time between the development of knowledge and the translation of that knowledge into improvement in the care of patients. One can contrast the extraordinary length of time it took for our understanding about electrolyte problems and their cause of disease and death in patients to become part of the clinical care of patients with the RH work, which began in 1962 and was in place at the bedside by 1968. I believe it is the performance of biomedical research in a setting in which care of patients is underway, which facilitates the translation of knowledge into the actual care of patients. The presence in a single institution of this continuum, from the most fundamental research to the responsibility for the actual care of patients makes this rapid translation of medical progress possible.

And last, but by no means least, the next generation of our physician-scientists, the men and women who will be in the position to ask the questions and examine the unresolved problems, are people who are in educational institutions. If one lists as an alternative to the conduct of biomedical research in academic medical centers, the establishment, let us say, of free-standing research institutes, then you preclude the influence of biomedical research on the young men and women upon whom all of us, and all of the citizens of our nation will be dependent for these efforts in the future.

^{2/}Finn, R., Krevans, J.R., and Clarke, C.S. (1962). An approach to the prevention of RH hemolytic disease. Journal of Clinical Investigation (abstract), 41, 1358.

I will speak very briefly about the interrelationship that I feel exists between the presence of an active research program in an educational setting and education. We all agree that in medicine one of our major educational objectives is to produce an individual who will remain to a substantial degree a scholar all of his or her life. Because of the rate of change of new knowledge, the actual care of the physician's patients is going to depend upon a scholarly attitude. What better way of imbuing the physicians we educate with the attitude that scholarship is important than to have the faculty who are responsible for helping them learn medicine be creative scholars in their own right? I think the difference between preaching scholarship, and allowing the students to share in the practice of scholarship, is a self-evident and an extraordinarily important reason why medical research should be conducted in institutions in which educational programs are underway. This more clearly than any other mechanism that I can think of, exposes the student to that interface between what we now know and what we do not know and imbues the student with the appropriate attitude and insight about how to solve the unsolved problems. The whole approach to accumulating data, sorting the data and evaluating data, which is the bedrock of the practice of good medicine is best taught by illustration with the faculty actually involved in that same kind of process.

One could list a brief set of negatives for biomedical research occurring in an academic medical center. I think it is susceptible to abuse. One could see the possibility that the research might crowd out some of the other essential efforts that ought to be going on in academic medical centers. One could envision that major research efforts might distort some of the academic programs that a center would undertake. Cost sharing, which we shall hear more about later today diminishes, to some degree, institutional discretion. But all of these are abuses to which the system might be susceptible and from my point of view, are far outweighed by the advantages to both research and education in conducting biomedical research in the academic medical centers. I find it hard to envision what a medical school would look like of no research or very little research were undertaken in these centers. I think they would be sterile places producing individuals who would be obsolete almost from the day that they entered their active practice.

Finally, I would say that the rationale which led this nation to put in partnership a Federal initiative to expand our knowledge about medicine and the basic processes which affect

health and disease, with the academic institutions, was a very intelligent and a very successful decision. I think it would be a serious blow to both research and to the institutions which have shared in its development if there were major changes in this partnership.

ESTABLISHING A BIOMEDICAL RESEARCH BASE
IN DEVELOPING INSTITUTIONS

Andrew D. Hunt, M.D.
Dean, Michigan State University
College of Human Medicine

Since 1964, 24 new medical schools have been established and been awarded either full or provisional accreditation by the Liaison Committee on Medical Education. Several more are in various stages of development, with Letters of Reasonable Assurance, and going through the accreditation process.

The spate of new medical schools since 1964 has been in response to the publicly felt and expressed need for more physicians, and was stimulated first by the availability of Federal funds for construction (1963), and later by the addition of Federal start-up funds for operations in 1971.

These schools, in the main, are either intimately related to or sponsored by existing universities. A few have developed from large pre-existent medical centers or hospital consortia, adding medical education to a well established base of patient care and research.

Three of the 24 accredited new schools have hospitals which they own and operate. The others have various sorts of arrangements with community hospitals, which while creative and conducive to marked upgrading of quality of care, often pose significant constraints in the development of a clinical research base.

The deans of new schools are dedicated to the task of building institutions with the usual mission of teaching, research, and service. In response to the public's demand for more physicians, the top priority in these new schools has been enrollment of as many students as possible, and expansion of our student bodies has usually exceeded considerably the original plans.

The arguments advanced by Dr. Krevans for biomedical research in medical schools applies to the new schools as well as to the established ones. In replying to a brief questionnaire circulated to the 24 accredited new schools, all 16 deans who responded indicated that investigator-generated basic research is important in their institutional development, although four (4) stated that budgetarily such funding to date has been insignificant.

The General Research Support program was generally considered very important by the deans. In my own case it has been vital. We have largely focused its use on setting up newly recruited faculty in research, and I do not know how we would have gotten along without it.

Exceeding the \$100,000 base of Federally-supported research required for eligibility for GRS, however, has posed some difficulties, with the range of time needed varying from 1 to 8 years, with a clustering at 3 years.

Newness is a self-limited condition, and some of us in the group considered "new" for this presentation find ourselves increasingly identified with the established institutions. A new school, however, does have an opportunity, which occurs only once in its history, to set its own course, and establish goals and objectives uniquely appropriate for time and place, and it behooves society, once it has authorized a new medical school, to do what it can to assist it to achieve its goals and objectives. Some of these new schools have made a point of taking a whole new look at the educational process with major new curriculum developments; others have made a special effort to generate a plan for the community to which it relates, to improve the health in the ghetto area or in a regional area, such plans are being definitely a part of what the institution was mandated to do in the state in which it was established.

There is, however, a level of research essential for excellence in a medical school environment; although this is not to say that all schools should aspire to the same level of research intensity. I feel confident, however, that without such a research base, no school can long be successful.

Building this base has for many been difficult. The problem of the \$100,000 threshold for GRS, and difficulties of attracting investigators and grants to some new institutions has been mentioned. New schools would seem rarely to be in positions to respond to research contract RFPs, and are probably inappropriate sites for most clinical research centers or other types of center programs.

The proposed Biomedical Research Support Grant Program, and its accompanying Biomedical Research Development Grant are steps in the right direction. The Biomedical Research Development Program, funded from 10% of the funds available to the Biomedical Research Support Program would make research funds available on a competitive basis to "institutions that currently have limited involvement in biomedical and behavioral research". This would include some new medical schools.

We feel, however, that a Federal program to support research in new medical schools, should be a program in its own right, rather than one which borrows from, and hence weakens, another.

Thus, we feel that there should be a start-up program in biomedical research for new medical schools available from NIH and NIMH linked to the early stages of the accreditation process. Such a grant to a new school would be limited in time, and would yield at its expiration to the generally available successor to General Research Support. Thus, support for a level of research activity would assist states, which have the basic responsibility for establishing new medical schools, make optimal use of the funds they have available within the state for planning and implementing the new schools.

Some of what I have said could apply to established schools whose research base is deemed inadequate. The Minority Biomedical Support Program within the Division of Research Resources, NIH, has supported the development of the health sciences at predominantly Black colleges.

The NIGMS is funding a Minority Access to Research Careers Program designed to provide opportunities for pre-doctoral (Ph.D) support, pre-graduate support, faculty fellowships, and Visiting Scientist Awards in predominantly minority colleges.

I cite these actual and proposed programs as examples of specific NIH efforts to solve the problem of certain inequities in the area of research and research training.

The precedent, it seems to us, could be built upon to help specifically with the funding of research in needy new and under-developed established medical schools

OVERHEAD: RATIONALE AND REALITY IN 1975

Richard L. Seggel
Senior Professional Associate
Institute of Medicine

An important issue

Institutional overhead costs (i.e., indirect cost rates) charged to research funded by the Federal government are becoming a matter of increasing concern to both policymakers and research investigators alike. These costs have risen much faster than the direct costs of research over the past seven years. In Fiscal Year 1967, they averaged 18.6 percent of direct costs. In Fiscal Year 1974, they had risen to an average of 31.7 percent of such costs, and, for reasons I shall mention in a minute, this ratio will probably continue to go up. The impact on the Federal budget of this trend is all too obvious.

It has led to the revival of discussion of a Federal ceiling on indirect cost rates, not unlike that which existed 10 years ago when the Wooldridge Committee conducted its study of NIH and which was subsequently abolished at the Committee's recommendation. In particular, suggestions of such a ceiling have been raised by the staff of the Office of Management and Budget and the investigations staff of the House Committee on Appropriations. The question of indirect cost rates on Federally-sponsored research has clearly become an important issue which needs to be thoroughly discussed and understood.

Protection of the Federal interest

The college and university business officers recently stated that:

"It is difficult to find a subject which causes as much misunderstanding as indirect costs. On the other hand, there is no other area examined so thoroughly, nor has any other been so carefully defined and redefined through various revisions of the cost principles established by the Federal government."1/

In this connection, I would like to emphasize a few key facts. First, the cost principles and procedures applicable to both direct and indirect costs were developed by Federal representatives who are not known to be free with a Federal buck--namely, representatives of the Office of Management and Budget and the pertinent government

1/ NACUBO Studies in Management, Indirect Costs: A Problem In Communication, Vol. 4, No. 1, July 1974.

audit agencies, as well as the comptrollers and financial officers of the major Federal agencies sponsoring research. These Federally-established principles and procedures govern the identification of the pertinent indirect costs of the institution and the allocation of the appropriate share of these costs to Federally-sponsored research. For example, plant operation and maintenance costs are allocated to research, instruction and other institutional functions, on the basis of the space utilized for each of these functions, general administrative costs on the basis of salary costs of each of these functions, library services on the basis of usage, etc. This process excludes from Federal research support all indirect costs properly allocable to instruction and other non-research institutional programs.

Secondly, the rates are based on actual institutional costs, which are subject to Federal audit. The institutions may project justifiable increases over past-year costs for inflation as a basis for arriving at a current indirect cost rate but these are subject to subsequent adjustment based on the actual cost experience.

Thirdly, the rates are set by negotiation between the Federal government and the institution. In the case of biomedical research, the rates are negotiated by HEW staff operating out of its regional offices. HEW has approximately 35 to 40 trained professionals assigned to this function. In addition, it has hundreds of auditors who check the books of the institution to assure the accuracy of their cost figures.

I believe that, although there are undoubtedly some imperfections in the system and operational slip-ups here and there, it is fair to assume that the Federal investment in biomedical research is reasonably well protected in this process. On the other hand, we should most certainly recognize that the Federal policy of paying full indirect costs does not translate to full recovery by the institutions of all the costs they incur in the conduct of Federally-sponsored research.

Shortfalls to the institutions

For one thing, when the Federal government went on full indirect-cost basis nearly 10 years ago, the Appropriations Committees stipulated that there should be significant cost-sharing by the institutions. This has been interpreted to mean at least 5 percent of the total, with anything below that subject to question.

In the second place, the original flat limitation of 8 percent for indirect costs on research training grants has never been lifted. This is principally due to the fact that, over the years, these

grants have been under attack by the Office of Management and Budget and the estimates for their support seriously constrained in the President's budget.

And thirdly, the institutions have trouble keeping their indirect-cost proposals up with the pace of inflation. Some institutions still base their proposals on past-year costs; they are apparently reluctant to project higher indirect cost rates because of their concerns about the possibility of pricing themselves out of the competitive market for certain types of grants and contracts and stimulating adverse reaction on the part of their faculties. I need hardly add that "overhead" charges are no more popular with faculty than they are with Federal budgeteers, particularly at the rate they have been advancing in recent years. It is, therefore, especially important to come to grips with the reasons why indirect costs have gone up faster than direct costs.

Reasons for rapid indirect cost increases

There is no doubt that some of the recent increases in indirect cost rates is attributable to the developing sophistication of institutional management personnel concerning the indirect cost items which constitute properly allowable charges to research under the Federal guidelines. But, beyond this, there are many other important factors.

First, inflation has had a heavier impact on some indirect cost items than on the direct cost items which typically make up the research project budget. I need only mention such items as fuel and utility costs. In some cases, fuel costs have gone up 300 percent in the last year or so. I am advised, for example, that a rise in fuel costs alone required an increase in the case of one institution of 9 percentage points in the indirect cost rate and in the case of another of 4 percentage points.

Secondly, these rates are increasing as a consequence of a number of new requirements placed on the institutions. These include the Federal requirement for committees to review the protection of human subjects in clinical research, the more stringent standards established by the Federal government for the care of research animals, the additional standards imposed by the Occupational Safety and Health Act of 1970, the provisions stipulated by executive orders for affirmative action programs, added administrative requirements imposed by Federal regulations and Federal auditors, and others.

This list could be greatly extended, but I think I have said enough to illustrate the problem.

What do we do about it?

Placing an arbitrary limit on indirect cost rates is certainly not the answer. It would be the wrong way for the very reasons which moved the Wooldridge Committee 10 years ago to recommend the Federal allowance of full indirect costs on research. It would be highly inequitable, since indirect costs necessarily must vary widely from institution to institution, depending on the nature of their organization, the extent of their plant and grounds, etc. Above all, it would only compound the financial difficulties already confronting the institutions. Providing them with less than full indirect cost allowances for research projects, along with the current cost-sharing requirement and the low ceiling on rates for research training grants (not to mention other financial problems of the institutions) would undoubtedly be a further serious blow to the financial integrity of many of the institutions on which we depend for biomedical research in this country.

I see no recourse but to face this problem squarely in the budget process--i.e., to identify and explain the estimates for funds to cover full indirect costs and vigorously defend them as add-ons to the estimates for whatever program levels for research grants are decided upon. Every effort should be made to overcome the usual tendency of budgeteers to slough off the impact of increased costs of such an item by requiring their absorption within a fixed budget total and thus obscuring significant reductions in the level of the program actually being approved. In my judgment, the only way in which this tendency can be overcome is to insist in every way possible that the indirect cost-requirements be met head-on as a separately identifiable item in the budget.

THE IMPACT OF CENTERS AND TARGETED RESEARCH
ON ACADEMIC MEDICAL CENTERS:
INSTITUTIONAL ISSUES

Stuart Bondurant, M.D.
Executive Vice President and Dean
Albany Medical College

Most would agree that the nation's programs of biomedical research need both aggregated and targeted research effort as well as dispersed and investigator initiated research. In many instances the large aggregations of targeted effort may provide the best setting for extending research programs into developmental efforts and to selective and effective articulation with the health care delivery system. These large and often complex programs have specific management, personnel and resource requirements which are not entirely congruent with those of the educational institutions in which many of them are best housed. The complexity of these programs is nicely illustrated by Figure I, a diagram of the functional relationships of a center which was kindly provided to me by Dr. Richard Ross, Dean-elect of Johns Hopkins Medical School and was designed by the Ad Hoc Working Group on Centers of that institution. However, the institutional issues which arise from the articulation of centers and targeted research are not new and they have been addressed with varying degrees of success by many groups which have worked out arrangements which are viable if not entirely satisfactory. While there are substantial issues, I believe them to be resolvable, provided there is recognition and initiative on the part of the institutions and full use of the administrative flexibility which exists under Federal laws and regulations and provided that Federal policies of research support include appropriate provision for sustaining the identity and integrity of the host institutions as well as the entire national biomedical education enterprise.

There are, of course, several kinds of centers and programs of targeted research. I will not discuss the characteristics of specific kinds of centers or the real questions of the appropriate size and number of centers or mechanisms of support for targeted research. Rather I will try to order the institutional issues which are associated with centers and targeted research and to offer comments as a guide to resolution of the issues.

To guide our discussion I will list six categories of institutional issues which are associated with centers and targeted research.

The first set of issues are those which concern or influence the systems of governance and management of the institutions. These issues derive from the need to integrate the dissimilar functions of governance and management associated with education

and with large or targeted Federal programs. They result in part from operational restrictions upon the centers and targeted research which are necessary to assure the accomplishment of goals. Targeted or center programs are associated with Federal management procedures which may in effect cause Federal participation in decisions which concern institutional governance and management. The systems of institutional governance need to provide a basis for reconciling the needs of both the centers and the institutions. A specific example of this problem is the occurrence of branched or ambiguous lines of institutional authority. Centers and targeted research create new intra- and extra-institutional constituencies to which institutional systems of governance must adapt. This is often reflected in the creation of a center advisory board which does not fit into the institutional decision-making procedure.

A second set of issues are those which concern criteria for allocation of institutional resources. From the time that an institution considers applying for a center or targeted research program through the operational phase of the programs there exists pressure to redirect institutional resources of space and personnel. Many institutions have committed multiple man-years of effort to writing applications with little chance of funding. Both the research programs and the institutions will benefit from considered criteria for allocation of institutional resources to centers and targeted research.

An extensive set of issues concern personnel policies and practices of the institutions. For example, there is the issue of the need for long-term commitments to personnel who are supported by programs without long-term guarantee. There is the institutional issue of the distortion of the mix of personnel types in the institution through the acquisition of staff with disciplines and goals which are not entirely congruent with the primary institutional purpose. There is the problem of the divided loyalty of staff supported by centers and targeted research projects. The existence of divergent goals may lead to dual personnel practices such as criteria for promotion or other advancement.

Fourth, there is a set of issues which concern the educational programs of the institutions. There may be a real or apparent relative de-emphasis of the primary educational purpose of the institution. There may be reordering of educational priorities according to available resources rather than according to institutional purpose. There may be an inappropriate mixture of programs or contamination of effort of both the research and institutional programs. The institution may come to provide an inappropriate mix of role models for its students.

A fifth set of issues are those which concern the relationship of academic institutions to the variety of other institutions which may be involved in the centers and targeted research programs. In some situations the academic institutions need to work with industry, community-based organizations, State and local governments or elements of the health care delivery system. Some center proposals have involved as many as 10 institutions, thereby posing issues of shared responsibility and authority as well as resource allocation.

The sixth group of issues are those which result from changes in the system of Federal research support. For example, the creation of large centers and programs of targeted research acts to aggregate the biomedical research and education enterprise of the nation with resulting substantial disadvantage to those institutions which do not house centers and targeted research. The disadvantages concern both resource development and staffing and ultimately influence the effectiveness of both the institutions and the Biomedical Research Programs. If the nonparticipating institutions are to sustain a base of research or scholarly endeavor, they must find other means of support. The creation of centers and targeted research programs can have profound effects upon the health care delivery system in that it can cause dislocations in the flow or distribution of patients among institutions. These changes can have broad social and economic impact on states and regions. Through participation in centers and targeted research, institutions can acquire a public responsibility which transcends their original societal role and may in fact exceed the limits of corporate charters. Finally, the increasing emphasis on centers and targeted research creates a capability and need for flexibility in the Federal programs which conflicts with institutional needs for stability.

I would offer the following suggestions:

1. That some, but not necessarily all, centers and programs of targeted research be housed in educational institutions.
2. That Federal planning for centers and programs of targeted research include explicit definition of the goals and requirements of the programs and a systematic analysis of the projected impact of each proposed program upon the educational institutions and enterprise.
3. That Federal policies of support and funding for centers and targeted research include explicit provision for sustaining the integrity of both host and non-host educational institutions (perhaps like some of those of NASA in the days of Webb and Dryden).

4. That the disadvantages of over-aggregation of the research effort be recognized through purposeful encouragement of dispersion of effort where appropriate.
5. That there be developed a set of guidelines--not regulations--for the design of systems of governance and management which must meet the needs of both the institution and the program.
6. That a description of institutional impact including that on the systems of governance and personnel policies be a consideration in the award of centers and large programs of targeted research.
7. That the goals and requirements of centers and targeted research programs be stated in advance in sufficient detail to reduce the number of non-responsive applications or proposals.
8. That the establishment of institutes or other dedicated corporate entities be explored to host certain specific types of targeted research. (Models being Lincoln Labs, Jet Propulsion Lab)

THE FUNCTIONS OF A COMPREHENSIVE CENTER

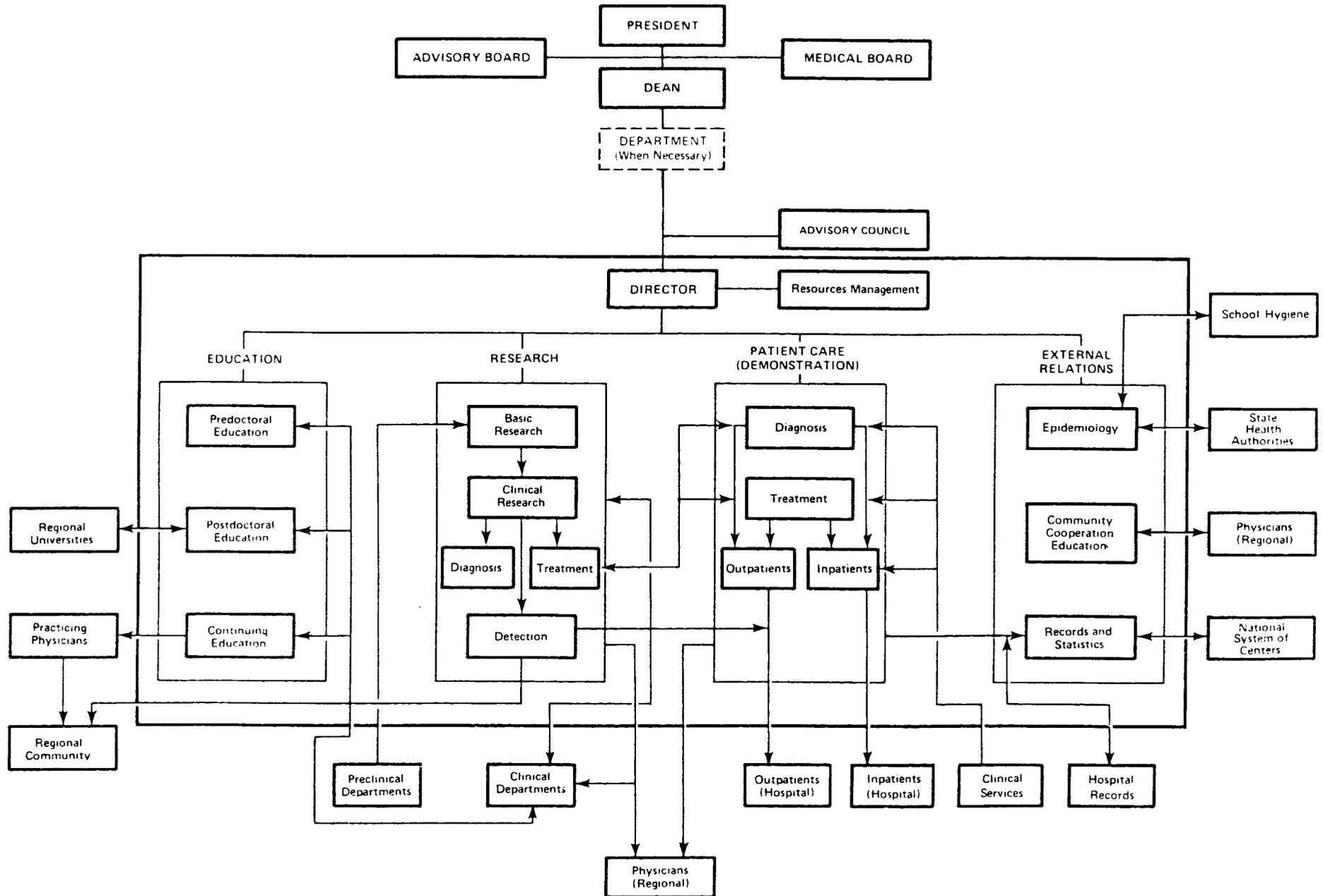


Figure 1

Centers and Targeted Research - Institutional Issues

Group I

Issues which concern systems of governance and management of institutions

1. Reconciliation of governance of institution and center.
2. Resolution of branched lines of authority.
3. Accommodation of new intra- and extra-institutional constituencies.
4. Resolution of differing management requirements.

Group II

Issues which concern criteria for allocation of institutional resources

1. Priority of institutional versus center development.
 - a. Physical plant.
 - b. Personnel - numbers and types.

Group III

Issues which concern personnel

1. Provision for long-term support.
2. Mix of types of personnel in relation to institutional purpose.
3. Divided loyalty and motivation.
4. Institutional criteria for advancement.
5. Reconciliation of disparate wage and salary scales.

Group IV

Issues which concern educational programs

1. Pressures to reorder educational priorities to align with resources.
2. De-emphasis of education.
3. Contamination of some educational programs with inappropriate content.
4. Imbalance in role models available for students.

Group V

Issues which concern relations to other groups and institutions

1. Development of effective working relationships with:
 - a. Industry.
 - b. Community-based agencies.
 - c. State and local government.
 - d. Health care providers.

Group VI

Issues which result from changes in the pattern of Federal research support

1. Aggregation versus dispersion of research and academic enterprise.
2. Influence on health care delivery system - patient distribution.
3. Broadened institutional corporate responsibility.
4. Institutional need for stability.

BIOMEDICAL RESEARCH RESOURCES: GENERAL SUPPORT,
CLINICAL RESEARCH CENTERS, RESEARCH
MANPOWER TRAINING

Chandler A. Stetson, M.D.
Dean, University of Florida
College of Medicine

The topic assigned to me clearly represents a problem area, or it would not be on today's agenda. The nature of the problem, however, is perhaps not quite so clear. Is there something intrinsically unsound about these programs, or something wrong with them, or is something else the matter? Speaking from my own experience, I can say that general research support grants have occasionally been misunderstood if not misused by one dean or another; a few general clinical research centers have sometimes seemed to have more "clinical" and "research" value and less "general" than the NIH staff would have wished; and there are some training grants with records of having produced no single truly productive investigator. But that is not the problem. Characteristically, these programs are sound and productive. Ten years ago, as panel members working for the Wooldridge Committee, many of us had an opportunity to view closely and evaluate a sizeable and statistically valid sample of these programs, and it was perfectly clear to all of us and to the Committee that, in sum, these programs of institutional support were well conceived, well administered and contributed in major ways to the stability and strength of the institutions and to the production and productivity of biomedical research manpower. Later, when I had the privilege of serving as chairman of the National Institutes of Health General Clinical Research Center Committee and still later of the Medical Sciences Training Program Committee, I had an opportunity to get a thorough overview as well as a detailed close look at these programs and I became even more convinced of their generally high quality and of their tremendous value to the institutions.

So what is the problem? Clearly what Dr. Bennett had in mind in placing this item on the agenda was in fact that these institutional support programs have been the object of particular attack by the OMB, and I think that the basis for this, the real problem posed by these programs, is that they are parts of positive feedback loops. That is, they are components of a system in which management intervention can have results (either desirable or undesirable) which are amplified by positive feedback in the system. For example, a management decision to increase funding of training of research manpower will result in a larger population

of scientists which will in turn lead to an increased demand for research project funding which in turn leads to an enlargement of the overall research enterprise which in turn tends to create an increased demand for training of more research manpower, and so it goes. Now viewed from the perspective of the biomedical scientist, management's decision to increase funding of training grants would be considered entirely proper and would be followed by entirely desirable positive feedback; but viewed from the perspective of the administration, quite a different management decision has appeared to be in order. Training grants must look pretty bad--even dangerous--to a manager who doesn't really know much about research, who does not have any particular confidence in the scientific method, or any particular appreciation of the long-term investment in biomedical research that will be needed to finally rid man of the burden of disease. Such a manager would be quick to see that a decrease in funding of training grants will lead in turn to a decrease in the number of scientists, a decrease in the size of the research establishment, etc., etc.

Having begun with training grants, let me say that I think that it has been a mistake in discussions of training grants to lump predoctoral graduate education and post-M.D. research training. The predoctoral trainee, aimed at a career in basic research, has a breadth and depth of scientific training that gives him a versatility that can scarcely be acquired in any other way. The post-M.D. trainee, on the other hand, usually has a different objective--primarily clinical research, where his knowledge of human pathobiology gives him a distinct advantage in a unique place. While we have had a reprieve with respect to starting new predoctoral training grants for fiscal 75, it now appears definite that no new predoctoral training grants will be started next year. This clear discriminatory decision, against predoctoral training grants, flies in the face of advice from many informed members of the biomedical scientific community, who are evidently assumed to be suffering from what I think Charles Hitch calls the "nobility of purpose self-delusion." But 6000 principal investigators of NIH research grants today are holders of the Ph.D. degree, many of them among the nation's most distinguished and productive scientists. It seems clear to many of us that the national interest requires that there be a mechanism for replacing them with equally well-trained young men and women who will be their counterparts in the future. Rather than eliminating predoctoral training grants, I would join those who argue for their continuation with the emphasis on quality rather than on numbers. Highly competitive programs of high quality would not need to be evaluated from year to year and would generally have a half-life long enough to be a positive stabilizing factor in the institutions. Holding these programs to a low number of high quality students would seem to pose no threat or risk of saturating the field.

Just in passing, I am a bit surprised that the administration has not been more impressed by the simple economics of predoctoral versus postdoctoral research training. Training of predoctoral students probably costs roughly half as much as that for post-doctoral students. At least, the costs to NIH today are of that order of difference, according to the following rough calculation. If one takes the total dollars requested in the Kennedy-Rogers bill and divides by the total number of trainees authorized, one comes out with a figure \$8,100. One arrives at the same figure, too, by working backwards: typically a predoctoral student receives a stipend for around \$3100, tuition averages \$2800, a travel allowance of around \$200, and an institutional allowance of \$2000, all adds up to the same figure.

Contrasted with that \$8,100, the best available figure for the cost of training a post-M.D. research fellow seems to be around \$17,500. Most stipends for such fellows are in the \$10,000-\$14,000 a year range, and \$12,500 is now the average stipend in those programs administered by the National Institute for General Medical Sciences. Around \$4,400 goes to the institution, another \$200 for travel and a small average allowance for tuition and we are up to \$17,500.

Lest I be misunderstood, I am not trying to make a case against postdoctoral research training grants; on the contrary, I do not see how we could mount or continue an effective clinical investigation program without them. I am saying, as many have said more eloquently before, that predoctoral fellows typically have different career goals, typically receive a qualitatively and quantitatively different educational experience, typically are better equipped to undertake basic as opposed to clinical research, and go on to form a substantial and distinguished segment of the biomedical research community, and costs much less to educate! I believe that this is a most serious policy issue, and to any members of the Panel who may be of like mind, I apologize for whipping what you may feel is a dead horse.

With respect to alternative methods of supporting research training, these are so cogently summarized and evaluated in Dr. Eugene Braunwald's research paper in the New England Journal of Medicine of February 6, 1975 that I can do no better than refer you to it, and perhaps to quote the following paragraph in which he makes reference to the direct individual fellowship mechanism:

"Fellowships that are paid directly to the recipient have many of the same advantages as do training grants, but they do not provide a means for developing and maintaining an effective training environment. Essentially no funds are provided for the

specific purpose of strengthening the responsible department or of lending stability to the program. Further, direct fellowships do not bolster the department or division through the process of continual national review. Also inherent in the direct fellowship mechanism of support are wide year-to-year swings in the number of trainees in any department or division. Fellowship programs primarily judge trainees, with less emphasis on the training environment. They do not establish the national standards for excellence in training that are a characteristic of the training-grant mechanism and therefore can provide little assurance of the quality of the training environment." Braunwald also addresses the mechanism of training support through research grants and contracts, and concludes (for many compelling reasons) that "research grants and contracts are poor substitutes for stipend support through training grants."

I would next like to address the General and Biomedical Support Grants. Here, too, I believe that these institutional support grants are all too clearly tied into a variety of positive feedback loops that are in fact desirable from our point of view and enable us to further our institutional objectives, but make OMB personnel very nervous. It would be presumptuous of me to review for this group the origin and uses of this program. It is unnecessary for me to emphasize the continuing need for flexible funds and slack resources, to permit us to intelligently and effectively manage our institutional development. I would like to emphasize one aspect of the problem that has perhaps not received enough attention, and that has to do with institutional and faculty morale. Somehow it is immensely reassuring to a young investigator to be able to turn to his institution for seed money research support and get it--from the institution! If this mechanism is not continued, I am very much afraid that the perceptions of young faculty will be that their only avenue of funding for their research is to be found in the oppressively competitive national research grant marketplace. But with institutional funds, there can be a very meaningful direct exhibition of the institution's commitment to research and scholarship, and I submit that this is a not insignificant aspect of this problem. Balancing programs, extending research support into areas not even close to the big centers, providing some insitutional cohesion, these are all functions which the GRS grant mechanism has served and which may well in the future become even more important than they are today, if the tendency to concentration of research funds in large centers continues. Current administration plans are to discontinue this program--I think that the vast majority of my colleagues feel that this would have a profoundly demoralizing and destructive effect on our institutions and would rather quickly result in an unfavorable change in the quality and quantity of young professionals opting for research careers.

Finally, a few words about the General Clinical Research Centers. If we did not have them, we would certainly have to invent them now. Current costs of hospitalization are so high and utilization review so stiff and agency constraints so tight that it would simply be unthinkable to attempt to do most kinds of clinical investigation today without these discrete, specialized units. They are to the clinical investigator what the laboratory is to the basic scientist. They have been chronically in trouble from the administration, largely because of something less than maximum use of beds in at least some centers. I believe that the management of this program has been excellent--pruning of less productive centers has enabled survival of the majority in the face of level funding for several years with no provision for inflation. Quality control has been careful and well-documented. Productivity of the centers has been gratifying, and one can only hope that their future stability can be assured.

In conclusion, let me say that I am afraid much of the criticism levelled against these institutional support programs has come not because of any intrinsic defects in them but as a result of their visibility as control points that could be used by management to "turn up" or "turn down" the level of research activity of the nation. Such "on-again off-again" management has clearly not been good either for the biomedical research enterprise of the nation or for the institutions that train the necessary manpower and do most of the research. Indiscriminate turning off of all predoctoral training grants simply cannot be supported as a defensible rational management decision, either as a short-range or long-range strategy. It is precisely in the control of these areas that we need stability, wisdom, informed leadership and clear national policy.

ORGANIZATION OF FEDERAL RESEARCH:
POLICIES AND LEADERSHIP

Ivan L. Bennett, Jr., M.D.
Dean, New York University
School of Medicine

Mr. Chairman and members of the Panel:

Thank you for this opportunity to appear before you.

I can begin by stating one of the rare, indeed, perhaps unique, unanimous views of the Council of Deans: the biomedical research enterprise of the National Institutes of Health (NIH) represents one of the proudest accomplishments of the Federal establishment. The record is clear and, despite the fact that the last few years have seen erosions of various etiologies of both the extramural and intramural efforts, we believe that the situation can be controlled and ameliorated by readjustment in policies, clarification of priorities and goals, and sharpening and updating of implementation mechanisms. Remember, then, that we are dealing with Federal institutions whose programs have placed the United States in the forefront of biomedical science, institutions that are the envy of and emulative model for other nations of the world. The continuing excellence of these institutions, even during the constraints and crippling of recent years, indicates clearly that they merit no huge overhaul, but merely repair, support and sustenance for rehabilitation rather than replacement.

In this terminal, brief presentation, I will attempt to identify certain shortcomings in the present Federal system of support for biomedical research that we hope that the Panel will study and examine closely. After a background statement, I will suggest certain recommended areas for investigation and consideration by the Panel.

The tenor of the times has brought into question or, at least, placed under searching reassessment, the emphasis formerly accorded the biomedical research enterprise within the Federal budget. Until now, the scientific aspects of this reassessment have been severely compromised by the economic and political realities of apparent competition for limited funds between research and medical care. This competition, which is as artificial as it is injurious, must be reduced and eliminated, since medical care and biomedical research are inseparable features of a single national objective - better health for all.

The problems of health care delivery in this country are urgent and complex and require their own solutions but a cutback in research and research training is a mortgaged solution with unacceptable consequences; namely, reductions in the rate of medical progress and the quality of medical education and, ultimately and inevitably, an unfavorable impact on the health of the American people. It is primarily research that creates such medical care as there is to be delivered and, hence, shapes the delivery system.

Research, unfortunately, has become a scapegoat or perhaps a better analogy is a penance-machine for many of the nation's social faults.

The support of biomedical research should be consistent. Erratic changes in the level of funding of the scientific enterprise destroy morale, vitiate planning, and waste human and capital resources.

In recent years, great uneasiness has developed among the various elements and levels within the government and at the vital interfact of government and the private sector. This has given rise to an almost incalculable loss of mutual trust and confidence, not merely from constraints and fluctuations in program funding but, to an even greater extent, from the way decisions are made and programs are managed.

There is, for example, increasing and counterproductive cynicism both within and outside of government over:

- Annual and semiannual vacillations in policy,
- Frequent and ineffectual reorganizations of health-related agencies,
- Lack of clarity about priorities at every level,
- Instability of scientific leadership,
- Lack of adequate means for science advice to the President and to the Congress.

Optimal results from biomedical (and other) scientific efforts can be achieved only through calculated programs of long-term support with sufficient flexibility to encourage and to exploit promising new ventures and creative ideas. This, in turn, requires effective communication between scientists and federal program managers, between program managers and higher decision-makers in the Executive branch, and between the latter and the Congress. It is the universal view of the biomedical scientific community, intramural and extramural, that these lines of

communication, once so clear and effective, are now blurred, confused, or extinct. The picture is one of changing signals, uncertainty, instability, and frustration. There have been erratic shifts in program emphasis with political bets placed prematurely on pursuit of quick solutions to problems which, at present are so little understood that no coherent scientific approach to them is yet possible. There has been continued fragmentation of categorical programs (in the form of new institutes, etc.) with overhead and administrative costs eating into program dollars. Program imbalance has become a way of life rather than a transient perturbation. There have been repeated attacks on the peer review system, some well-intentioned, many not, and all of them uninformed, divisive, and disruptive. There have been precipitous and distinctive changes in leadership and all without any appearance of urgent concern for the maintenance of continuity of the research effort.

The vast majority of this wallowing and drift results from political and economic considerations to the virtual exclusion of scientific considerations. The annual budget for NIH programs must traverse a tortuous route through the bureaucracy as it wends its way to Congress. Confronted by a welter of differing views and levels of managerial and scientific expertise, it is viewed in the context of all of the health missions, ranging from health care delivery to the regulation of drugs. Next, research must compete economically and politically within HEW with the requirements of programs of education and of welfare. In this latter process, research must survive competition with the major uncontrollables in the Federal budget, including health care reimbursement. As it has developed, science program judgements receive increasingly meager attention in this cacophony and, more and more, top priorities are assigned to the political grabbers.

Innovation in health care for the next generation depends upon today's basic research. Such research takes place chiefly in NIH itself and in the universities, colleges, and their related medical centers. These centers are under pressure to expand their responsibilities by increasing class size, training more paramedical personnel, and by assuming a larger and more central role in the health care enterprise itself. These forces, largely unaccompanied by corresponding funds for implementation, coupled with the vagaries of research funding, place severe financial strain on these essential institutions. I will only mention to underline it the additional impact that inflation has had on the centers.

If the science and technology of medicine were static and unchanging and if we were to expect no new advances in the understanding and therapy of human disease, new physicians could be trained by book and by rote, as technicians, to learn today's methods for management of disease as doctrine and dogma. But if scientific understanding of disease is to be continually enhanced and improved, and new forms of prevention and treatment are to be based upon this new information, there can be no question about the "relevance" of biomedical research to the education of physicians.

The Council of Deans suggests to the Panel some areas for attention, study, and changes:

Within the Federal establishment:

1. We suggest a thorough examination of the missions of NIH and NIMH as they relate to the total Federal health enterprise in order to determine the best ways for these agencies to interface with programs of health care delivery and regulation without compromising their biomedical research responsibilities.
2. We suggest appointment of the NIH director for a specified term, subject to renewal, similar to the appointment of the Director of the National Science Foundation (NSF).
3. We urge that salaries of key scientific leaders be made competitive with the private sector. This has been accomplished for the new Uniformed Services Medical School, may soon be implemented for the Veterans Administration Hospital system, and should certainly be possible for NIH.
4. We suggest that a top-level scientific policy group should be reestablished for science and technology, including biomedical research, at the level of the Office of Management and Budget (OMB) and the Domestic Council to advise the President, OMB, and the Congress. Such an arrangement would be more objective than program managers, could facilitate cooperation among various government agencies supporting biomedical research, and would promote clarity around scientific issues and priorities at various levels in government. We doubt that the present arrangements involving the Director of NSF and the new Congressional Office of Technology Assessment (OTA) will suffice to carry out these functions, especially in the biomedical research field.

5. We urge that the Panel become aware of increasing encroachment by legislation upon biomedical and other research and seek protective measures. For example, the so-called Baumann amendment, passed by the House, is the severest challenge in history to peer-review and to scientific quality. In no way would it increase quality or relevance of research, would it speed innovation, or would it even increase public accountability. A second example, from many which could be cited, should suffice to emphasize this concern. Traditionally, the portion of research grant applications submitted to NIH/NIMH containing the investigator's research protocols, hypotheses, and designs has been treated as confidential by the funding agency. This practice was challenged by the Washington Research Project-Children's Defense Fund in a suit against DHEW under the Freedom of Information Act. The Association of American Medical Colleges (of which this Council of Deans is an integral component) submitted a brief Amicus Curiae to the U.S. Court of Appeals for the District of Columbia Circuit in support of the government's contention that these documents were exempt from disclosure under the act on the grounds that they contained material of a proprietary and confidential nature procured by the government under an assurance that they would be treated as confidential. The government lost the case on this issue and was ordered to disclose the documents.

Subsequently, the AAMC has proposed that the Public Health Service Act be amended to preserve the confidentiality of such information for a period of at least 12 months after funding of the application, except in the case of human experimentation in which case the 12 month period could be waived by the Secretary of HEW.

We would hope that the Panel would look at the area represented by these examples and devise insulation against such encroachments for the future.

More generally we suggest for the Panel's consideration:

1. The objective of highest priority for Federal health programs should be the elimination or mitigation of major human diseases for which effective technologies for prevention, cure, or control do not now exist.

2. Biomedical research should be recognized as a central element in Federal health programs and stable levels of fiscal support should be provided for it. The annual amount of support to be proposed by the Executive Branch should be extrapolated from existing levels and projected for three years. In the absence of significant new initiatives, the rate of change, up or down, in this indicative budget should be gradual and should be considered in relation to its impact on universities, colleges and their related medical centers.
3. When an area of undifferentiated research has progressed to the point that it appears to be possible to focus the work, support for the "targeted research" phase of the program by expansion of grants and contracts should be derived from appropriation of new funds rather than from a trade-off against ongoing programs of basic project research, research training, or intramural programs.
4. The General Research Support Program should be resuscitated and continued as the mortar between the bricks of grant and contract projects.
5. We urge the Panel to examine the accretion of measures and regulations originating in both Legislative and Executive Branches that have steadily eroded the flexibility and the effectiveness of the working arrangements between NIH program managers and principal investigators, the essential interface that finally gets the work done. The increasing tendency of Congress to legislate not only what should be done, but to legislate also in infinite, uninformed detail exactly how it is to be done, imposing uniform requirements and stereotyped programs upon a system whose unmatched success has been possible because of its pluralism, diversity, and flexibility should be brought forcibly to the forefront as a counterproductive trend that in no way increases accountability and assures eventual, uniform mediocrity.
6. We urge the Panel to clarify mechanisms for setting priorities. We recognize the political imperative of setting priorities in consideration of public wants and opinions. We also recognize the economic imperative imposed by limited or finite resources. Our plea is for some mechanism that will reintroduce scientific facts and judgements into the process of decision-making at an appropriate and effective level. The present

confusion of political realities, economic realities, and scientific realities at all levels in government seriously disrupts the continuum of productive effort at the laboratory bench and in the clinic, whether intramural or extramural.

Mr. Chairman, this concludes my statement and the presentation by the Council of Deans. We would be pleased to respond to questions or to discuss other topics of concern to the Panel. I would hope that the Panel would not limit its questions to those of us who have made these formal statements but would take advantage of the expertise provided by the presence of so many of our other colleagues from the Council of Deans.

Thank you.

ACADEMIC DECISION-MAKING:

ISSUES AND EVIDENCE

Proceedings of the Association of American Medical Colleges
Council of Deans Spring Meeting, April 26-30, 1975.
Key Biscayne, Florida.



ASSOCIATION OF AMERICAN MEDICAL COLLEGES
SUITE 200, ONE DUPONT CIRCLE, N.W., WASHINGTON, D.C. 20036

RE: ACADEMIC DECISION-MAKING: ISSUES AND EVIDENCE

ERRATA

This compilation of the papers presented at the 1975 Spring Meeting of the Council of Deans contains material which we believe is of interest to many people. We regret that in printing this document many typographical errors were not corrected. Since a change at this time would delay the publication and increase the expense, we are distributing the document in its present form. We believe that these errors have not altered the meaning and hope that they will not detract from the significance of the papers.

Joseph A. Keyes
Director
Division of Institutional Studies

ACADEMIC DECISION-MAKING:

ISSUES AND EVIDENCE

Proceedings of the Association of American Medical Colleges
Council of Deans Spring Meeting, April 26-30, 1975,
Key Biscayne, Florida.

January, 1976

Association of American Medical Colleges
One Dupont Circle, N. W., Suite 200
Washington, D. C. 20036

CONTENTS

EDUCATING PHYSICIANS AND SCIENTISTS 1

W. D. McElroy, Ph.D.

The need for a solid scientific education for physicians is now beyond dispute. The chancellor of the San Diego campus of the University of California says, however, that the amount of science in the medical student's early years probably is being overdone, and he offers a proposal to alleviate the situation.

EVALUATION FOR DECISION-MAKING: A CONCEPTUAL FRAMEWORK 7

Christine McGuire

A 1974 survey conducted by the AAMC Council of Deans forecast that academic medical centers will have new responsibilities for both graduate and continuing medical education. The author cautions that an expansion can be undertaken without a dilution of quality only if rational choices are made, and to that end she outlines the indispensable elements of the decision-making process and the support system required by educational administration.

INFORMATION SYSTEMS FOR MONITORING STUDENT PERFORMANCE 19

James V. Griesen, Ph.D.

Too often, large amounts of money and faculty time are devoted to the implementation of methods for evaluating student performance which are incomplete and inefficient. The author stresses the inherent value of developing *systems* for monitoring student performance, and he illustrates his presentation with examples from the Ohio State Independent Study Program.

EVALUATION OF PROBLEM SOLVING 29

Lee S. Shulman, Ph.D.

The author reviews a number of approaches to the evaluation of clinical problem solving in medical students. He concludes that no single method possesses a full range of virtues without attendant liabilities and supports a *program* of clinical evaluation combining many methods.

AN ANALYSIS OF STUDENT PERFORMANCE AS REFLECTED IN THE NON-THREATENING
AUDIT OF THE STRUCTURED MEDICAL RECORD

37

Anthony E. Voytovitch, M.D.

The author describes a non-threatening audit of the structured medical record which grew out of his experiences with third-year students. Success with the techniques suggests a new direction in the analysis of clinical performance: peer- and self-evaluation in a program of continuing education.

THE EVALUATION OF FACULTY

43

Hilliard Jason, M.D., Ed.D.

Faculty evaluation is made difficult by the relative absence of a tradition for gathering systematic data and by the substantial absence of mechanisms for analyzing or interpreting data. The author says that while the elimination of uncertainty is not a reachable objective, a search should be undertaken for a method that can facilitate judgment-making and optimize the utilization of those data that are available.

SOME THOUGHTS ON MEDICAL EDUCATION

67

Steven Muller, Ph.D.

Although the medical schools turn out some superb physicians, problems do exist in premedical and undergraduate education. The president of The Johns Hopkins University calls for a new look at the admissions process and for the integration of all basic sciences within the university.

THE PRODUCT OF OUR MEDICAL SCHOOLS IN PERSPECTIVE

73

John W. Williamson, M.D.

Important sources of data and information are becoming increasingly available which can contribute to decision making in the field of health services. Building on the most solid facts should be of great help in formulating policies that inevitably will have increasing impact on efforts to meet the health needs of society.

EDUCATING PHYSICIANS AND SCIENTISTS

W. D. McElroy*

I appreciate this opportunity to talk with you. With your assistance, I hope to raise questions and stimulate discussion about the education of new physicians.

All of us here are aware of the major issues facing medical schools and their associated health care programs. Certainly there has been no lack of public discussion on these matters, and health issues have risen high on the national agenda. I can add very little new to this subject, and indeed, I suspect that these critical aspects of policy and financing may now be largely out of the hands of the medical community and will be primarily decided by people and institutions outside that environment.

CURRICULUM DESIGN

In the past, curriculum design in undergraduate medical education has been almost exclusively controlled by the medical faculty, subject only to board accreditation guidelines. But even here there are disturbing signs of external influence. While the initial federal legislation for construction grants specifically forbade governmental agencies administering the program to interfere with the school's curriculum, several states have legislated departments, and at least one pending bill in the Congress calls for specific curricular changes. If the schools wish to stop this insidious encroachment and return to the sensible tradition of faculty control of the medical curriculum, the schools will have to demonstrate and prove their sensitivity to some of the deep national currents concerning the education of physicians.

That education is my topic here. I am the first to acknowledge the pitfalls and difficulties encountered by a nonphysician on this subject and ask for your indulgence if I obviously overlook a critical point.

My tentative thesis is simple to state: science content in the education of physicians requires a critical reexamination to determine if it matches the anticipated realities of the future. Put in a more pointed form, certain to stir up the traditionalists: are we teaching the right science at the right depth at the right time? Is the common core of science -- that which every medical student is expected to know -- too extensive for some students, not enough for others? Is it programmed at the most appropriate time, when motivation to learn is the highest?

Before I develop my argument, let me reassure you that I am not challenging the need for a solid scientific education for physicians. Flexner's tenets, in this regard, are as valid today as they were 65 years ago. But, unlike Flexner, we now have positive proof of the effectiveness of science when applied to certain problems of disease.

Having bowed sincerely as well as traditionally to the *Flexner Report*, I advance no startling new idea when I say that from one point of view, the history of American medical education in the past 20 or 30 years has been one attempt after the other to break from certain curriculum strictures of that famous report. For example, in

*Chancellor, University of California, San Diego.

Medical Education Reconsidered, the result of a 1965 gathering of distinguished medical educators, this point is faced squarely: "The medical schools face a new orientation as radically different as that of 1910. The palliative measures of the last 20 years have accomplished little. Fundamental changes are once more required."

MEDICAL CURRICULUM

Of course, we have seen modifications in the medical curriculum in the past decade, and most medical educators I talk to consider them favorable and a step forward. And yet, when measured against the accelerating pace of change we see in the total American health system, our curriculum response appears to be relatively modest.

Let me review briefly several of the most pertinent changes on the medical scene since 1910. These are all too familiar to most medical deans, but seemingly their totality -- and implications for the medical curriculum -- has not clearly registered on all of our *faculty* colleagues. The most striking developments include:

1. *The increased scope and complexity of science.* This point seems too obvious to belabor. If you have any doubts about specifics, just consider what was known about the basic biology of the cell when you were in medical school versus what is known now. In so doing, I think you will also appreciate the complexity of modern biology.

2. *The expansion of research within the medical school.* Since World War II, all of our medical schools have increased enormously their research efforts. In some medical schools, from the standpoint of time alone, research is probably the principal activity of the majority of the professionals in the institution.

3. *The improved science education background of medical students.* Since the launching of Sputnik I, there has been a marked improvement in the teaching and sophistication of science instruction. We all know of the advanced science courses commonly taught in the secondary schools and in collegiate programs. One dean has told me that his entering students know more science than he did when he was graduated from medical school 20 years ago. Furthermore, even though a small percentage of first-year students have weak backgrounds in science, many medical schools have developed remedial programs to correct these deficiencies in a relatively short time.

4. *Specialization now a fact of life for all advanced medical students.* Every M.D. graduate now spends at least three to five years in clinical specialty training. This is a perfectly natural development and simply reflects the fact that no individual can expect to have more than a passing knowledge of the entire range of medicine.

5. *The changing patterns of medical practice.* The solo practitioner, unconnected with a group or clinic or hospital, is fading like an old soldier, even in rural areas. University medical centers are increasingly involved in the practice of medicine.

6. *The increased public policy and financial involvement with fundamental aspects of medical care.* Both Senator Kennedy and Dr. Edwards have spoken bluntly on this recently. Only an ostrich -- or a completely out-of-touch faculty member -- could ignore the Congressional and public mood for changes in our health system.

Taken together, these perfectly obvious developments in recent years have resulted in a distinct liberalization of the medical curriculum. We now understand phrases like "core and elective," "earlier clinical experience," "the biology of disease," "independent research," and "the continuum." Virtually every medical school, I understand, reflects some of these changes and a handful of schools have made very considerable modifications in the style and substance of undergraduate medical education.

Worthy as this direction is, we have not yet come to grips with what I believe to be the heart of the matter

-- the content and timing of the medical student's science education. Let me approach this point through a consideration and comparison of the education of physicians and scientists.

FLEXNERIAN MODEL

Not a long time ago, in what some might term the apogee of the Flexnerian model, there were more than a few scholars who believed that a medical student's preclinical education should be similar, if not identical, to that of a biology graduate student. I'm reliably informed that at least several schools attempted to teach the same curriculum with the same faculty to both first- and second-year medical students and graduate students. That this experiment had a certain logic is acknowledged, but the fact is, it proved unworkable, and to the best of my knowledge this concept has been quietly dropped.

We can learn from this experiment, for it does seem to illustrate that the education of physicians should be neither identical nor totally dissimilar from the education of graduate students in science. A scientist requires an education of very great depth in a special field and must demonstrate the capacity for independent judgment. (When I say special field, I really mean a sub-field of a larger discipline. For example, a biochemist today can no longer be an expert across the board in biochemistry.) A physician, it seems to me, needs a broader science education encompassing a number of fields and sufficient to achieve a level of scientific literacy adequate to keep up with advances in those fields. A physician who wishes to concentrate on advancing the science of medicine will need both types of training; I will have more to say on that presently.

By drawing a distinction between the science education of physicians and graduate scientists, I make no revolutionary statement -- this is certainly the conventional wisdom as clearly evidenced by a comparison of the respective curriculums. Here it should be noted that no one that I know is comfortable in stating exactly what science is needed by the practicing physician. Despite the increased options afforded medical students, the general medical curriculum seems to have expanded more, in Julius Richmond's word, by accretion than by careful analysis of the student's need. Perhaps such an analysis is what is required; it might be an excellent foundation-sponsored project. If done properly, I believe it could significantly influence faculty and departmental attitudes.

Short of such a study, let me explore in more detail the question of the science education of physicians from the perspective of a former scientist. As Hippocrates once said, "Extreme remedies are very appropriate for extreme diseases." Safely wrapped in these words, I advance a straw man for your consideration. In the interests of fair play, I shall mention only the positive aspects of my proposal -- the negatives will come soon enough.

BASIC SCIENCE

My first point is that alluded to earlier -- that we are probably overdoing the amount of basic science in the medical student's early years. Even with the liberalization of the curriculum, it still seems for true than false that the average student faces the unbearable task of mastering an enormous range of knowledge that he may never use in his practice or research career. When I look at the biochemistry textbooks used in medical schools, I have to assume that the faculty are training practicing biochemists. The same point can be made, I understand, after examining pharmacology and microbiology texts. And, no matter how strong the medical school research environment, only about 10 percent of a school's graduates choose an academic research career.

It would seem that there are at least two ways of alleviating this situation, both of which could possibly

lead to an improved educational experience. One way -- largely a continuation of a current trend -- would be to increase the number of courses and electives, thereby constructing, in effect, a series of "majors" fitted closely to an individual's interests and talents. In 1967 Paul Sanazaro described this approach: "In the place of the fixed medical curriculum, we see the emergence of multiple majors within medicine. Each department within the college becomes, in essence, a school offering courses at the undergraduate and graduate levels. Some of these would be required of all candidates for the M.D. degree. Others would be offered for special students or for selected majors." Assuming a student has a reasonable idea of what he wants to do, this concept would appear to have some attraction, although the faculty teaching time would be increased considerably.

ANOTHER APPROACH

Another approach -- my straw man -- would be quite different, especially in the first year. I propose that **all** medical students take several newly designed "science foundation" courses which would review and relate the current science underpinnings of the practicing fields of medicine. In many ways, these courses might be best taught by clinicians, preceding from the known problem to the known and unknown aspects of the solution, supplemented by appropriate lectures by the preclinical faculty. I emphasize that these courses should not be what we disparage as "appreciation" or "survey" courses; the fundamental science should be taught, but emphasizing the broader principles and not the labyrinth of detail. Perhaps three courses would be of this nature, one each in the biological, physical, and social sciences, leaving adequate time for other materials as well as an introduction to the clinical aspects of medicine. These courses might also reduce some of the problems associated with under-prepared students.

In the second year of this curriculum, the student could begin to point toward his broad area of specialization through a series of courses specifically designed to differentiate between those choosing primary care and those choosing any other specialty. For a year or more, the student would have instruction within one of these two broad clusters of courses, eventually, of course, the student would settle into a specific specialty track.

Once in that final track, the student would, in addition to the regular clinical work, begin intensive class and lab work in the basic science specifically related to his anticipated specialty. It is at this point that the student would receive intensive, indepth scientific training, plus independent research, if appropriate.

TRUE CONTINUUM

In my straw man I assume a true continuum between undergraduate and postgraduate medical education. This implies that some of the advanced science course work might well take place during the formal residency period. Here, I might add, the medical student would be in a graduate level course, sitting side by side with Ph.D. candidates. For those students choosing research as a primary medical career focus, the 10 percent, my model would encourage them to obtain a joint M.D.-Ph.D. For those wishing to teach on a medical faculty, the joint degree would probably be absolutely essential.

This is obviously a very sketchy exposition of my thesis. I do not believe it is appropriate to go into any more detail, for if the idea has any merit there are far more qualified people than I to do just that. All I wish to do here today is to encourage you to take a fresh look at what science is taught when to medical students.

It would be inappropriate to consider some of the ramifications of this different approach to medical education. In the first place, I would hope that it could free some of you from the tyranny of the examinations of the

National Board of Medical Examiners, at least Part 1. As you know, there are discussions underway now to test knowledge more closely at the beginning of post-M.D. training. To me, the basic science material is more logically covered there than during the second year of the formal curriculum.

But the "National Boards" are a relatively minor ramification to the logical progression of my straw man. Under this curriculum, the traditional basic science departments would become less critical in the teaching aspect of a medical school. I have already suggested that perhaps the first-year "foundations" courses might be taught by clinicians because they should be able to relate the material better to actual health problems. Conversely, I believe the advanced science courses, what might be called graduate level, should be handled by the basic scientists. In this topsy-turvy situation, however, there seems less justification for medical basic science departments. In my opinion, the individuals in those traditional departments might find themselves both more comfortable and more productive if they were integrated into arts and science departments or, depending on individual choice and interests, into the clinical departments. The University of California, San Diego, School of Medicine has no basic science departments but very strong basic science within the clinical departments. Using participating faculty members in the arts and sciences and the basic science members of the clinical departments works very well for us. Of course, even one not on a medical faculty can appreciate the difficulty of implementing this suggestion; certainly, there would have to be an evolutionary period, keyed to the proposed change in undergraduate curriculum.

Please note that I am not advocating an abbreviated medical curriculum. I do believe, however, that more time could be available for serious study of the behavioral sciences, the social sciences, and the humanities, a point the *Journal of Medical Education* reprints a majority of medical deans strongly favor. I would like to see the curriculum include courses on, for example, the social and legal aspects of practice, nutrition, and sex and population dynamics. Furthermore, if one of your goals is to break the lock-step environment and turn to the less structured graduate atmosphere, would there not be merit in spreading these courses throughout the total curriculum rather than concentrating them entirely in the four-year medical school period?

Assuming my basic proposal has interesting possibilities, what might be done to see that it is brought to the serious attention of medical school faculties? As an understatement, I might say that recent history suggests that these institutions, never known for rapid change, are somewhat on the conservative side, particularly when issues about curriculum are discussed. Despite this, as long as we cling to the belief that rationality sometimes prevails, a careful study of what science and what level of science are required for the superior performance of each specialty certainly would be in order. If we begin with the assumption that we really do not know these answers, if we reject the notion that every physician should know the latest developments *in detail* in every field of science, I believe we could obtain a great deal of useful information. Such a study, by the way, may demolish my straw man rather than support it.

Although this may not be the most popular suggestion to a group of deans, it may be well for presidents and chancellors to get involved with issues of the type I have discussed. There has been a tendency of late for medical schools to have considerably more administrative latitude than other schools of the university. While this may be desirable from some viewpoints, it does tend to remove the university's chief executive officer from frequent contact with the medical school. I am not advocating any official change in these relationships, but merely stating that in the long run, interest and involvement from the president or chancellor on this issue may accelerate

a closer consideration of the medical curriculum.

I have tried to be provocative. I hope I have challenged you enough to consider some different approaches to science content within the total spectrum of medical education.

EVALUATION FOR DECISION-MAKING: A CONCEPTUAL FRAMEWORK

Christine McGuire*

The 1974 Delphi Survey (2) conducted by the AAMC Council of Deans, limns a composite picture of the universe of medical education as envisioned by that group 10 years hence. The forecast emanating from that study depicts a world in which academic medical centers will be involved in developing medical education as a continuum and will have new responsibilities for both graduate medical education and continuing medical education that rival in importance those which they now assume for undergraduate medical education. The study further suggests that these new functions can be performed without decline in the quality of programs, despite substantial changes in financing and less vigorous growth in funding that virtually everyone anticipates.

In the face of the predicted zero growth in resources, a significant expansion in responsibilities can be undertaken without deleterious effects on quality only if priorities are set and the resulting choices are enforced. Making such choices rationally, and explaining and defending them under difficult circumstances, leads to increasing dependence on information systems which can be relied on to yield comprehensive, fresh, accurate, and meaningful data. It is the purpose of this paper to outline the indispensable elements of the educational decision-making process and of the information and evaluation support system required to inform that process and to maximize its benefits.

AN EVALUATION MODEL -- OVERVIEW

The decision-making process and the support system requisite to educational administration are completely analogous to those required in any other managerial situation -- from general institutional management to patient management; only the *objects* of the decision-making differ. In the case of educational decision-making, it is the student, the faculty and the educational program which require evaluation. With respect to each of these three objects, evaluation is always undertaken for a purpose: either that of making inescapable "go/no go" decisions and/or that of monitoring progress in order to counsel improvement.

Student Evaluation. With respect to students, the "go/no go" decisions are concerned with admission and selection, promotion and graduation, licensure and certification. Increasingly, for both legal and ethical reasons it is necessary to be able to document such decisions with detailed, reliable, valid and objective data that accurately describe each student's characteristics and competencies. Student evaluation for purposes of counseling requires similarly precise measurement of each student's level of functioning with respect to numerous requisite cognitive and non-cognitive skills, in order to arrive at an accurate educational diagnosis and to develop a specific, individual educational prescription that will assist each student in improving his or her professional health.

Faculty Evaluation. With respect to faculty, the "go/no go" decisions for which sound data are required are those having to do with hiring, firing, promotion and assignment. To avoid capricious application of idiosyncratic

*Professor of Medical Education, associate director and chief for research and evaluation, Center for Educational Development, University of Illinois College of Medicine.

standards in this arena, it is essential that each institution develop clear, explicit definitions of what an effective scholar is, what the intervals of evaluation will be, and how performance will be measured. Only then will it be possible to meet the demands of both faculty and public for administrative accountability -- demands that have been fully legitimated by repeated court decision. Similarly, when it comes to monitoring faculty performance for purposes of counseling, an information system that helps individuals to identify their own strengths and weaknesses, that aids them in satisfying personal desires for self-improvement, and that provides reliable data for planning and implementing appropriate faculty development programs is indispensable.

Program Evaluation. Finally, with respect to the evaluation of educational programs, the "go/no go" decisions are obvious: at the national level they are most often concerned with issues of accreditation -- institutional approval and disapproval. At the institutional level such decisions include not only those relating to the traditional programs of undergraduate medical education but, as corporate responsibility for graduate programs is implemented, it will be necessary for each institution to make a "go/no go" decision about each of its graduate programs which, though approved by the accrediting agencies, may or may not be compatible with institutional goals and standards.

Somewhat disguised, but even more common in program evaluation are those "go/no go" decisions relating to allocation of funds, space and personnel among approved programs that compete for these scarce resources. Rational decision about these matters requires a system of monitoring the relative efficiency and effectiveness of the several approved programs in order to determine cost-effectiveness for purposes of better resource allocation. Such monitoring also serves the counseling purpose by providing data that assist in identifying strengths and weaknesses of the respective programs and that suggest modifications in educational settings and strategies that are demonstrated to be less than optimal.

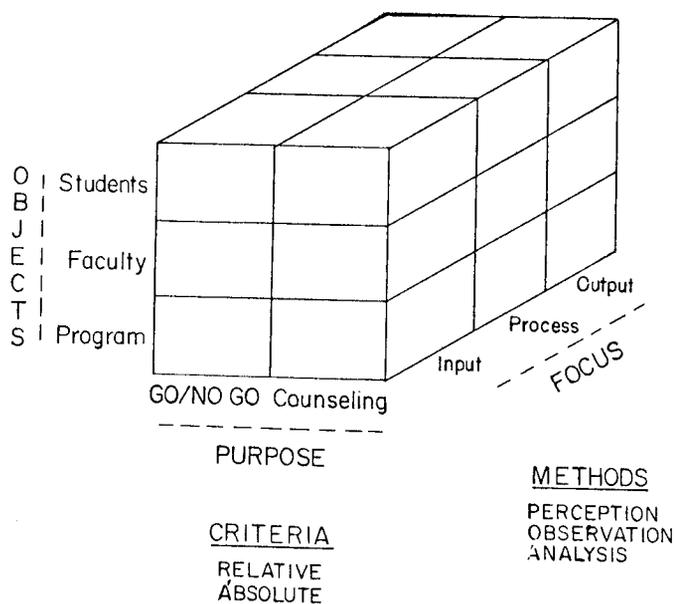
FOCUS OF EVALUATION

These then are the objects and purposes of educational evaluation. To complete this part of the model, it is necessary to consider one additional dimension (Figure 1): namely, the *focus* of the data collection.

Input Variables. In some instances that focus may be largely or exclusively on the organizational structure and resources of the object -- i.e., the *input* variables. For example, in their preoccupation with data about the number of certified faculty, the faculty/student ratio, the laboratory and clinical facilities, the total instructional hours available to each discipline and the like, traditional procedures for accrediting programs are illustrative of such a focus. Conventional student admissions procedures have similarly stressed the structure and resources of the applicant: his scholastic aptitude, his personality structure, his values and attitudes.

Process Variables. In contrast with the prevailing approaches to accreditation of graduate and undergraduate programs and to student admission, evaluations of programs of continuing education and of faculty contributions to institutional programs at all levels have all too often been limited to collection of data about process variables -- that is, the systematic documentation of what actually occurred in a particular program or in a series of encounters between teacher and learners. Where this approach has been properly employed, the evaluator has attempted to gather comprehensive data about the extent to which the actual conditions for learning, and the interaction between the learner and the environment are congruent with those deemed optimal for achieving a given set of objectives. This

Figure 1: An Evaluation Model for Educational Decision-Making



approach is now increasingly employed in higher education in the evaluation of faculty, where numerous instruments are being developed for use in gathering data about the nature of the learning environment created by a particular teacher, and the quality of the communication -- i.e., the interaction between students and teacher -- which takes place in that environment.

Fortunately, both students and patients are a hardy lot. Were it not so, it is difficult to imagine how either could have survived the noxious interventions to which generations of both have been subjected. While more critical attention to the process variables might have spared society some of the deleterious effects of misguided "therapies" -- both medical and educational -- it is doubtful whether process assessment has yet developed to the point where it can be used for purposes of "go/no go" decisions. Until such time as the basic connections between process and product are unequivocally established the primary utility of process assessment is in support of counseling decisions.

Output Variables. Clearly, the ultimate test of any system is its output -- that is, the product which it produces. In the case of an educational system that output is the performance of its graduates. Typically, however, such performance data have been used chiefly in the evaluation of students, primarily for decisions associated with promotion, graduation and certification. But if it is true that all teaching and all educational programs are undertaken to facilitate student learning it follows that the effectiveness of both teachers and programs can ultimately be assayed only in terms of the extent to which this intended outcome is achieved -- i.e., in terms of changes in student performance. The techniques now available for assessing performance (4) range from conventional tests of information, through sophisticated simulations of clinical problems (3, 6, 10, 11), to longterm, systematic observations of the learner in varied professional settings.

However, since the evidence is mounting that, given adequate time, personal instruction and self-instructional resources, most anyone of normal intelligence and emotional stability can master any element among the objectives

of medical education, no educational evaluation is complete in the absence of cost data. In the final analysis, therefore, educational evaluation for decision-making necessarily entails economic considerations in determining whom a country can afford to educate and at what level of mastery that is, at what cost in resources. Unfortunately, at present the only usable cost data tend to be those concerned with the direct money costs attributable to the construction and maintenance of specialized facilities (e.g., lecture halls, student laboratories) and materials (educational films, slides, self-instructional programs). What is urgently needed in order to evaluate faculty and programs in cost-benefit terms is a detailed recording of the amounts of all resources that go into any instructional effort, including, for example, the increased laboratory and hospital costs attributable to the use of patients as "teaching materials." And, if the notorious unreliability of retrospective data is to be avoided, this implies the necessity of undertaking diary studies of representative samples of students and faculty to determine how much of their time is actually being devoted to some kind of instructional activity, the exact nature of that activity and the way in which the quantity and distribution of such activities change with different instructional methodologies.

In short, it is proposed that both program and teacher effectiveness can be judged only after the costs are documented and charged against the desired changes in student knowledge, judgment, skills, habits and attitudes which are realized. Two types of data are therefore indispensable in focusing on output variables: that derived from comprehensive measures of relevant aspects of student performance, and that obtained from careful cost accounting of resources consumed and of undesired side effects induced by the educational interventions.

SPECIAL CONSIDERATIONS IN APPLYING THE MODEL

Employment of the model outlined above requires brief attention to two additional sets of issues: first, issues relating to methods of collecting data on input, process and output; and second, issues concerned with the nature of the criteria to be used in decisions relating to the counseling and disposition of students, faculty and programs.

METHODOLOGICAL CONSIDERATIONS

Sources of Data. With respect to the first, provision needs to be made for utilizing three types of data sources: one, *perceptions* of participants in the educational program; two, *systematic observations* (using the term in its technical sense to include all types of objective measurements) of input, process and product; and three, various kinds of *analyses* based on appropriate manipulation of the measurements.

The Use of Consultants. It is in relation to these methodological issues that the role of the professional educator is best defined in terms of the five-fold contribution he should be prepared to make to the decision-making process: first, as technical advisor in helping to identify the payoff variables whether the focus be on input, process and/or output; second, as a collaborator in the development of the most reliable instrumentation for systematic accumulation of data about those variables; third, as technical expert in designing the techniques for collecting and analyzing the data; fourth, as a primary technical resource in the development of a total information system; and, finally, as technical consultant in making the findings of current research in medical education regularly available to inform the decision-making process.

Related Educational and Psychological Research. In this connection five general areas of research in medical education should be noted as especially relevant to educational decision-making: research on the goals of medical

education; research on the characteristics of medical students; research on the setting for learning, curricula and instructional methodologies; research on teaching and research on methods of evaluating professional competence (9).

Research on the *goals of medical* education has led to the development of a number of procedures for collecting scientific evidence about what competent physicians need to know and need to be able to do in order to deliver responsible health care. Three of these methods -- the critical incident technique (5), the method of task analysis and the analysis of epidemiological data (16) -- are of special interest in furnishing an empirical basis for a behavioral description of the essential components of professional competence to guide a faculty in setting institutional goals and designing institutional programs.

In contrast with the relatively limited and only recent attention that has been given to research on the goals of medical education, *student characteristics* have long been subject to intensive research and study (7), particularly as these characteristics relate to the problems of selection and attrition. However, recent research on student selection has focused less on the intellectual and academic attributes of the applicant population and very much more on the personality, values and interests of that population. This shift in emphasis has been accelerated by the conviction that, in order to have any significant impact on the geographic and specialty distribution of health manpower, it will be necessary not only to redefine the goals of medical education but also to reassess criteria for student selection and to look more carefully at the attitudinal and other noncognitive variables which influence career choices.

Related to the new direction which research on student characteristics has recently taken is a concurrent upsurge of interest in looking at the implications of research on the relation between environmental press (12) and personality needs for decisions about *curricular organization* (e.g., integrated versus disciplinary) and *institutional strategies* (e.g., lecture versus small group discussion versus programmed instruction, etc.). Earlier control studies of the relative effectiveness of alternate methodologies have produced conflicting evidence and inconclusive results (1, 13). What has emerged from more recent research is clear evidence first, that students learn in many different ways and at very different rates; second, that the appropriate instructional strategy depends on the nature of the educational objectives sought; third, that whatever the methodology utilized, the greatest learning takes place when students are actively involved in the learning process and when the material to be learned has the greatest apparent relevance to the students' own objectives; and fourth, that the most effective program is one which provides genuine opportunity for these individual differences in learning style and learning speed, and in which the particular materials and instructional techniques have been chosen with a view to specific, explicit objectives that are shared by faculty and students (1, 13).

Like research on curriculum and instructional methodology, *research on teaching behavior* per se has so far produced only the negative finding that the presence or absence of any particular teacher behavior cannot be demonstrated to be generally associated with augmentation of student learning. Rather, research in this field has yielded a variety of useful approaches and instruments for documenting the perceptions of, and transactions between, teachers and learners. This documentation has proved increasingly effective in counseling faculty who desire to improve their professional performance (1).

And now to complete the circle: this brief digression into the application of research findings to

educational decision-making was introduced with a consideration of research on the goals of medical education; it will conclude with a note about research on the outcomes of medical education. Research on *techniques of evaluating professional competence* has taken two forms: first, careful, systematic analysis of existing instruments available for use in assessing student achievement and educational outcomes and second, the development of new and more effective methods of evaluating those outcomes of special importance. The emerging methods encompass new types of paper and pencil tests (4, 17), unconventional oral and practical examinations (4, 6), reliable record audits (15), objective observations of performance in actual hospital and clinical settings and, of special significance, a fascinating variety of simulation techniques (3, 6, 11). Together, these techniques furnish a broad array of new instruments useful not only for purposes of evaluating most important aspects of student and physician competence, but also for monitoring the effectiveness of educational programs and identifying promptly any areas of deficiency.

PHILOSOPHIC ISSUES

Discussion of the methodological issues noted above still leaves unresolved certain philosophic issues in applying this or any other model for decision-making. Of particular significance is a consideration of the nature of the criteria to be employed. Both for purposes of making "go/no go" decisions and for purposes of counseling toward improvement, questions are sometimes posed in the form "Is X better or worse than Y?" Other times they take the form: "Is X good enough?" In short, for certain decisions it will be necessary to use *relative* criteria and the data on which those decisions are based must be adequate to distinguish better from worse, and must be organized in a manner that facilitates comparison on all significant parameters. For other decisions it will be necessary to apply *absolute* criteria in determining whether or not X meets pre-determined standards, irrespective of its position relative to A, B and C. Under these circumstances the data must be sufficiently relevant, reliable and comprehensive to justify that hardnosed judgment without reference to actually existing alternatives.

CASE ILLUSTRATIONS

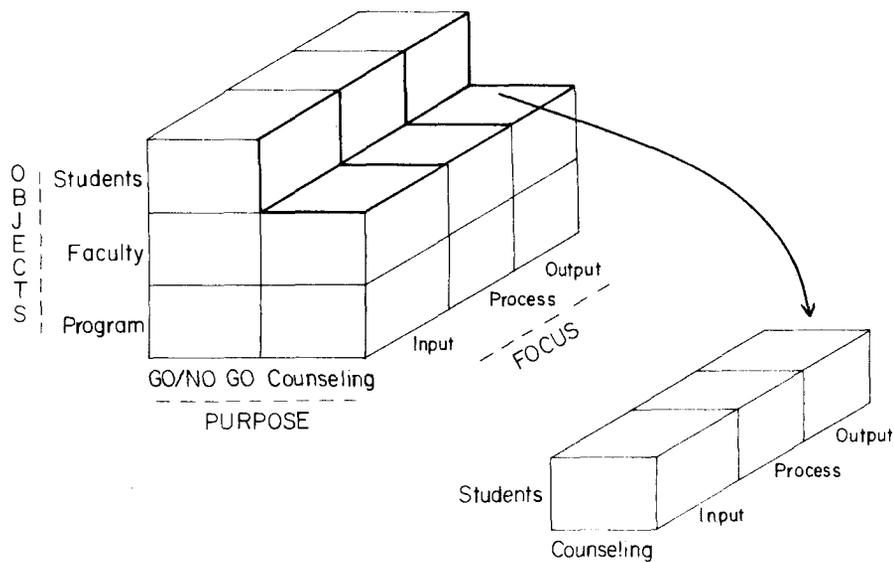
Rather than elaborating each of the elements of the model, the balance of this monograph is devoted to a series of brief case presentations illustrative of information and evaluation systems to support educational decision-making with respect to each of the objects: students, faculty, and programs.

STUDENT EVALUATION

To illustrate evaluation of students three cases are presented. The first (Figure 2) describes a total information system for monitoring student input, process, and output for purposes of counseling students and guiding their learning. It is a computer managed system designed to support the experimental program of the Pilot Medical School at Ohio State University. Though fully computerized by its developers the system is amenable to other management modalities.

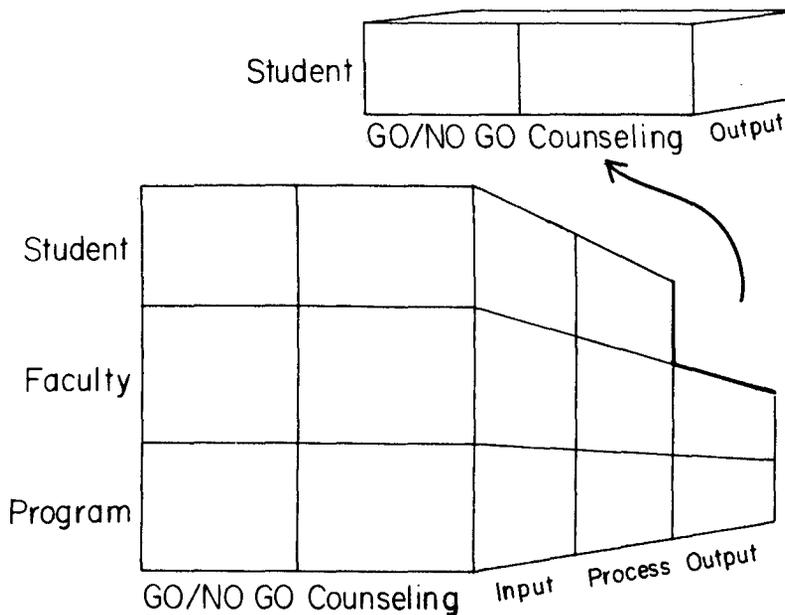
However, the utility of even the most elegant information system obviously depends on the quality of the data fed into it, and this, in turn, is a function of the reliability and validity of the data collection instruments employed. As noted earlier it is in the field of instrumentation that the professional educator can be especially helpful and it is in this area that research has yielded important new tools which need only be applied. The other two cases concerned with student evaluation report on some of these innovative techniques for gathering data on

Figure 2: Student Evaluation for Counseling Purposes



output variables (Figure 3). In the first the relative merits of different types of tests for the measurement of diagnostic and problem-solving skills are compared. But, no matter how closely a test situation simulates a life situation, there is always a nagging concern about the extent to which performance in the former predicts the latter. The last case presentation on student evaluation therefore describes experience with record audits in the instruction and evaluation of medical students, for both counseling and certifying purposes.

Figure 3: Output Measures of Student Performance



FACULTY EVALUATION

Many different cases might also have been selected to illustrate that part of the decision model concerned with faculty evaluation. The most appropriate would depend not only on the range of relative values assigned to research, patient care, service and teaching in a medical school, but also on the underlying concept of the responsibilities of the teacher in a medical faculty since, clearly, the proper method for evaluating teachers is dictated by the accepted concept of his/her role.

The teacher's role in the medical setting has been variously described as analogous to that of a physician, an expert consultant, a master craftsman, a counselor, a guide and a shepherd. In a delightful essay on the subject, Dr. Richard E. Snow (14), presents the metaphor of "The Teacher as Bayesian Sheepdog." In elaborating that metaphor he notes the role of the teacher in "herding" students along a path from their existing location to some presumed desired location, rounding up strays and bringing the whole "flock" together first on one side of the optimal route and then on the other. As he describes it, "the key features of this metaphore are: a two dimensional space in

which a direction vector represents the teacher's chosen path and a flock of points represents pupils (Figure 4); a measure of the average flock direction which the teacher estimates periodically; the difference between average flock direction and the teacher's direction vector, which the teacher tries to minimize; and a measure of flock density which the teacher tries to maximize. Using the pedagogical analogs of circling movement, feints, and charges, along with iterative sampling and estimation of location and direction, the teacher moves his or her class through a field, physical or psychological." (14, page 90) If one accepts the analogy between "teaching" and "herding" then the evaluation of the teacher, either for the purpose of counseling or for the purpose of promotion, becomes a conceptually simple problem of obtaining two measurements: one of flock density and the other of average flock deviation from the optimal path.

Rather, however, than selecting a specific example of teacher evaluation which would, of necessity, be based on an arbitrary and possibly irrelevant concept of the role of the teacher, the case presentation on faculty evaluation has been selected to focus attention on a particularly troublesome area of decision-making -- namely, evaluation of faculty for purposes of promotion (Figure 5). Many questions have been raised about the legitimacy of traditional procedures for personnel

Figure 4

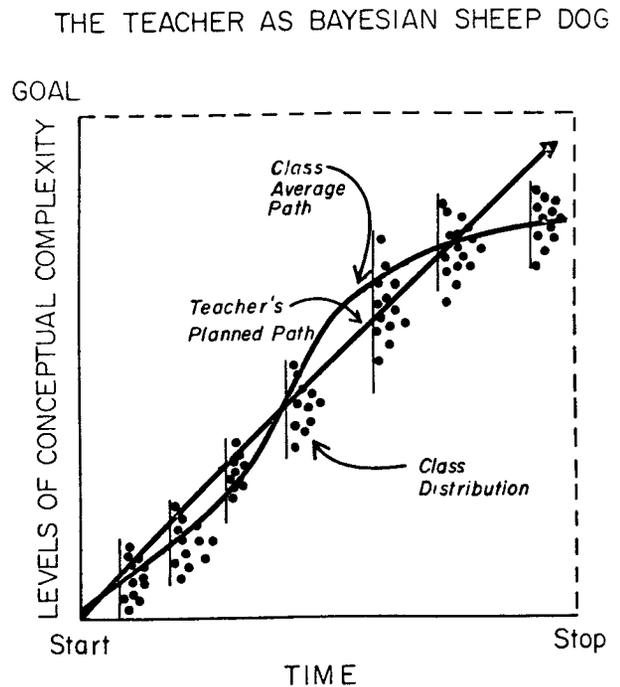
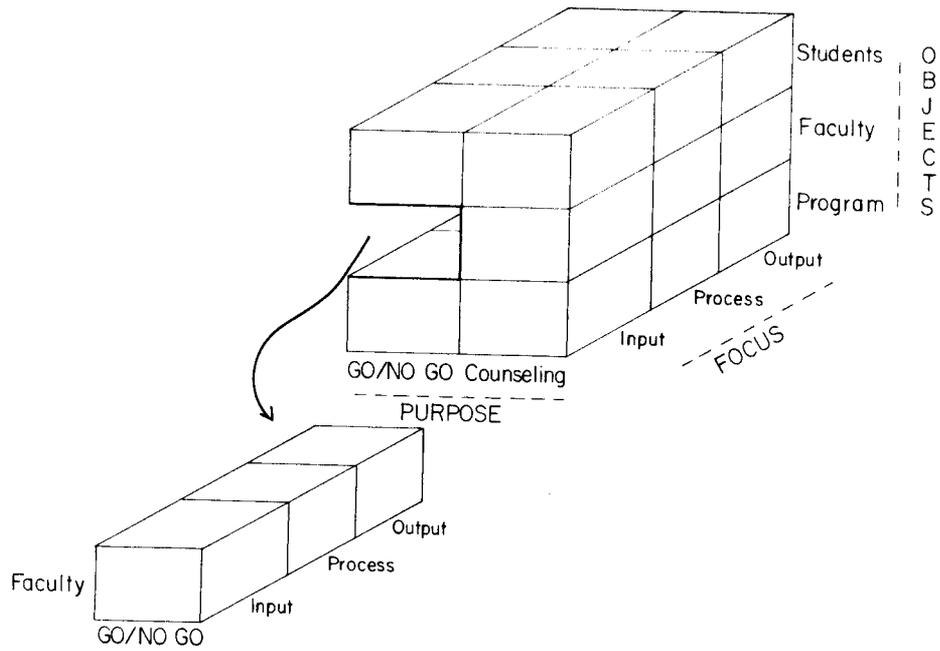
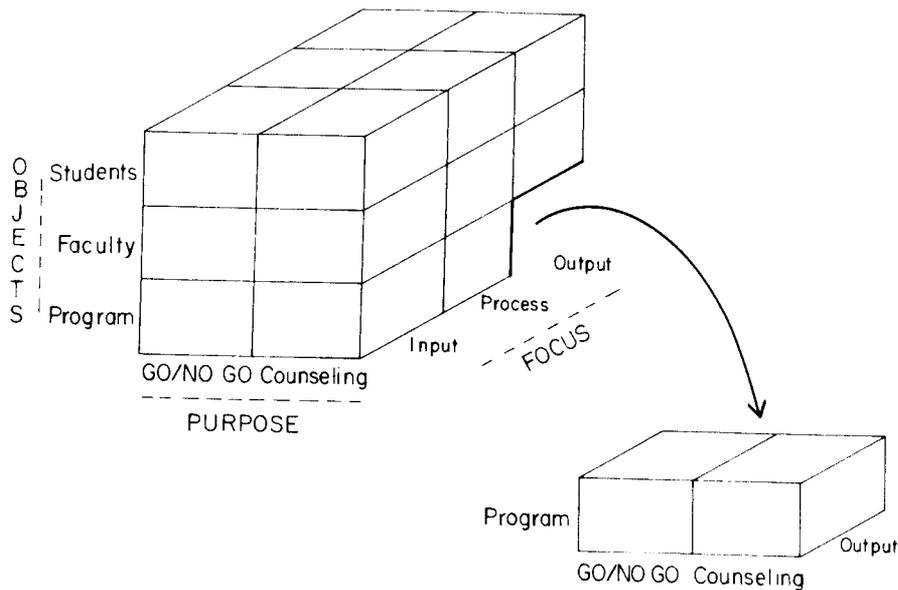


Figure 5: Faculty Evaluation for Promotion



evaluation, depending as they do on a subjectively administered standard of quality, applied through the process of peer review. The case presented below has been chosen to illuminate the assumptions and procedures typical of traditional methods of faculty evaluation and to examine the feasibility of utilizing more systematic techniques of measuring faculty performance.

Figure 6: Output Assessment in Program Evaluation



PROGRAM EVALUATION

The final example of the model of educational decision-making -- a case concerned with program evaluation -- illustrates the linkages between educational evaluation and evaluation of patient care (Figure 6). It is designed to demonstrate how epidemiological and patient care studies can be used in the determination of educational goals and priorities and in the evaluation of educational programs (8). First, an educational need is identified through a patient care study; second, a program is designed to meet the defined need; finally, a patient care study is used to determine the extent to which the educational objective has been achieved by the program provided. This methodology was originally employed in reviewing local hospital practice as a basis for setting educational priorities in, and evaluating programs of, continuing education (16). The concluding chapter of this volume presents sample data from recent national surveys of both ambulatory and hospital practice and shows how such data can be used in evaluating the product of educational programs at all levels.

SUMMARY COMMENT

In the five case studies which follow the purpose, focus and methodology of the evaluation are merely representative of a whole host of relevant issues with respect to student, faculty and program evaluation for purposes of educational decision-making. These concrete illustrations, provided to illuminate the model outlined in this chapter, are offered as suggestive of ways in which, in collaboration with appropriate experts, those who are responsible for policy can establish the type of information system requisite for more rational decision-making.

REFERENCES

1. American Educational Research Association, *Handbook of Research on Teaching*. N.L. Gage (ed.), Chicago: Rand McNally & Co. (1963).

American Educational Research Association, *Second Handbook of Research on Teaching*. R.M. Travers (ed.), Chicago: Rand McNally & Co. (1973).

These two volumes are only for the reader who desires a comprehensive, sophisticated discussion of the merits and limitations of alternative solutions to the problems of research design in this area.

2. Association of American Medical Colleges, *Zero Institutional Growth, Proceedings of the 1974 Council of Deans Retreat*, Washington, AAMC (1974).

A compendium of provocative essays on the prediction of, and implications for, administrations of colleges of medicine in a period of limited resources and minimal expansion.

3. Barrows, H. *Simulated Patient*. Springfield, Illinois: Charles C. Thomas (1972).

The author outlines the methods of creating simulated patients and discusses his extensive experience in using them in both the instruction and evaluation of medical students.

4. Charvat, McGuire and Parsons. *A Review of the Nature and Uses of Examinations in Medical Education*. W.H.O. Public Health Paper #36, Geneva: W.H.O. (1966).

Chapter IV provides a succinct summary of new techniques for product assessment, and a description of certain approaches to process assessment, other than interaction analysis.

5. Flanagan, J.D. "The Critical Incident Technique." *Psychological Bulletin*, 5:327 (1954).

Fivars, G. and Gosnell. *Nursing Evaluation: The Problem and the Process*. New York: The MacMillan Co. (1966).

The 1954 article furnishes a general description of the methodology; the later volume expands on, and illustrates more recent experience with the method.

6. Levine, H. and McGuire, C. "The Validity and Reliability of Oral Examinations in Assessing Cognitive Skills in Medicine." *Journal of Education Measurement*, 7:63 (1970).

Levine, H. and McGuire, C. "The Use of Role-playing to Evaluate Affective Skills in Medicine." *Journal of Medical Education*, 45:700 (1970).

The two articles describe the design, characteristics and results of nontraditional types of oral examinations for assessment of both cognitive and noncognitive components of competence at the specialty board level.

7. McGuire, C. "Research on Identification of Student Needs." In *Proceedings Fourth Annual Pan-American Conference on Medical Education*. Toronto (1972).

A detailed summary on concepts and findings of recent research on student needs, variously defined as "felt need," "unconscious demands on the environment," "inadequacy in preparation" (deficiency in entry behavior), and "lack of mastery" (deficiency in terminal behavior).

8. McGuire, C. "Educational Program Research and Development." In *HMO Program Development in the Academic Medical Center*, Robert H. Kalinowski and Stephen J. Ackerman (eds.) *Journal of Medical Education*, Part 2, 48:137-143 (1973).

A conceptual framework for utilizing patient care research in the definition of the goals, and evaluation of the effectiveness, of educational programs in the academic medical center.

9. McGuire, C. "An Overview of Applied Research in Medical Education: Problems, Principles and Priorities." In *Proceedings, 1974 Workshop on the Needs for Research in Medical Education*, Alexandria: Eastern Mediterranean Regional Office of the W.H.O. (1975).

A brief summary of the major directions and findings of research in medical education over the past decade, with comment on the implications of significant findings for educational planning, and selected references relevant to the principal research areas.

10. McGuire, C. and Solomon, L. *Clinical Simulations*. New York: Appleton-Century-Crofts (1971).

Ramsay, D. and Solomon, L. *Simulations in Dermatology*. New York: Appleton-Century-Crofts (1974).

These selected problems in patient management illustrate the use of written simulation in the instruction and assessment of medical students.

11. McGuire, C., Solomon, L., and Bashook, P. *Handbook of Written Simulations: Their Construction and Analysis*. New York: Harcourt-Brace-Jovanovich, (in press).

The authors present detailed instructions for selecting, writing, scoring, and analyzing problems in patient management designed to simulate clinical decision-making.

12. Pace, C.R. "College Environments." In *Encyclopedia of Educational Research*, Fourth Edition. Robert L. Ebel (ed.), New York: MacMillan Co. (1969).

A comprehensive evaluation of significant findings from research on the relation between student characteristics and environmental press, with extended bibliography.

13. Snow, R.E. "Representative and Quasi-representative Designs for Research on Teaching." *Review of Educational Research*, 44:265-291 (1974).

A comprehensive summary of findings from both parallel studies and alternative research designs on effectiveness of different instructional strategies, with extensive bibliography.

14. Snow, R.E. "Theory and Construction for Research on Teaching." *The Second Handbook of Research on Teaching*. R.M.W. Travers (ed.), Washington: American Educational Research Association (1973).

A description and critical evaluation of alternative models and conceptual structures for evaluation of, and research on, effectiveness of alternative instructional strategies.

15. Voytovitch, A.

An interesting report of this investigator's experience in utilizing medical records, for both instructional and assessment purposes with groups of medical students.

16. Williamson, J.W., Alexander, M., and Miller, G.E. "Continuing Education and Patient Care Research." *Journal of the American Medical Association*, 201:938 (1967).

Williamson, J.W., Alexander, M., and Miller, G.E. "Priorities in Patient Care Research and Continuing Medical Education." *Journal of the American Medical Association*, 204:303 (1968).

These two companion articles describe the general conceptual framework for utilizing results of patient care research to determine educational priorities and to evaluate programs in continuing education. Results of application of the technique to a particular hospital setting are discussed.

17. World Health Organization. *Development of Educational Programmes for the Health Professions*. W.H.O. Public Health Paper #52, Geneva: W.H.O. (1974).

World Health Organization. *Educational Strategies for the Health Professions*. W.H.O. Public Health Paper #61, Miller, G.E. and Fulop, Tomas (eds.) Geneva: W.H.O. (1974).

These two compendia contain useful articles on program planning, diagnostic and certifying examinations, assessment of attitudes, simulation and teacher evaluation in the health professions.

INFORMATION SYSTEMS FOR MONITORING STUDENT PERFORMANCE

James V. Griesen*

In 1962 the late chairman of anatomy at The Ohio State University, Dr. Grant O. Graves, launched an experiment in medical education which was to have a great impact upon the future of the College of Medicine (1-3). Displeased with the traditional methods by which anatomy was taught, Dr. Graves developed an independent study track which allowed students a considerable amount of freedom in meeting the course objectives. Lectures were replaced by small group discussions and individual tutorial sessions. Dissection exercises were made optional. Individual research projects and clinical experiences were arranged to provide the students with a source of motivation and a sense of relevance for the subject.

While difficulties were encountered, the independent study track improved and became more popular with each new class. By 1966 the majority of the entering class was electing independent study, and the logistical problems encountered in managing the track became more noticeable. With a large number of students progressing at independent rates, the faculty found it almost impossible to monitor student progress and identify areas of learning difficulty.

The students expressed a need for a method of testing themselves to gain assurance that they had mastered the objectives of the course. They wanted a reliable evaluation mechanism which would be readily available -- at night, on weekends, or during the lunch hour. They wanted a system which would adapt to their individual rates of progress, allowing *them* to decide when they were ready to be evaluated. They wanted an evaluation mechanism which would give them instant feedback about their performance, provide helpful advice, and prescribe further study if needed.

With these problems in mind, we began work in 1967 to develop a computer-based system of self-evaluation. After experimenting with a variety of strategies, we finally adopted a basic format for self-evaluation items, and labeled the approach a Tutorial Evaluation System (TES).

In actuality the format is quite simple, as illustrated in the schematic diagram presented in Figure 1. TES simulates the type of dialogue that a faculty member would engage in if he were conducting a student evaluation session in his office. In fact, Dr. Beth Wismar first described the item strategy to us in relating how she quizzed students in histology laboratory on the recognition of body organs.

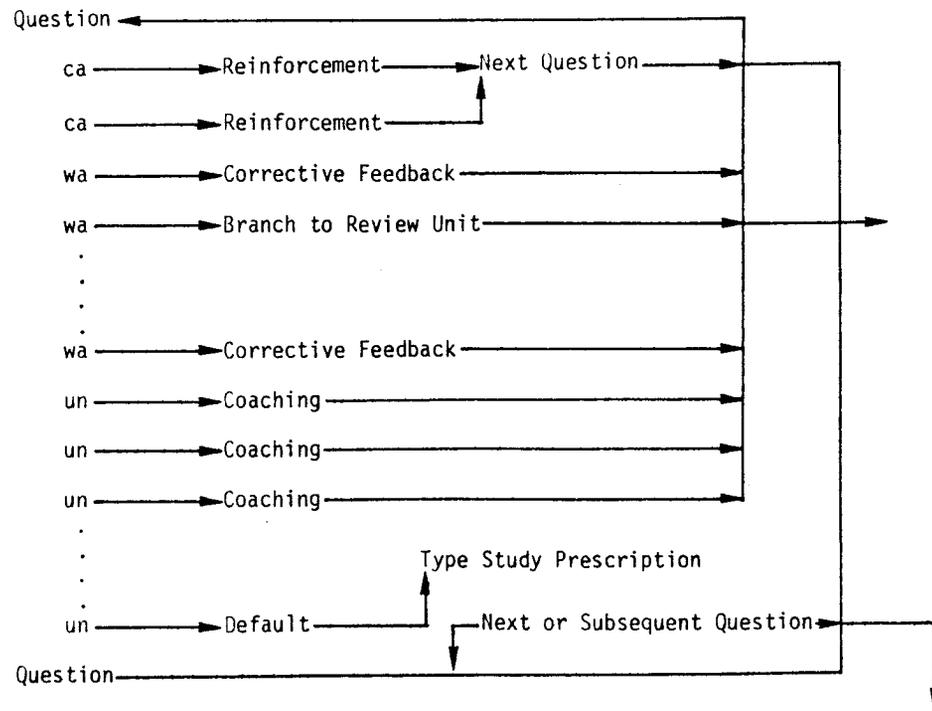
PROGRAMMED EXERCISE

Let us look at a sample item from an exercise that Dr. Wismar programmed in 1968, and is still using (and revising) today (4). The student is seated at a computer terminal which has a free-standing slide projector and viewing screen beside it. The exercise begins with a question being posed to the student. In the sample item presented in Figure 2, the student is instructed to view two photomicrographs of an unspecified body organ. The slides present two different magnifications of the histologic material, and the student is free to study them for whatever length of time he desires. Upon arriving at a judgment as to which organ is depicted in the photomicrographs,

*Director, Office of Educational Resources and Research, the University of Michigan Medical Center; former director, Division of Research and Evaluation in Medical Education, The Ohio State University.

the student merely types his response on the keyboard. The first three student replies illustrated in Figure 2 are identified as unanticipated responses, and the feedback is intended to "coach" the student into the range of anticipated responses. With each unanticipated response the feedback becomes increasingly specific. In his fourth reply our sample student finally enters an anticipated response, and even though it is misspelled the computer is able to properly interpret the organ name. Since this response (Duodenum) represents a common student error, the author has programmed a specific tutorial feedback statement intended to guide the student to the logically correct answer (being careful to properly spell the organ name in the feedback statement). Indeed, our sample student got the message and came up with the correct organ. The computer program was written to include the logic which would enable it to recognize any of the acceptable ways of stating the various correct and incorrect anticipated responses (e.g., large intestine instead of colon).

Figure 1: Diagram of a Tutorial Evaluation System Item



Where: ca = correct answer, wa = wrong answer, un = unanticipated answer

It was the success of the anatomy independent study track and the potential of TES as an instructional evaluation and management tool that convinced us to develop a plan for a comprehensive independent study program in the medical basic sciences (the traditional first two years of medical school). This plan was submitted to the Bureau of Health Manpower in the spring of 1968, and received full funding in the spring of 1969 (5). Since most of you are now familiar with the Ohio State Independent Study Program (ISP), further discussion about that program will be confined to the "information system for monitoring student performance" which was incorporated in the design. A complete description of the ISP -- its history and philosophy, curriculum, and evaluation results -- will be available soon in book form (6).

The Tutorial Evaluation System which had been developed for the Anatomy Department served as the basis for

Document from the collections of the AAMC. Not to be reproduced without permission.

Figure 2: Illustrative Tutorial Evaluation System Item From a Histology Organ Identification Exercise

COMPUTER: Study slides 1 and 2 and then type the name of the organ.

STUDENT: (After reviewing these slides on the rear-screen projector next to the terminal, he types his answer on the terminal keyboard)
I don't know.

COMPUTER: Your answer was not anticipated. Please check your spelling and try again.

STUDENT: I do not know the answer.

COMPUTER: No. This is a tubular organ. Think of possible tubular organs which would display the characteristics shown by the slides and try again.

STUDENT: I can't think of any.

COMPUTER: No. Observe the four layers in the wall of this organ -- mucosa, submucosa, muscularis, and serosa. This is highly indicative of the alimentary canal. Name the organ.

STUDENT: Duodenim.

COMPUTER: No. You could be misled if you assumed that the structures were villi when they are actually tubular glands. If these slides depicted the duodenum you would expect to see Brunners Gland; notice the absence of glands in the submucosa. Try again.

STUDENT: Colon.

COMPUTER: Correct. The characteristics you should have observed were crypts of lieberkuhn with a large number of goblet cells present as well as the absence of villi.

Figure 3: Illustrative Tutorial Evaluation System Item From the Independent Study Program

COMPUTER: Name the pathological conditions within the brain caused by blockage of the lateral and medial foramina of the inferior medullary velum.

STUDENT: I don't remember?

COMPUTER: Are you having difficulty placing the location of the inferior medullary velum and/or the significance of the foramina?

STUDENT: Yes.

COMPUTER: The inferior medullary velum forms part of the roof of the fourth ventricle. The foramina communicate with the cisterna magna and/or cisterna pontis. What is contained in these cisterna?

STUDENT: Cerebral spinal fluid (CSF).

COMPUTER: Correct. Now consider blockage of the foramina and answer the original question.

Name the pathological conditions within the brain caused by blockage of the lateral and medial foramina of the inferior medullary velum.

STUDENT: Water on the brain.

COMPUTER: Not acceptable. It's time to become more professional in your terminology!!

STUDENT: Hydrocephaly.

COMPUTER: Very good!!! The obstruction will cause an increased quantity of cerebrospinal fluid in the ventricular system.

the instructional information system utilized in the ISP. To further illustrate TES, an actual ISP student encounter which occurred at a computer terminal is presented in Figure 3. Interactions such as these comprise about an hour and ten minutes of the average ISP student's academic week. From the computer-stored records of these tutorial interactions it is possible to generate a variety of instructional management reports which aid faculty members in advising students and handling the logistical arrangements required by the program.

The intent of this presentation is *not* to persuade each of you to develop Computer-Assisted Instruction (CAI) as an evaluation or instructional management tool at your school. Rather, it is intended that this example will serve to illustrate the inherent value of developing *systems* for monitoring student performance. Too often we devote large amounts of money and faculty and staff time to the implementation of evaluation methods which are incomplete and inefficient. Reflecting on the example at hand, consider the advantages of the information system developed for monitoring student performance in the ISP:

1. The self-evaluation system was *developed as an integral part of the instructional program*, complementing the other main elements -- instructional objectives, learning resources, and formal examinations.
2. The system *provides students* :
 - a means of self-assessment
 - tutorial instruction (if appropriate)
 - immediate diagnoses of learning deficiencies and prescriptions for further study (if needed)
3. The system *provides faculty* :
 - a means of monitoring individual rates or progress
 - indications of areas in which individual students need tutorial assistance
 - logistical information to guide the scheduling of laboratory exercises and group discussions
4. The system *provides the program administrators* :
 - information about the effectiveness of the other major program components -- objectives, learning resources, and formal examinations
 - feedback regarding the effectiveness of the self-evaluation items themselves (through item analysis statistics)
 - logistical data necessary to manage program operations

A system for monitoring student performance can service a wide variety of informational needs in a comprehensive approach to managing the teaching/learning environment. If properly designed it will provide feedback regarding its own effectiveness, thereby providing a basis for continued improvement of the system.

To further illustrate the use of information systems in monitoring student performance, two different clinical rating forms are presented in Figures 4 and 5. These forms are utilized in clinical clerkship evaluation systems at The Ohio State University and The Wayne State University, respectively. In each case the form serves as a device for recording faculty perceptions of the student's abilities in various performance categories. While they clearly

Figure 4: Clinical Evaluation Form -- The Ohio State University (Page 1)



THE OHIO STATE UNIVERSITY
COLLEGE OF MEDICINE
CLINICAL EVALUATION FORM

27405469

DO NOT WRITE IN THIS AREA

STUDENT'S NAME		
DEPARTMENT	HOSPITAL	
REPORT COVERS	WEEKS, ENDING	MONTH DAY YEAR
STUDENT IS ENROLLED IN:		
<input checked="" type="checkbox"/> A REQUIRED CLERKSHIP	IF AN ELECTIVE, INDICATE:	
<input checked="" type="checkbox"/> AN ELECTIVE →	COURSE NAME	COURSE NUMBER
EVALUATOR'S LAST NAME	FIRST INITIAL	MIDDLE INITIAL
EVALUATOR'S STATUS: <input checked="" type="checkbox"/> ATTENDING STAFF <input checked="" type="checkbox"/> HOUSE STAFF		

CHARTS					
NOT OBSERVED <input checked="" type="checkbox"/>	UNSATISFACTORY WORK: POOR RECORDS OF HISTORIES AND PHYSICALS. <input checked="" type="checkbox"/>	CHARTS COMPLETE BUT UNORGANIZED AND REFLECT INADEQUATE UNDERSTANDING OF PATIENT. <input checked="" type="checkbox"/>	CHARTS PROMPTLY AND CAPABLY DONE. <input checked="" type="checkbox"/>	CHARTS CONCISE: REFLECT GOOD UNDERSTANDING AND FOLLOW-UP OF PATIENT. <input checked="" type="checkbox"/>	CHARTS ARE OUTSTANDING WRITTEN PRESENTATION OF THE CASE. <input checked="" type="checkbox"/>
INITIATIVE					
NOT OBSERVED <input checked="" type="checkbox"/>	NOT WELL MOTIVATED, UNPRODUCTIVE: AVOIDS "DOING" WHENEVER POSSIBLE. <input checked="" type="checkbox"/>	JUST GETTING BY: ACCEPTS REQUESTS BUT FREQUENTLY FAILS TO FOLLOW THROUGH. <input checked="" type="checkbox"/>	CARRIES HIS SHARE OF THE WARD LOAD; ACCEPTS REQUESTS AND SOMETIMES VOLUNTEERS. <input checked="" type="checkbox"/>	DOES MORE THAN HIS SHARE OF THE WORK; FREQUENTLY VOLUNTEERS. <input checked="" type="checkbox"/>	EXCEPTIONALLY HARD WORKER. CONSCIENTIOUS, VOLUNTEERS AND IS WILLING TO TAKE ON EXTRA WORK. <input checked="" type="checkbox"/>
PROFESSIONAL APPEARANCE					
NOT OBSERVED <input checked="" type="checkbox"/>	UNSUITABLE FOR THE OCCASION. <input checked="" type="checkbox"/>	USUALLY SUITABLE FOR THE OCCASION. <input checked="" type="checkbox"/>	SUITABLE FOR THE OCCASION. <input checked="" type="checkbox"/>	USUALLY CAREFULLY GROOMED AND APPROPRIATELY ATTIRED. <input checked="" type="checkbox"/>	ALWAYS CAREFULLY GROOMED AND APPROPRIATELY ATTIRED. <input checked="" type="checkbox"/>
KNOWLEDGE					
NOT OBSERVED <input checked="" type="checkbox"/>	SHOWS VERY INADEQUATE KNOWLEDGE OF MEDICAL SITUATIONS RELATED TO THE PATIENT'S PROBLEMS. <input checked="" type="checkbox"/>	SHOWS LESS THAN MINIMAL AMOUNT OF KNOWLEDGE OF MEDICAL SITUATIONS RELATED TO THE PATIENT'S PROBLEMS. <input checked="" type="checkbox"/>	SHOWS ADEQUATE COMPREHENSION OF BASIC MEDICAL PRINCIPLES AND RELATES THEM TO THE PATIENT'S PROBLEMS. <input checked="" type="checkbox"/>	SHOWS COMPLETE KNOWLEDGE OF THE BASIC MEDICAL PRINCIPLES AND DISPLAYS GREAT INSIGHT IN RELATING THEM TO THE PATIENT'S PROBLEMS. <input checked="" type="checkbox"/>	DEMONSTRATES INTELLECTUAL AGGRESSIVENESS WHICH QUALIFIES HIM AS THE LOCAL EXPERT ON THE PATIENT'S PROBLEMS. <input checked="" type="checkbox"/>
PROFESSIONAL JUDGMENT					
NOT OBSERVED <input checked="" type="checkbox"/>	DECISIONS AND RECOMMENDATIONS OFTEN WRONG OR INEFFECTIVE. <input checked="" type="checkbox"/>	JUDGMENT IS USUALLY SOUND BUT MAKES OCCASIONAL ERRORS. <input checked="" type="checkbox"/>	SHOWS GOOD JUDGMENT RESULTING FROM SOUND EVALUATION OF FACTORS. <input checked="" type="checkbox"/>	SOUND, LOGICAL THINKER; CONSIDERS ALL FACTORS TO REACH ACCURATE DECISIONS. <input checked="" type="checkbox"/>	CONSISTENTLY ARRIVES AT RIGHT DECISIONS EVEN ON HIGHLY COMPLEX MATTERS. <input checked="" type="checkbox"/>

Copyright College of Medicine, The Ohio State University

Reproduced with permission

Document from the collections of the AAMC. Not to be reproduced without permission.

Figure 4: Clinical Evaluation Form -- The Ohio State University (Page 2)

CLINICAL EVALUATION FORM (PAGE 2)

DEPENDABILITY AND ATTENDANCE					
NOT OBSERVED <input type="checkbox"/>	USUALLY ABSENT OR TARDY; CANNOT BE COUNTED ON FOR ADEQUATE PATIENT CARE. <input type="checkbox"/>	USUALLY PRESENT BUT SOMETIMES SPORADIC IN ATTENDANCE OR TARDY; GIVES PATIENT LESS THAN MINIMAL CARE. <input type="checkbox"/>	PRESENT AND CONSCIENTIOUS IN PATIENT CARE. <input type="checkbox"/>	ALWAYS PRESENT AND PUNCTUAL IN GIVING THE PATIENT COMPLETE CARE. <input type="checkbox"/>	ALWAYS PRESENT AND PUNCTUAL; SPENDS ADDITIONAL TIME IN WARD PROVIDING OPTIMAL PATIENT CARE. <input checked="" type="checkbox"/>
TEAM PARTICIPATION					
NOT OBSERVED <input type="checkbox"/>	MEMBER OF THE TEAM WHOSE BEHAVIOR UNDERMINES TEAM EFFORT. <input type="checkbox"/>	MEMBER OF THE TEAM WHOSE BEHAVIOR SOMETIMES UNDERMINES TEAM EFFORT. <input type="checkbox"/>	AN ACTIVE MEMBER OF THE TEAM WHO WORKS WELL WITH THE OTHER MEMBERS. <input type="checkbox"/>	AN ACTIVE MEMBER OF THE TEAM WHO ELICITS THE CO-OPERATION OF THE OTHERS. <input type="checkbox"/>	AN EXCEPTIONALLY ACTIVE MEMBER OF THE TEAM WHOSE LEADERSHIP QUALITIES ARE SOUGHT BY OTHERS. <input checked="" type="checkbox"/>
PROFESSIONAL RELATIONSHIPS					
NOT OBSERVED <input type="checkbox"/>	BEHAVIOR IS UNACCEPTABLE TO COLLEAGUES; DOES NOT COOPERATE; MAKES A POOR IMPRESSION. <input type="checkbox"/>	BEHAVIOR IS USUALLY ACCEPTABLE TO COLLEAGUES; COOPERATES WHEN NECESSARY; MAKES LITTLE IMPRESSION. <input type="checkbox"/>	MAINTAINS ACCEPTABLE AND WORKABLE CO-WORKER RELATIONSHIPS. <input type="checkbox"/>	ESTABLISHES ATMOSPHERE OF MUTUAL RESPECT AND DIGNITY WITH CO-WORKERS. <input type="checkbox"/>	COMMANDS ADMIRATION AND RESPECT OF CO-WORKERS; CONDUCTS HIMSELF AS A TRUE PROFESSIONAL. <input checked="" type="checkbox"/>
RELATIONSHIP WITH PATIENTS					
NOT OBSERVED <input type="checkbox"/>	UNABLE TO ESTABLISH APPROPRIATE RAPPORT WITH THE PATIENT. <input type="checkbox"/>	FAIR RAPPORT ESTABLISHED; SOMETIMES SEEMS TO BE A LACK OF COMMUNICATION. <input type="checkbox"/>	GOOD RAPPORT LISTENS AND COMMUNICATES HIS CONCERN FOR THE PATIENT'S PROBLEMS. <input type="checkbox"/>	NOT ONLY LISTENS AND COMMUNICATES WELL BUT INSTILLS CONFIDENCE IN HIS ABILITY. <input type="checkbox"/>	PROFESSIONAL ATTITUDE CONVINCES THE PATIENT OF HIS EXPERTISE AND PUTS THE PATIENT COMPLETELY AT EASE. <input checked="" type="checkbox"/>
OVERALL EVALUATION					
UNSATISFACTORY <input type="checkbox"/>	EFFECTIVE AND COMPETENT <input type="checkbox"/>				OUTSTANDING <input checked="" type="checkbox"/>

GENERAL COMMENTS (REQUIRED):

IF YOU HAVE ANY REASON TO QUESTION THAT THIS STUDENT WILL BE ABLE TO PERFORM CREDITABLY AS A PHYSICIAN, PLEASE CHECK THE BOX AT THE RIGHT AND EXPLAIN UNDER GENERAL COMMENTS.

Document from the collections of the AAMC Not to be reproduced without permission

Figure 5: Clinical Evaluation Form -- Wayne State University

WAYNE STATE UNIVERSITY SCHOOL OF MEDICINE
YEAR III CLINICAL EVALUATION FORM F

SEE THE BACK OF FORM
FOR INSTRUCTIONS

Nº 3902

STUDENT'S NAME: _____
 RATER: _____ RATER'S POSITION: _____
 DEPARTMENT: _____ HOSPITAL: _____
 PERIOD FROM: ___/___/___ TO: ___/___/___ SIGNATURE OF RATER: _____

DIAGNOSTIC ABILITY

Consistently unable to make use of available data to reach a differential diagnosis and plan logical diagnostic investigation.	Indulges in hunches and intuition while ignoring available data.	Eventually reaches a reasonable diagnosis and plan of management but excessive use of stressful diagnostic procedures.	Makes good use of data and arrives at logical differential diagnosis with use of only indicated diagnostic techniques.
--	--	--	--

Not observed

MOTIVATION

Does less than prescribed work, and that is incomplete and/or unsatisfactory	Does only prescribed work. Difficult to determine interest. Contributes occasionally verbally. Just enough to get by.	Does only prescribed work which is usually satisfactory, appears to work hard. Contributes verbally regularly and asks questions.	Does some extra work. Displays considerable interest and curiosity.
--	---	---	---

Not observed

SENSE OF RESPONSIBILITY

Displays little interest; needs repeated reminders of assigned duties.	Displays initial interest but usually does not follow through.	Displays initial and sustained interest in patients. Generally dependable with assigned duties.	Displays unusual interest in patients; performs duties promptly; makes unique contribution to patient care.
--	--	---	---

Not observed

ORAL VERBAL ABILITY

Frequently inarticulate.	Sometimes imprecise, obscure, ambiguous, or verbose.	Expresses himself in reasonably understandable terms.	Articulate even under stress.
--------------------------	--	---	-------------------------------

Not observed

EFFECTIVENESS WITH PATIENTS

Often antagonizes or generates a negative reaction from patients.	Patient relationships are superficial; has difficulty establishing effective rapport.	Able to relate effectively to most patients, but has yet to learn how to handle difficult situations.	Able to establish effective rapport with all types of patients. Wins the confidence and cooperation of all.
---	---	---	---

Not observed

ABILITY TO TAKE A HISTORY AND DO A PHYSICAL EXAMINATION

Physical examinations have major deficiencies in technical quality and thoroughness as well as incomplete and/or inaccurate history.	Minor features of history omitted and/or minor deficiencies in the physical exam.	Items of history and physical accounted for but has difficulty integrating the findings obtained from the history and physical.	Able to elicit a comprehensive history from patients and performs a technically accurate and complete physical.
--	---	---	---

Not observed

COOPERATION WITH PERSONNEL

Frequently does things which are thoughtless and causes unnecessary work for other personnel.	At times, does things which are thoughtless and creates unnecessary work for other personnel.	Generally does own work but neither helps nor hinders the work of others.	Carries full share of responsibility and is thoughtful, and concerned about helping other personnel to do their job effectively.
---	---	---	--

Not observed

DESCRIPTIVE COMMENTS: _____

Document from the collections of the AAMC. Not to be reproduced without permission

are not analytical assessment devices, each of the forms is structured to elicit faculty judgments about student performance with a reasonable degree of objectivity. The faculty members who developed the scales attempted to describe student performance in behavioral terms at various points on each continuum. While not absolute standards, they provide the individual faculty raters a clear and understandable description of performance at each point on the scale.

PERFORMANCE

At both schools the data provided via these forms are complemented by other performance assessment techniques. The Wayne State Medical School also administers patient management problems and standard objective examinations to clerkship students, and systematically reports composite profiles of student performance. The Ohio State Medical School utilizes oral examinations and standard objective tests to compliment the clinical performance assessments. In both cases there are faculty committees which oversee the development and operation of the clerkship evaluation system, and these committees are continually striving to improve the validity, reliability, and comprehensiveness of evaluation system. Both schools have well established offices of "medical education" which provided professional guidance in the initial development of the instruments and processes and then coordinated the administrative operations and continued revisions of the evaluation systems (The Division of Educational Services and Research at the Wayne State Medical School, and the Division of Research and Evaluation in Medical Education at Ohio State).

While still far from perfect, these information systems have inherent features which are clearly desirable:

1. They generate feedback to students at regular intervals, providing reinforcement for good performance and advising students of areas in need of improvement.
2. They generate current and cumulative performance records for each department and the appropriate administrators and committees.
3. They provide a means for determining patterns of student performance over long periods of time (several different clerkships).
4. They are designed to minimize paperwork.
5. They operate in a routine fashion and are well understood by both faculty and students.
6. They provide faculty members with detailed evaluation data and analytical summaries which will enable them to identify areas of instruction that need improvement.
7. They routinely provide statistical analyses of the evaluation measures themselves, thereby encouraging faculty to continually improve the system.

Again, the key concept in these examples is that an information system has been developed. Such systems need not involve the use to expensive computer equipment or elaborate and sophisticated procedures. They must, however, incorporate a comprehensive approach to satisfying a variety of information needs. They must also operate in a routine and consistent fashion, and provide feedback relative to their own effectiveness. If properly designed, they will generate an evolutionary process of continually improving effectiveness and efficiency in the instructional program and the evaluation system itself.

REFERENCES

1. Grant O. Graves, et al. "Need for a Multiple-Track System in Medical Education." *Journal of Medical Education*, 44, no. 5:344-348 (May 1969).
2. Grant O. Graves and Ralph W. Ingersoll. "Comparison of Learning Attitudes." *Journal of Medical Education*, 39, no. 2:100-111 (February 1964).
3. Lloyd R. Evans, Ralph W. Ingersoll, and James V. Griesen. "CAI at Ohio State University." *Computer-Assisted Instruction in the Health Professions*, Lawrence M. Stolurow, Theodore I. Peterson, and Anne C. Cunningham (eds.), Newburyport, Massachusetts: ENTELEK Incorporated (1970).
4. Beth L. Wismar and G. Ronald Christopher. "Computerized Histology." (Abstract) *Journal of Medical Education*, 44, no. 10: 983-84 (October 1969).
5. *Computer-Assisted Independent Study: A Pilot Program*. National Institutes of Health Grant No. 5 R01 PM 00040 (Bureau of Health Manpower Education, Department of Health, Education and Welfare, 1969).
6. Robert L. Folk, James V. Griesen, and Robert L. Beran. *Individualizing the Study of Medicine*. Oak Lawn, Illinois: Educational Products Inc. (1975).

EVALUATION OF PROBLEM SOLVING

Lee S. Shulman*

In this paper I shall review a number of approaches to the evaluation of clinical problem solving capabilities in medical students. The types of approaches will be described in the following ways: (a) important features of methods for evaluating clinical problem solving; (b) characteristics of each of the seven types of evaluation instrument selected for discussion; and (c) some of the virtues and liabilities associated with each approach. Clinical problem solving will be defined, for our purposes, as:

ACQUIRING)	
INTERPRETING)	INFORMATION
ORGANIZING)	

in order to make

DIAGNOSTIC)	
MANAGEMENT)	DECISIONS

I refer specifically to *clinical* problem solving in order, for purposes of this discussion, to exclude problem solving in the basic biological behavioral sciences, which may manifest somewhat different characteristics. The basic features of methods used to evaluate clinical problem solving are:

Fidelity of Situation -- the degree to which the situation which serves as the examination setting resembles the actual situation in which clinical performance occurs;

Similarity of Performance -- the degree to which the examined performance -- perceptual, intellectual, emotional, interpersonal, manipulative -- corresponds to the actual clinical performance in the field;

Nature of Feedback -- the purpose, focus and timing of feedback to the examinee of the results of the evaluation.

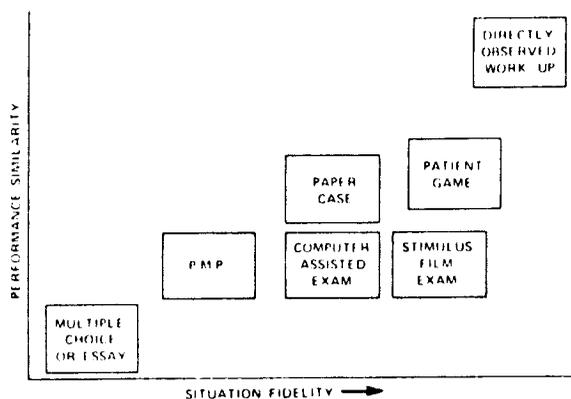
The first two features of evaluation methods can be arrayed (see Figure 1) as co-ordinates on which the seven types of evaluation discussed in this paper can be depicted.

In all these cases, *feedback* to students can vary according to its *purpose* -- diagnostic or certification decisions; its *focus* -- the processes students employ to solve a clinical problem or the correctness of the solution they reach; and its *timing* -- whether immediate, delayed (not necessarily inferior to swift feedback, and frequently superior), or totally absent.

I shall now review these seven varieties of evaluation, a sampling of their characteristics, virtues and liabilities.

*Professor of Medical Education, Office of Medical Education, Research and Development, Michigan State University.

Figure 1: Evaluation Methods as Co-ordinates of Types of Evaluation



THE DIRECTLY OBSERVED WORK-UP

The observed work-up can be conducted in many ways. First, the observation itself can vary from immediate, real-time supervision of a work-up, through the mediation of television or videotape, all the way to written records of a work-up, in the form of student notes to be reviewed or a formal patient record to be audited.

Second, the patient whose work-up is observed can be authentic, currently under examination and care; a chronically ill individual with stable signs who volunteers to assist in an evaluation setting; or a programmed patient, an actor or actress trained to simulate the history and, where possible, the physical signs of the condition represented.

The virtues of directly observed work-ups are their unparalleled situation fidelity and performance similarity. The liabilities consist of the costs of recruiting, training and re-training simulators; the costs of live observers, raters and video equipment; the time involved for all participants; the uncertain reliability of ratings made by poorly trained observers; and the need for a large repertoire of patients or simulated cases if the student is not to be limited to a narrow range of case contents.

In summary, this approach can involve alternative modes of observation (or approximations thereto) and several types of patients or patient simulators.

Modes of Observation

- Direct Observation (or live TV)
- Videotape and Playback
- Review of Notes on Work-up
- Review of Chart

Patient Types

- Real Patient
- Programmed Patient
- Chronically Ill Patient

VIRTUES

- * High Situation Fidelity
- * High Performance Similarity

LIABILITIES

- * Cost and Headaches
- * Questionable Reliability
- * Narrow Sampling of Content

MULTIPLE-CHOICE EXAMINATIONS

A variety of item types are possible. Among the many options are the standard, choose-the-best-answer item; the item which asks students to rank-order the alternatives according to some criterion; and the item which asks, along with selection of a best alternative, an estimate of the degree of confidence the respondent feels regarding his choice. This latter represents an important aspect of good clinical judgment, accurate estimation of likelihoods.

While lowest in both fidelity and performance similarity, the multiple choice examination does have the virtues of reasonable costs (though writing truly excellent items can require a surprisingly long time), opportunity to sample broadly across areas of clinical content and known, or at least knowable, test reliability. The major liability of these tests is that they tend to assess *what* someone knows far better than his ability to solve problems based on that knowledge. While in the hands of the most experienced and skilled item writers this may be far less a problem, the bulk of today's multiple-choice items measure knowledge of facts or simple principles.

It is difficult to imagine a comprehensive test of clinical problem solving without multiple-choice items. It is equally difficult to imagine an adequate test composed only of such questions.

In summary, the virtues and liabilities of these exams are:

VIRTUES

- * Broader Sampling of Clinical Content
- * Known (or Knowable) Reliability

LIABILITIES

- * Minimum Fidelity and Performance Similarity
- * "Knowledge" Oriented

PATIENT MANAGEMENT PROBLEMS

Patient management problems (PMP) were introduced some 15 years ago to provide an objectively scorable, paper-and-pencil complement to the multiple-choice examination, which would assess skill at acquiring and processing information in clinical judgment. They now exist in many forms, including the branching format employed by McGuire and her co-workers at the University of Illinois, the linear PMP typically used by the National Board of Medical Examiners, and the format employing a fixed list of information and decision alternatives, such as that of Helfer.

Their virtues include moderately high fidelity and performance similarity, relative ease of development and economy in use, and their provision of useful performance feedback on the basic components of the clinical work-up. Their growing use in broad certification examinations testifies to their usefulness. For example, the recent 1975 administration of the American Board of Internal Medicine examination devoted six hours to PMPs, up from four hours in previous administrations.

Their liabilities are few, albeit important. First, controversy remains on the effects of cueing -- the availability to the examinee of a list of alternative items among which he can select his choices. Second, there remain problems of scoring, that is, determining the most useful scoring categories and the best ways of establishing criterion keys. Finally, as with most methods of clinical problem solving evaluation, there remain questions of generalizability. How closely does performance on the PMP represent the examinee's likely performance in the real world? Research continues in a number of places to answer these questions. At the moment it appears that,

while capability to perform successfully on a PMP is no guarantee that the student will also behave effectively with real patients, *inability* to perform on the PMP provides strong evidence that outside performance will also be inadequate.

- Begins with brief patient characterization
- Involves sequential information: gathering and management
- May be linear or branching
- Provides both summary scores (e.g., Thoroughness, Efficiency, Accuracy) and optimal routes as feedback

VIRTUES

- * Moderate Fidelity
- * Ease of Development
- * Useful Feedback on Components of Work-up

LIABILITIES

- * Cueing
- * Unknown Generalizability

COMPUTER-ASSISTED EXAM

The computer-assisted exam comes in essentially two varieties -- the cued clinical problem, which essentially resembles a computerized PMP, and the uncued problem in which natural language is the median of communication between examinee and computer terminal. The essential features, virtues and liabilities of the computer-assisted exam are summarized briefly below.

- Comes in many varieties
- Often like a computerized PMP
- Can be cued or uncued
- Realistic time factor can be included
- Scoring based on criterion group

VIRTUES

- * Flexibility of Branching
- * Can Be Uncued, When Using Natural Language

LIABILITIES

- * Cost of Development
- * Technological Unreliability
- * Unknown Generalizability

PAPER CASES

Paper cases resemble natural language computer-based problems. They are uncued, i.e., students must supply their own questions and inquiries. They are also linear, that is, all students pursue essentially the same path through the problems. They thus combine some of the virtues of the computer-based exam (of course lacking advantages of flexibility and automation which characterize the computer), without the high costs and technical problems.

I briefly characterize the features, virtues and liabilities of the paper case exam below.

- Open-ended, uncued, linear patient problems
- Student formulates own responses
- Information supplied as needed to pursue case

- Feedback available during and after case
- Criterion group for scoring

VIRTUES

- * Ease of Construction
- * Higher Response Similarity
- * Flexible Format

LIABILITIES

- * Linear
- * Unknown Generalizability
- * Scoring Problems
- * Can Be Uncued, When Using Natural Language

STIMULUS FILM EXAM

The stimulus film can constitute the basis for an entire examination, or it can serve as the "front-end" for almost any other exam format -- multiple choice, PMP, etc. The features on *one type* of stimulus film exam are briefly reviewed below.

- "Physician's eye" view of patient
- First five minutes of work-up
- Student called upon to interpret and organize (not acquire) information
- Criterion group scoring
- Feedback possible during and after presentation

VIRTUES

- * High Performance Fidelity
- * Can Be Used As "Front End" for Other Techniques
- * Focuses on Specific Component Clinical Skills
- * Student Constructs Own Response
- * Effective Feedback for Students

LIABILITIES

- * Costs and Complications of Production
- * Single Interview Sequence or Style Per Film
- * Limited Response Categories for Students

PATIENT GAMES

Patient games are simulations which combine interaction with an examiner, slides, films, photos, physical models, written records and other formats to represent the initial encounter, work-up and subsequent course of a single case. The cases used thus far have been of an emergency nature, hence the format must also provide for changes in signs, symptoms and lab values over time. The examiner's response to the students' (uncued) questions, investigations, management decisions or referrals is controlled by an elaborate set of time-dependent flow diagrams. This method, like most of those discussed earlier, can be employed both for instruction and evaluation. The essential features of Patient Games are:

- Multi-media simulation for situation fidelity
- Multi-modality responses available to students
- The situation is fully interactive
- Circumstances change realistically over time

-- The approach is both sequential and branching

VIRTUES

- * High Fidelity
- * Realistic Time Change
- * Branching, Uncued
- * Constructed Responses
- * No Unrealistic Writing

LIABILITIES

- * Costs and Difficulties of Production
- * Training of Examiners
- * Some Inflexibility of Routes of Inquiry Open to Students
- * Scoring Problems

In conclusion, no single method for evaluating problem solving possesses the full range of virtues without attendant liabilities. In order to optimize across such characteristics as scope of content, fidelity, performance similarity, effective feedback, easy and economy of construction, flexibility of format and the like, a *program* of clinical evaluation is needed which combines many methods -- mixing, matching, and marrying techniques until a maximally virtuous, minimally liable program is achieved.

RESOURCES AND REFERENCES

The following list is not exhaustive, or even necessarily representative. Those named below are people (or books) I know and whose work I respect.

DIRECTLY OBSERVED WORK-UP

1. Programmed Patients (direct or videotape and playback)

- * Dr. Howard Barrows
McMaster University
Hamilton, Ontario
- * Dr. Norman Kagan or Dr. John Schneider
Michigan State University
East Lansing, Michigan

2. Chronically-Ill Programmed Patients

- * Dr. Alex Bryans
Queens University
Kingston, Ontario

MULTIPLE-CHOICE EXAMINATIONS

(Many books have been published, of which the following are particularly recommended)

1. Ebel, Robert L. *Essentials of Educational Measurement, 2nd Edition.* Prentice-Hall (1972).
2. Hubbard, John P. *Measuring Medical Education.* Lee and Febiger (1971).

PATIENT MANAGEMENT PROBLEMS

The following book presents samples of excellent PMPs.

1. McGuire, C. and Solomon, L. *Clinical Simulations.* New York: Appleton-Century-Crofts (1971; new edition forthcoming).
2. Christine McGuire
Abraham Lincoln Medical School
University of Illinois
Chicago, Illinois
3. Ray Helfer
Department of Human Development
Michigan State University
East Lansing, Michigan

COMPUTER-ASSISTED EXAM

1. William Harless
Pacific Medical Center
San Francisco, California
2. Octo Barnett
Harvard University (MGH)
Boston, Massachusetts
3. Richard Friedman
University of Wisconsin
Madison, Wisconsin

PAPER CASES

1. Sarah Sprafka
Office of Medical Education, Research
and Development
Michigan State University
East Lansing, Michigan

2. Lee Shulman
Office of Medical Education, Research
and Development
Michigan State University
East Lansing, Michigan

PATIENT GAMES

1. Jack Maatsch
Office of Medical Education, Research
and Development
Michigan State University
East Lansing, Michigan
2. Gerald Holzman
Office of Medical Education, Research
and Development
Michigan State University
East Lansing, Michigan

The only published reference to Patient Games, which can be ordered from Dr. Maatsch, is:

Fundamentals of Patient Games
Office of Medical Education
Research and Development
Michigan State University, 1974

AN ANALYSIS OF STUDENT PERFORMANCE AS REFLECTED IN THE NON-THREATENING AUDIT
OF THE STRUCTURED MEDICAL REPORT

Anthony E. Voytovich*

While at Case Western Reserve University between 1971 and 1974, I became indebted to Dr. C. H. Rammelkamp for providing me with the unique opportunity of spending a large amount of time in the planning and administration of, and experimentation with, the day-by-day operation of a basic clerkship in medicine for third-year students at the Metropolitan General Hospital. It was a ripe opportunity, and I undertook the task with all the bristling momentum of a just-completed chief residency. I felt heavily "armed" with a familiarity with the system -- plenty of free time -- and a burning memory of a lack of direction as a third-year student myself. The clerkship would be my laboratory!

Two facts seemed clear at the outset:

1. Day-by-day (if not hour-by-hour) evaluation with a ready feedback-loop is the best teaching there is. It has immediacy, it is most likely to be at the cutting edge of the students' understanding, and most importantly, it involves what they actually do rather than simply what they know!

2. Since you cannot be with all the students all the time, the system for recording what has been occurring and for communicating those thoughts must be optimum. I personally felt that verbal presentation was not the answer. That requires the simultaneous presence of both parties and a bilaterally optimum frame of mind. Most of all, I fondly recall steering the course of teaching or attending rounds myself as a house officer by what I chose to present or not to present or color as important. The direction in many of those verbal teaching sessions I remembered as distinctly resembling Brownian movement with a lot of the learning a product of chance collision rather than deliberate analysis.

STRUCTURED RECORD

It seemed that certain aspects of a properly structured medical record might serve as the all important medium. Such structure provides goals and objectives not only for a single patient but for single events within that patient's course at a level of detail nearly impossible in any other way, and with a simultaneous capacity for rapid overview completely impossible in any other way. It appeared that we had a very neat sort of "zoom" lens on the students' day-by-day performance.

The structured record then provided the rules by which we played the game. It had always seemed to me that the players were active; the ball was in motion; evaluation was occurring, but that the rules were vague. The students and I now had a contract. They generated and structured their information in a pre-agreed format -- with pre-agreed standards -- and I would take the time to tell them how they did. We both knew what was expected beforehand.

*Assistant professor of medicine, School of Medicine, University of Connecticut Health Care Center.

The elements of structure were several and clear enough that, within the first week, everyone understood them (1,2).

The interaction was called an "audit." I would go to the students' notes, sometimes invited, sometimes not, and could generate a two-to-three page written critique which I would seal in an envelope and send to the student for consideration at his leisure. I asked for a written reply point-by-point. This served to:

1. take away the feeling of helplessness and make the student/auditor interaction more of a two-way street, with no one feeling like a victim;
2. provide me with reinforcement and with some guarantee that my efforts had an impact!

The hundreds of recorded interactions which I have provide some fascinating insights into the struggle of a junior clerk in medicine.

I'd like to select just one pair of the rules by which we play to illustrate how much insight into the students' thinking they provide.

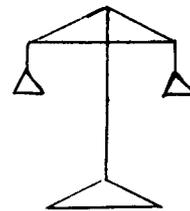
TWO RULES

Figure 1 shows the two rules by which it becomes possible to analyze the students' concept of how his abnormal data fits together. His data and his conclusions must balance precisely, for omissions and guessing in any quantity permit facets of incomplete understanding to become obscured in the haze!

Figure 1: Complete Precise Problem List

RULE 1 - All the abnormal data shall be accounted for.

RULE 2 - Each problem shall be defined as precisely as the data at hand will allow -- Neither above nor below.

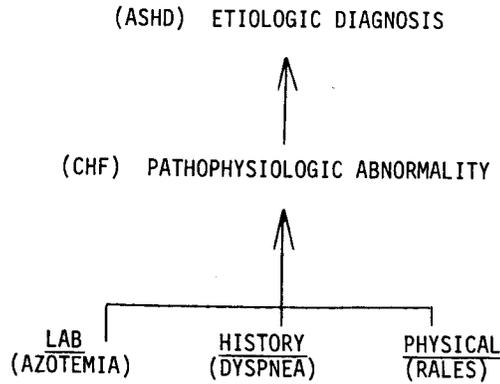


The two rules, and the symbolic balance stressing their implication may, at first glance, seem obvious and descriptive of any work-up in any system. However, like a balanced accountant's ledger or the superb performance of a piano concerto, there is a huge difference between "nearly correct" and "correct." Measurement requires strict, almost rigid application and so it is with the structured record. Failure to include a bit of proteinuria in the problem formulation, like allowing your checkbook to be \$2.32 off balance, creates the uneasy feeling that there may be, or have been, mistakes of a more serious nature along the way.

Leaving out clues in formulating the problems, or arriving at conclusions unjustified by the data at hand, even if minor, *softens* the measurement and allows gaps in understanding to be lost in the ambiguity; therefore, the rules are rigidly agreed upon, even to the level of trivial detail since this is where the cutting edge of the understanding frequently lies!

Figure 2 represents the framework within which the rules apply. Problems may be defined at any one of the three levels of understanding from the lowest (isolated observations), through initial synthesis (pathophysiologic abnormalities), to a definitive conclusion (etiologic diagnosis). The assembly of the bits and the movement

Figure 2: Levels of Definition



upward through higher levels of definition then became a disciplined exercise driven by the students' logic and understanding as the most important variable in the system.

If we agree on a perfect balance between the abnormal data on the one hand, and the way it is clustered and labelled on the other, few students do it perfectly at first, if ever. They make mistakes. Our studies have characterized the breakdown in the application of the "rules" and isolated these behaviors as specific targets in our teaching. Figures 3-6 describe the four specific errors in problem formulation that we have observed in hundreds of audit situations.

Figure 3 describes "omission" or the simple act of failing to include, in the analysis, some abnormal bit of information elicited and recorded earlier in the work-up. The applicable rule from Figure 1 is included. The underlined words stress the break in contact.

Figure 4 describes an error in analysis in which the student's conclusions go beyond what his data allow. Again, the appropriate rule is indicated and underlined. We call this "premature closure."

Figure 3: Errors in Formulation

#1 OMISSION

Lab Data)	Unused data . . .
Symptom)))	unaccounted for on
Physical Finding)	problem list

 RULE 1: All the abnormal data shall be
accounted for.

Figure 4: Errors in Formulation

#2 PREMATURE CLOSURE

Overuse of data -- unjustified conclusions drawn

PULMONARY INFILTRATE + COUGH = PNEUMONIA

(OR)

ANEMIA + LOW RETICS = Fe. DEFICIENCY

 RULE 2: Each problem shall be defined as precisely as the data at hand will allow
 -- Neither above nor below.

Figure 5 describes "inadequate synthesis" (the opposite of premature closure), in which the available facts allow conclusions well beyond where the student's reasoning has taken him.

In Figure 6, we describe "incorrect synthesis" in which not only do the data and the conclusions fail to justify each other, but in fact, are mutually exclusive or incongruent.

Figure 5: Errors in Formulation

#3 INADEQUATE SYNTHESIS

Underuse of data (Timid use of information)
 MYOCARDIAL INFARCTION + ABNORMAL ANGIOGRAMS =
 "ORGANIC HEART DISEASE"
 (OR)
 MICROCYTIC HYPOCHROMIC ANEMIA + ABSENT Fe.
 STORES = "ANEMIA"

 RULE 2: Each problem shall be defined as precisely
 as the data at hand will allow -- Neither
 above nor below.

FOUR BEHAVIORS

These four behaviors, then, represent the ways in which students find themselves unable to adhere to the agreed-upon rules. Our arrangement disallows such excuses as: "It isn't important anyway," or "I was too busy to think about that," or "Oh, I know that, I just didn't put it down." The agreement between student and auditor allows these errors to be analyzed as such. As in any system of measurement, one simply cannot proceed until the units are identified and agreed upon.

Given that these mistakes occur, how common are they? Are some more common than others? Figure 7 represents the distribution of errors in 100 student audits. The range per case study and the total number of work-ups in which such errors appear are indicated. Clearly, omission is the most common and, in this sample, "premature closure" is more likely than "inadequate synthesis."

Figure 7: Distribution of Errors in Synthesis in 100 Audits (Students)

	Total	Range	Students Involved
Omission	107	0-3	59
Premature Closure	55	0-3	39
Inadequate Synthesis	33	0-2	30
Wrong Synthesis	4	0-1	4

Figure 6: Errors in Formulation

#4 WRONG SYNTHESIS

Conclusions based on erroneous interpretation
 HCT. 32 + RETICS. 9.5% = "APLASTIC ANEMIA"
 (OR)
 pH 7.41 + HCO₃ - 28 + Tr KETONURIA =
 "KETOACIDOSIS"

 RULE 2: Each problem shall be defined as pre-
cisely as the data at hand will allow
 -- Neither above nor below.

Who makes these mistakes? Are they distributed across the classes in general, or do students who appear to be having difficulty in the traditional system of evaluation contribute more heavily to the total?

Figures 8 and 9 represent an analysis of the occurrence of errors in synthesis as correlated with clerkship performance measured in the traditional way. (Summary evaluation by two sets of house staff, attendings, and preceptors).

Both the raw data in Figure 8, and the calculation of errors-per-student in Figure 9 suggest that those who fare badly in the traditional evaluation system tend to have more difficulty in assembling their facts according to the agreed-upon rules than those who do well in the traditional system!

Figure 8: Distribution of Errors by Clerkship Grade

	Honors (11)	Satisfactory (32)	Failing (7)
Omission	6	45	18
Premature Closure	4	22	6
Inadequate Synthesis	3	12	2
Wrong Synthesis	<u>0</u>	<u>1</u>	<u>3</u>
TOTAL	13	80	29
TOTAL PER STUDENT	(1.18)	(2.50)	(4.14)

Figure 9: Per Student Errors By Grade

	Honors	Satisfactory	Fail
Omission	0.55	1.40	2.57
Premature Closure	0.36	0.69	0.86
Inadequate Synthesis	0.27	0.38	0.29
Wrong Synthesis	0.00	0.03	0.43

Some might react to the correlation between traditional evaluation and results of audit by saying, "so what!" "It is no surprise that poor students make more mistakes than good ones -- you simply have validated the status quo!" It is true we always knew that good students do better and poor students worse . . . but in what exact ways? We needed a higher level of resolution if we were to transcend the gross observation that Student A is better than B. We need a higher grade of data. The audit technique does more than corroborate traditional methods -- it *extends them* and provides a higher degree of resolution of what is good and what is bad -- not only in a given individual -- but in a given individual at a certain point in time with a particular patient and over a particular issue. With evaluation at this level of detail, fed back, day-to-day, efficiently, and with immediacy, we can do more than evaluate; we can *close the loop* and teach as part of the process.

STUDENT/TEACHER INTERACTION

It has been difficult for me to dwell on the detached numeric analysis of what I consider to be a highly personal and sensitive kind of student/teacher interaction (3). We have striven to provide a kind of coaching atmosphere of invited trust and I have spent considerable effort at determining the best way to make the audit a non-threatening highly constructive kind of detailed evaluation. Each of the elements in these figures was delivered in an individualized, deliberate, calculated attempt to *stimulate insight* rather than point out a mistake.

Errors were not treated as cold facts, rather, the audit would ask: "might it help us to know the serum calcium?" Such phrasing, carefully calculated, tends to result in a response reflecting the stimulation of curiosity and the pleasure of insight and discovery, rather than embarrassment at being "found out."

I remain indebted to Dr. Thomas Hale Ham for his warm and valuable encouragement and advice in the evolution of this approach.

AUDIT TECHNIQUE

I realize that this audit technique is not a widely and immediately applicable one on a broad scale, for it requires time and continuity. Bearing this in mind, we experimented with the concept of asking students to audit one another. This was an immediate success. The students participated willingly both as auditors and subjects. They were excited at being able to help one another and amazed at the insight they gained by the critical and tactful analysis of the efforts of their peers. Early success with this "round robin" approach has certainly stimulated our interest in looking toward a new direction in the analysis of clinical performance -- namely peer and self-evaluation in a program of continuing education.

REFERENCES

1. Voytovich, A.E. "The Student and the Problem-Oriented Record: Bane or Blessing." *Bulletin of Case Western Reserve Medical Alumni Association*, 37, no. 1:4-6 (1st Quarter, 1973).
2. Voytovich, A.E. *Understanding the Problem-Oriented Record*. Press of Case Western Reserve University (1974).
3. Voytovich, A.E. "Evolution of a Non-Threatening Audit System in a Large Teaching Hospital." *Applying the Problem-Oriented System*, Hurst, J., and Woody, M. (eds.), New York: MEDCOM Press (1973), pp. 307-315.

THE EVALUATION OF MEDICAL FACULTY

*Hilliard Jason, Henry B. Slotnick, Helen Eden, and Luis Patino**

Faculty evaluation is a matter of both academic and high practical concern to medical school Deans. It is an issue that has legal as well as institutional implications. Part of the problem in faculty evaluation--probably larger than the problems which we face with student evaluation--is the relative absence of a tradition for gathering systematic data, and the substantial absence of mechanisms for analyzing or interpreting data. Yet there is the inevitable, persistent need for decisions to be made; decisions on who gets hired, who gets promoted, who receives tenure and does not, who, if the institution has adopted such practices, should be advised to shift their emphasis in their activities, time, or focus. Ultimately, all of these decisions are the responsibility of the dean, whether or not he delegates the process to others.

There are a growing number of medical schools where these decisions are being decentralized, in various ways, through the committee structure. But, it is fair to say, in virtually all institutions these decisions are being made in behalf of the dean, not to the exclusion of the dean; and he is still expected to at least ratify, and occasionally challenge, these decisions. It is with this in mind that the topic of faculty evaluation was included in the spring, 1975 program of the Council of Deans meeting.

The decisions to be made in the faculty assessment area illustrate most vividly the essential ingredient of the exercise of leadership in human institutions: decisions must repeatedly be made with insufficient data. Basically, it is not a decision, in the mathematical sense, that is being made; it is judgment that is being exercised. Such judgments include the weighing of imponderables and the dependence on intuition. We simply do not approximate the kind of precision that society would like us to have in these situations. Such precision would lead to an elimination of the ambiguities and uncertainties that result in law-suits and intra-institutional confrontations. As nice as it would be to avoid disputes, the elimination of uncertainty is not a reachable objective at this time, if ever. Judgment will likely continue to be necessary for as long as we have human institutions. The only way to eliminate ambiguities and uncertainties is to choose to attend exclusively to easily quantifiable attributes; thereby simplifying the decision-making process and doing violence to the people involved, to the ultimate detriment of the institution.

*Respectively, director, evaluation coordinator, research assistant, and workshop coordinator, Division of Faculty Development, Department of Academic Affairs, AAMC.

The primary issue in the exercise in which the deans engaged is a search for a method that can facilitate judgment-making; that can optimize the utilization of those data which are available. Our hope is to minimize, although we cannot eliminate, the uncertainties that are involved in the array of decisions surrounding faculty evaluation.

It was our judgment that the best way to embark on an analysis of these very difficult issues would be to invite the assembled deans actually to participate in a decision-making exercise. It would be done under circumstances that would give all of us an opportunity to understand some of the essential ingredients of decision-making and to help us move toward the reduction of uncertainty and an increase in systematization. The deans were asked to engage in a simulation. Inescapably, it was artificial. If it were real it would take more time and resources that we had at our disposal. They were asked to eliminate from consideration factors that are of vital importance in their own institutions. They were asked, for example, not to consider the special characteristics of their medical schools or the unique features of the departments that these decisions would affect. They were asked exclusively to make decisions about individual people.

The scenario was that, in the round of promotion recommendations this year, the clinical departments of their school had put forward the names of nine candidates for promotion from assistant professor to associate professor with tenure. Budgetary considerations being what they are, only two people could be funded for promotion and tenure. The Promotion and Tenure Committee had succeeded in eliminating five of the nine candidates and the dean was assumed to be satisfied that these eliminations were justified. The Committee was now deadlocked on the remaining four candidates, and turned to the dean to arbitrate the decision as to which two of the remaining four deserved to be promoted and receive tenure. The deans were faced with unavoidably limited amounts of data on these four candidates, and were asked to make a straightforward decision: which two will be recommended for promotion. The situations with which they were confronted were derived from the actual records of real faculty members that had been under consideration for promotion in real medical schools. The records were, of course, reworked to protect the anonymity of both the individuals and the schools.

In addition, the deans responded to the exercise anonymously. The results were analyzed and are reported only in terms of the overall group.

There were some purposeful characteristics of this activity that were not evident from the single task with which each individual dean was faced. Each, randomly, had a somewhat different challenge from each other. There were no "tricks" involved in this design. It was a straightforward effort to simulate the kind of difficult problems that deans face frequently: with insufficient, or occasionally contradictory, data. This exercise permitted us to examine some of the elements of this decision-making process.

The three variables of interest (research, service, and teaching) were represented dichotomously: candidates were described as being "acceptable" or "exceptional." "Acceptable" meant that the candidate had met the requirements for promotion to Associate Professor: his

research activities resulted in the publication of four articles in refereed journals with his being the senior author on two of them and the presentation of papers at annual meetings; his teaching activities were said to be respected by his colleagues; and the candidate was described as serving successfully on university committees and having volunteered his services to community agencies.

"Exceptional" meant the candidate had done everything attributed to acceptable candidates and more: the candidate had published seven articles in refereed journals (being the senior author in four of them), had presented six papers, and was currently the recipient of a federal grant; the candidate was described as well respected by his colleagues and having participated in some non-teaching educational activity (such as materials development); and the candidate had served on one or two university committees and had also actively participated voluntarily in various national organizations.

Each dean received a packet containing "dossiers" on four candidates. Each dossier contained a letter of nomination from the candidate's department chairperson, which included an opening paragraph, communicating both support for the candidate's promotion and an indication of the candidate's adequacy in terms of patient care, and three paragraphs describing the candidate's contributions in terms of research, service, and teaching, and indicating five years' membership on the faculty. A bibliography was also included in support of the information on research described in the letter.*

The dossiers were assembled in packets of four for each dean. The characteristics attributed to the candidates were systematically varied so as to assure that among the four candidates being considered by any one dean two would be "adequate" and two would be "exceptional." The particular mix of the exceptional and adequate attributes in each letter and packet was determined by lottery. Thus, research, service and teaching achievement were independent of each other, and no set of four dossiers were identical to any other set; everything was done using random selection without replacement.

Half the packets had a rating form, which was a table designed to help organize the data (Fig. 1). Many people find that they are helped by the kind of visual organization that comes from being able to lay out an array of observations, so as to have a simultaneous contrast of each of the options with each other. We were trying to examine, in this very small way, whether the rating form helped their decision-making process. The basic task was the same for the half of the group that did not have the rating form. We subsequently learn from some of the deans who had not been given a rating form that they created one of their own. Apparently, experienced decision-makers, like many of the assembled deans, learned the value of rating forms and often generate one in their heads or on paper, when faced with decisions involving several variables.

*Bibliographies cited publications appropriate to the candidate's field. They varied, in terms of numbers of publications and in terms of the prestige of their journal locations, according to whether the candidate was an "acceptable" or "exceptional" researcher.

FIGURE I

The following rating scale was designed to help your assessment of the strengths and weaknesses of each of the candidates, and to facilitate your final decision. Put a check indicating your rating for each candidate on each variable.

<u>TEACHING</u>			<u>RESEARCH</u>			<u>SERVICE</u>			<u>PATIENT CARE</u>		
1	2	3	1	2	3	1	2	3	1	2	3

Dr. Miller
 Dr. Lewis
 Dr. Lorimer
 Dr. Spindler

*NOTE: 1 = Strong, 2 = Moderate, 3 = Weak

We have further divided each of the half-groups into three subgroups, in terms of the kind of information provided about each candidate's instructional strengths. One-third of the group had only the descriptive paragraph about the candidate's teaching contribution in the letter of recommendation, with no supporting data. Another third of the group had what might be called nonspecific or general data in support of the narrative evaluation of their teaching. This was a summary of student reaction forms, giving only minimally helpful information about the teacher (Fig. II).

FIGURE II

SAMPLE STUDENT REACTION FORM
 "NONSPECIFIC INFORMATION"

Medical Student Council
 December 31, 1974

Dear Dr. _____:

Based on the 20 third year students completing the Student Opinion/Faculty Teaching forms describing you as an attending physician, the following summary statements can be made:

* Student Responses

	<u>Generally</u>	<u>Often</u>	<u>Seldom</u>
Dr. was a good teacher	70	25	5
This Dr. should teach more.	60	35	5
More Drs. should teach the way this one does.	75	20	5
Dr. knew how to deal with medical students.	75	25	0
Dr. understood medical students' needs.	65	30	5
Dr. was pleasant.	75	20	5
Outside work seemed appropriate.	30	45	25

When a group of students tells you that a faculty member was a "good teacher," you cannot know the basis for each individual's decision. Were some approving of the teacher because he is an easy grader, or because he gives free beer at evening sessions, or because he truly teaches

well? This might be considered general, nonspecific data, which is only slightly better than no information.

One-third of the group got what we have called specific support data on teaching. This is the kind of information that is now beginning to be available in some institutions, derived from well-designed student reaction forms (Fig. III). If a student says he felt respected as a person, or found the teacher's questions insightful and understandable, that is useful information. Whether it is critically important or not to the student's learning is a separate question, but it is descriptive, reliable information about that teacher's contribution. This kind of information is considered by some authors (Slotnick and Durkovic, 1975) to be useful in the improvement of teaching because it provides instructors with specific feedback they can use to examine and alter their teaching. Participation in non-teaching educational activities (e.g., development of instructional materials) was also reported.

We were interested in understanding the ways that you might differ in terms of the relative importance you attach to the major components of a faculty member's contribution: research, teaching and service. To cross-check these factors against each other, we created eighty different folders. Each had a unique combination of data (Fig. IV). Half of the candidates that were being evaluated had descriptions and support data about their research contributions which might be considered just adequate for someone ready to be an associate professor (identified as "R-", in Fig. IV).

FIGURE III
SAMPLE STUDENT REACTION FORM
"SPECIFIC INFORMATION"

	% Student Responses		
	<u>Generally</u>	<u>Often</u>	<u>Seldom</u>
<u>Student/Faculty Rapport</u>			
Students felt they were respected as persons.	70	25	5
Students were comfortable during the Dr.'s rounds.	60	35	5
Students respected the Dr.	75	20	5
Dr. addressed the needs of the students as physicians-in-training.	75	25	0
Dr.'s expectations for the students were reasonable.	65	30	5
<u>Outside Work</u>			
Outside readings were useful.	75	20	5
Outside work was not excessive given the time allocated.	30	45	25
<u>Dr. as Teacher</u>			
Impromptu lectures were well-organized.	55	35	10
Dr.'s questions were insightful.	65	30	5
Dr.'s questions were understandable.	60	35	5
Dr.'s answers to student questions were understandable.	65	30	5
Dr. summarized cases well.	55	45	0

FIGURE IV
SIMULATION MODEL

Rating	R-				R+			
	S-		S+		S-		S+	
	T-	T+	T-	T+	T-	T+	T-	T+
N								
G								
S								
Subtotal								
No Rating								
N								
G								
S								
Subtotal								
Total								

Note: R- = Adequate research
R+ = Exceptional research
S- = Adequate service
S+ = Exceptional service
T- = Adequate teaching
T+ = Exceptional teaching
N = No support data on teaching
G = General (nonspecific) support data on teaching
S = Specific support data on teaching

The other half of the teachers that were being evaluated had what might be considered a relatively higher quality and quantity of research accomplishments ("R+"). The differences, as you detected, were not dramatic. They were not intended to be. After all, the committee was deadlocked. These were the ones they couldn't distinguish among, and you were faced with an arbitration decision. If arbitration decisions were easy, being a dean would not be so challenging.

Similarly, the candidates were divided into half that would be considered just adequate in their contributions in the service area ("S-"), and the other half would be seen as moderately high in this aspect of their work ("S+"). Service was represented by two attributes: the candidate's committee work and community volunteer efforts. Patient care was not an experimental variable in this study; all candidates were described as having acceptably fulfilled their patient care responsibilities, as a way of artificially limiting the decision-making process to a more manageable number of variables. And, finally,

in the teaching domain half the candidates were described as having a strong record ("T+") and half had an adequate record ("T-").

Eighty-three different folders resulted from combining all these factors in a random fashion. The purpose of this entire exercise is to help us all get a little better handle on the process of making these very complex decisions. It will be interesting to see if there is any discernible trend to the relative value assigned by the deans of this country to the areas of research, teaching and service. Finally, it is emphasized that the fundamental rationale for this entire effort is: *the more that you are able to understand the factors which contribute to your decision making, the more control of your decisions you retain.* That is, the less you have to relinquish to the soft data that other people feed you in the name of information, and the less you are subjected to the constraints of externally imposed legislative dictates or legal rules, the more you can systematize your thinking and decision-making, the more you control it. That, we trust, would be consistent with your hopes and intentions.

It must be re-emphasized that this simulation is a minor representation of the complexities of the issues which are truly involved in the process of making promotion and tenure decisions. This was simply a compromise to save time. To limit the task so that it required only of your time demanded the telescoping of data, and a modification of the decisions to be rendered, in a way which lead to certain over-simplifications and, possibly, to some distortions. It is also emphasized that a major intent of this exercise was to stimulate reflection and discussion.*

LITERATURE REVIEW

A literature review indicated that debate exists about the correlations among the attributes with which we are concerned (research, teaching and service) and promotion decisions. The following findings describe the situation in colleges and universities only, since no published information of this nature exists regarding medical schools. Some workers indicate research activity correlates significantly with promotions decisions. Aleamoni and Yimer (1972) report a correlation between these two variables of .38 which is significant at the .01 level, although they do point out that even though individuals of higher academic rank publish more often than those of lower rank, it is unclear whether this is due to time (i.e., professors who have been active longer have published more) or whether rank is due, at least in part, to research activities. Both situations may, in fact, obtain. Hayes (1971) also reports similar findings using publication lists, receipt of research grants, and department chairpersons' estimates of both research

*In this paper, the topics and problems identified by the Deans' questions have been integrated into the general narrative, rather than presented as separate questions and answers.

ability and time spent on research. These measures all correlate positively and significantly with academic rank, though the significance may be due more to large sample sizes than to a large amount of common variance.

Teaching is supposed to correlate well with promotion because instruction is considered a major goal of higher education (Astin and Lee, 1967). However, while the Kulik and McKeachie literature review (1974) discusses the evaluation of teaching, they do not mention any studies describing the relationship of teaching (however measured) to promotions. Hayes (1971) does report that teaching is unrelated to promotion at his university without regard to whether students or department chairpersons' evaluations are considered. It is probable, though, that this is an after-the-fact phenomenon; Hayes reports conditions as they currently exist rather than looking at inputs into the promotion procedure.

Service is similarly not reported in the literature as it relates to promotion, research, or teaching.

Research and teaching activities are generally acknowledged to be independent of one another, or, if a significant correlation exists, only a small portion of the variation is shared. Aleamoni and Yimer (1972) report research productivity and ratings of teaching to be independent of one another and Voeks (1962) also reports that there is no relation between faculty recognition of research activity and student ratings. Stallings and Singhal (1970) report mixed findings between publication (as a measure of research productivity) and student ratings: at one university, the correlation is statistically significant but modest, in another, it is non-significant. More recently, Linsky and Strauss (1975) conclude that there is little relationship between teaching performance and research. Hayes (1971) looked at three measures of research activity (department chairpersons' estimates of research ability and time spent in research, and number of publications), which were all highly intercorrelated, and two measures of teaching ability (a student questionnaire result and a department chairperson estimate). Chairpersons estimates of teaching were significantly correlated with research activity measures while student questionnaire results were not.* Kulik and McKeachie (1974) summarize the literature on this general topic by suggesting that while the issue is unclear, a slight positive relationship may exist between research and teaching activities.

Astin and Lee (1967) report that chairpersons' offering anecdotal information about teaching is a major method of describing teaching for promotion purposes. They report that 85 percent of the 1110 colleges they polled used this procedure. In the same study, colleagues' opinion is reported to be a major contributor to the promotion decision in 49

*The department heads' estimates may include a halo effect because the same persons made both estimates. Further, Hayes does not report the correlations per se, but only the analysis of variance significance test results. Working backwards from these values, a correlation (η^2) of chairperson's estimate of teaching ability with chairperson's estimate of research ability is probably about .15. While statistically significant (because of the sample of over 310 cases), this relationship is far from strong.

percent of the schools they considered. Results of student evaluation questionnaires, on the other hand, were used in only 12 percent of those same schools. This contrasts to the 104 out of 117 medical schools reporting using devices collecting student opinions in their curriculum evaluation activities (AAMC, 1975). A fair number of these questionnaires are probably of a teacher evaluation character suggesting that medical schools are currently using this procedure for collecting feedback on instruction more frequently than were the colleges and universities examined by Astin and Lee.

ANALYSIS OF THE FINDINGS

Two considerations were central to the analysis used. First, each dean was assumed to differ only randomly from each other dean in the way he responded to the candidates in order to simplify the analysis reported below. Second, a linear dependency existed among the decisions made by each dean; if decisions were known about three of the candidates, the decision regarding the fourth was uniquely determined. This dependency was eliminated by randomly removing one dossier from each set of four after the deans had finished working and before the data were analyzed, producing two samples of data: a normative sample containing three of the four decisions made by each dean, and a validation sample containing the remaining quarter of the cases.

Normative sample data were used in the primary analysis reported here, and conclusions reached in that analysis were tested by making predictions about cases in the validation sample. The correctness of those predictions was a test of the accuracy of the initial analysis.

Correlation and regression techniques determined the nature and the strength of the relationships among the variables involved. The dependent (predicted) variable was the promotion decision scored 0 for no promotion, 1 for promotion. The independent (predicting) variables were described using analysis of variance terminology. Main effects included: research, service, teaching, the presence of a rating form, and the nature of the information presented on teaching. Research, service, and teaching were each described as "adequate" (scored 0) or "exceptional" (scored 1) and the presence of a rating form was noted the same way. Information on teaching was presented as a set of three variables: one variable meant that anecdotal information alone was presented to the dean, the second variable meant that anecdotal information was presented along with general student evaluation questionnaire information, and the third meant general anecdotal information along with specific student evaluation questionnaire data. The particular variable applying to each set of dossiers was indicated by a 0 or 1, indicating the variable did or did not apply, respectively. The information on teaching quality was described earlier in more detail. Interactions among the main effects (e.g., the combination of high teaching with high research) were also included as predictors. The presence or absence of such a combination was indicated by a 1 or a 0.

Pearson product-moment correlations among all levels of main effects, interactions, and the dependent variables were then computed.* The resulting matrix of correlations was used in multiple linear regression predicting the promotion decision from the levels of the main effects and interactions included in the analysis.

RESULTS AND DISCUSSION

Eighty-three deans participated in the exercise, though four did not complete the tasks requested. Thus, the 79 complete sets of responses, which can be seen as representing 68 percent of all United States medical school deans, provided the input into the analyses which follow.

The data collected were sufficiently reliable ($r_{xx} \geq .6$) to warrant confidence in the decisions made by the deans as a group, but not the decisions made by each dean individually (see Table 1). Reliability coefficients were estimated using the responses made by the random half of the deans ($n = 39$) asked to rate each candidate's research, service, teaching, and patient care efforts. Data in the normative sample was used in this analysis, and reliability was estimated using the coefficient of termination: validity was considered to be a reliability squared as a lower bound (i.e., the validity coefficient is limited by the reliability squared).*

Table 1. Reliability estimates and validity coefficients for the normative sample data. The lower bound of the reliability is estimated by the square root of the validity. N's vary because of varying rates of non-response.

<u>Attribute</u>	<u>Research</u>	<u>Teaching</u>	<u>Service</u>
Reliability	.739	.676	.637
Validity	.546	.460	.406
N	119	123	116

Validity coefficients were computed by correlating deans' decisions of "strong," "moderate," and "weak" with "exceptional" and "adequate" attributes built into the paragraphs describing research, service, teaching, and patient care in the nominations letters. The figures in Table 1 indicated that (i) deans' responses, as a group, were internally consistent enough to support further analysis of the data, and (ii) the deans' perceptions of research, service, and teaching reflected differences built into the dossiers with an acceptable degree of accuracy. A correlation of .4 or higher was required in the latter case.

The results in Tables 2-5 describing the deans' perceptions of research, service, and teaching as determined by this exercise, indicate that the deans responded in a different manner to each of these three attributes. The distribution of decisions on candidates, in

*Linear dependencies were removed by dropping one level from the set (e.g., the main effect "research" was represented by only one level, exceptional). Dropping these levels caused the matrix of correlations to cease being singular.

*Note that it is possible to have data which are very reliable, i.e., internally consistent, but display a relatively low validity; the square root of the validity is an estimate of the reliability, which is conservative.

terms of their having been regarded as "strong" versus "moderate" versus "weak" in the Research area (on the rating scale used by half the participants) indicates that the "adequate" researchers were generally split between being described as moderate and weak. This distribution of cases accounted for 90 of the 119 dossiers evaluated, and produced a clearly non-random distribution ($\chi^2 = 34.85$, $df = 2$, significant at the .01 level). It can be speculated that the deans are in greater agreement about what constitutes "strong" (i.e., "exceptional") research activity than they are about those activities which are only "adequate." They can, however, make the "exceptional" versus "adequate" distinction readily.

Table 2: Distributions of Deans' Perceptions of Exceptional and Adequate Research Activity

<u>Attribute</u>	<u>Exceptional</u>	<u>Adequate</u>	<u>Total</u>
Strong	41	7	48
Moderate	16	24	40
Weak	6	25	31
Total	63	56	119

The data in Table 3 indicate that teaching was perceived by the deans in a manner similar to research, with one exception: while they saw "exceptional" as corresponding to "strong" (on the rating scale), they perceived "adequate" as corresponding more specifically to "moderate" than they had for research. These results suggest research and teaching are perceived differently only when one considers "adequate" (rather than "exceptional") levels. The deans' responses were again non-random ($\chi^2 = 49.26$, $df = 2$, significant at the .01 level).

Table 3: Distribution of Deans' Perceptions of Exceptional and Adequate Teaching Activity

<u>Attribute</u>	<u>Exceptional</u>	<u>Adequate</u>	<u>Total</u>
Strong	48	40	58
Moderate	11	35	46
Weak	2	17	19
Total	61	62	123

The deans viewed service in a manner different from the way they perceived research and teaching, and the distribution was again non-random ($\chi^2 = 19.22$, $df = 2$, significant at the .01 level). However, "exceptional" service was split between "strong" and "moderate" while "adequate" was seen as being predominately "moderate" (89 of 116 cases were classified this way). In short, the deans responded consistently to "adequate" service while they were about evenly split in terms of how to evaluate "exceptional" service; they did not distinguish between these two levels consistently.

Table 4: Distributions of Deans' Perceptions of Exceptional and Adequate Service

<u>Attribute</u>	<u>Exceptional</u>	<u>Adequate</u>	<u>Total</u>
Strong	26	8	34
Moderate	28	35	63
Weak	3	16	19
Total	57	59	116

Patient care was also perceived in a non-random manner ($\chi^2 = 41.16$, $df = 2$, significant at $\alpha = .01$). Even though all descriptions of patient care were designed to be "adequate," they were perceived as being split between "strong" and "moderate," a pattern reminiscent of service. Only rarely were the descriptions considered indicative of weak patient care (see Table 5).

Table 5: Distributions of Deans' Perceptions of Adequate Patient Care

<u>Attribute</u>	<u>Adequate</u>
Strong	44
Moderate	51
Weak	3
Total	98

In summary, the deans showed variation in the way they evaluated research, service, teaching, and patient care. Research and teaching were similar in that dossiers designed to be "exceptional" were perceived regularly as "strong." "Adequate" teaching, however, was perceived uniformly as being "moderate" while "adequate" research was split between "moderate" and "weak." Service, however, was viewed consistently only when it was "adequate" in quality; deans split between calling it "strong" or "moderate" when it was designed to be "exceptional." Patient care was also viewed inconsistently; statements designed to describe "adequate" care were about evenly split between "strong" and "moderate;" they were rarely seen as weak performance, however.

The correlations reported here for normative sample data are based on 224 cases. The mean, standard deviation, and the correlation of each variable with the promotion decision is displayed in Table 6. Eleven of the correlations are statistically significant at the .01 level, though the value required for significance ($r = .171$) is very low due to the large sample size; statistical significance does not imply large portions of accounted for variability.

Table 6: Means, Standard Deviations, and Correlations with the Promotions Decision

<u>Variable</u>	<u>Mean</u>	<u>Std. Dev.</u>	<u>Correlation with Promotion</u>
No information on Teaching (N)	.333	.472	-.036
General information on Teaching (G)	.346	.477	.008
Specific info. on Teaching (Sp)	.321	.468	.029
Exceptional Research (R)	.509	.501	.385*

*Indicates significance at $\alpha = .01$.

Table 6 (Cont'd): Means, Standard Deviations, and Correlations with the Promotions Decision

<u>Variable</u>	<u>Mean</u>	<u>Std. Dev.</u>	<u>Correlation with Promotion</u>
Exceptional Service (Se)	.487	.501	.104
Exceptional Teaching (T)	.512	.501	.290*
N by Acceptable Research	.188	.392	-.200*
N by Exceptional Research	.158	.366	.154
N by Acceptable Service	.167	.373	-.126
N by Exceptional Service	.175	.381	.055
N by Acceptable Teaching	.188	.392	-.112
N by Exceptional Teaching	.158	.366	.060
G by Acceptable Research	.158	.366	-.128
G by Exceptional Research	.188	.392	.129
G by Acceptable Services	.154	.362	.097
G by Exceptional Service	.184	.388	-.058
G by Acceptable Teaching	.171	.377	-.205*
G by Exceptional Teaching	.171	.377	.227*
Sp by Acceptable Research	.154	.362	-.212*
Sp by Exceptional Research	.166	.373	.241*
Sp by Acceptable Service	.179	.384	-.069
Sp by Exceptional Service	.141	.348	.114
Sp by Acceptable Teaching	.136	.344	.095
Sp by Exceptional Teaching	.184	.389	.119
R by Se	.278	.449	.283*
R by T	.243	.480	.380*
Se by T	.239	.427	.212*
R by Se by T	.138	.340	.270*
Promotion Decision	.526	.500	1.000

The eleven identified correlations were logically related to one another. Two "main effects" in the model (teaching and research) correlated significantly with the promotion decision, bearing direct though modest relationships to promotions ($r=.290$ and $r=.385$, respectively).

The significant correlation of research with promotion is not surprising. Gustad pointed out that institutional and departmental goals are often externally oriented (1967); departments and schools look for the development of faculty members' national reputations, and reward their efforts in these directions (e.g., publications and grant awards) accordingly. While Gustad's comments were directed toward higher education generally, they probably do apply at medical schools. He also pointed out that research efforts are measured in terms of "yardage," that the researcher with the longest list of publications is most likely to be promoted. Such a criterion, clearly, is convenient because (i) evaluation of research quality can be assumed--if it was published in a reputable journal, someone must have already thought it was good--and (ii) it is objective--there is no dispute on the actual count of an individual's publications.

It should also be noted that the magnitude of the correlation found here is similar to at least one other reported in the literature (Aleamoni and Yimer, 1972).

The correlation of teaching with promotion suggests that contrary to some faculty member assumptions, the deans do respect teaching as an activity which can be used as a significant component of promotion decisions. Note, however, that less than 9 percent of the variance in

*Indicates significance at $\alpha = .01$

teaching is shared with the promotion decision. This is not a surprising finding, given the literature cited earlier. What is surprising is the fact that the difference between the correlations for research and teaching is non-significant at any of the usual alpha levels, even given the large numbers of cases in the sample ($T=1.22$, $df=231$, Guilford, 1965). Thus, in contrast to the studies reported earlier, the medical school deans that participated in this simulation did not value excellence in research more heavily in the promotion decision than excellence in teaching; they see the two as being about equally important, and the observed difference between them has to be attributed to random variation.

There are at least two explanations for this finding. The first is straightforward: teaching and research excellence are equally valued by medical school deans. The second is that this finding is an artifact of the methodology used to collect, analyze, and interpret the data. Thus, a review of the limitations of the methodology is in order.

First, the role of the department chairperson has been assumed to be constant for all the nominations. In the absence of additional information, the deans were forced to assume that each applicant was seen as being equally valuable by his own department chairperson. Hence, the nomination for promotion. In point of fact, this situation rarely obtains; deans are familiar with differing departmental needs resulting in varying support for applicants. Second, the deans were told nothing about the promotion committee's deliberations and why it was deadlocked on these particular candidates, and such information, if available, may well have influenced their decisions. Finally, though deans may not know each candidate personally, they do have the option of calling other faculty members to get additional information. It could be that such supplemental information (whether it bore directly on the candidate, his department, or whatever) could have influenced the promotion decision. The problem, then, may be with the information that was not available rather than with the information which was presented.

All the deans were asked to indicate their feeling about the quality and quantity of the information in the dossiers. In both cases, inadequacy was indicated at a rate of about 4 to 1 over adequacy. (A number of deans, however, reported that this promotion exercise pointed up the need for better information about the promotion procedure.)

The appropriate conclusion is that the data presented here indicate research and teaching excellence do not differ significantly as predictors of promotion *under the conditions described in this study*. This conservative approach is warranted, given the nature of this simulated activity.

The remaining nine significant correlations were all "interactions," involving research and teaching, either with each other or with other main effects. Thus, the observed significance could be due to either unique variation, contributed by each of these interactions, or to the inclusion of research and teaching as contributors to the interactions. This, in fact, was the case; when the variance in the promotion decision attributed to research and teaching was

partialled out, the remaining correlations contributed little in the prediction of promotion. The partial correlations are displayed in Table 7.

Table 7: Partial correlations among interaction predictor variables controlling for research and teaching. Abbreviations are defined in Table 6.

<u>Variable</u>	<u>Zero Order Correlation</u>	<u>Partial Correlation</u>
N by Acceptable Research	-.200	-.001
G by Acceptable Teaching	-.205	-.078
G by Exceptional Teaching	.227	.087
Sp by Acceptable Research	-.212	-.087
Sp by Exceptional Research	.241	.067
R by Se	.283	.069
R by T	.380	-.033
Se by T	.212	.024
R by Se by T	.270	-.005

It is interesting to note that neither information about teaching nor the presence of the rating form (for the deans' evaluations of research, service, teaching and patient care) bore a significant relationship to the promotion decision. In the latter case, this was interpreted as indicating that most of the deans were probably using some kind of evaluation rating system--if not the one provided, then one they created for themselves. In the preceding case, the lack of significance was taken to mean either that the deans were not accustomed to handling varying kinds of information about teaching (specifically, student questionnaire results) or that they felt the available information provided no useful information beyond what was known from the department chairperson's comments. The latter view is supported by Hayes' study (1971), where the ratings of chairpersons and students regarding teaching ability correlated at $r=.62$.

Astin and Lee (1971) point to survey results indicating that 44 percent of a sample of 1110 academic deans used "scholarly research and publications" as a basis for evaluating teaching effectiveness. Whether this is also the case in medical schools cannot be determined from the present experience. However, the presence of such a halo effect would not be surprising; department chairpersons know faculty more thoroughly than students, and it would not be surprising to find their judgment of teaching was influenced by outside information. If a halo effect did exist in this study, it would appear as a significant correlation for the interaction between research and teaching and the promotion decision (see Table 6). Such a significant correlation did exist; although, as shown in Table 7, the partial correlation was non-significant. In short, if there was any halo effect it was not an important factor.

The correlations among the predicting variables are uninteresting. Many of the correlations reflect the independencies built into the model (e.g., research is independent of teaching is independent of service) and the remaining correlations reflect built in dependencies (a negative correlation exists between specific information on teaching and general information on teaching because these two situations are mutually exclusive).

Only two variables, excellence in research and excellence in teaching, appeared in the multiple linear regression results. No other variables appeared in the regression analysis which yielded a multiple correlation of .503 (see Table 8 for a summary of results).

Table 8: Stepwise regression analysis results. The dependent variable intercept is .14964.

<u>Variable</u>	<u>Slope</u>	<u>Beta</u>	<u>Standard Error</u>	<u>F</u>	<u>DF</u>
Research Excellence	.41177	.41225	.05702	52.143	1,232
Teaching Excellence	.32487	.32519	.05703	32.445	1,231

The observed multiple correlation is large enough to be psychologically important. However it is based on the independence of research and teaching, an assumption which has been questioned in the literature: a slight positive correlation may exist between research and teaching, suggesting that for whatever reason (e.g., good teachers tend to be good researchers; a halo effect may exist), a certain amount of shared variance exists. Simple recalculation based on non-independence produced interesting findings. If the correlation is one of .20 (4 percent common variance) the multiple correlation would drop to $r = .44$ (higher than the value estimated for the Hayes data). If it were as high as .30 (9 percent common variance) the multiple correlation would drop to $r = .41$, values which do not depart appreciably from the observed multiple correlation of .503. In short, the assumption of orthogonality affected the model little.

The multiple regression analysis provided values which were used to predict promotion decisions for the cross-validation data. The predictions could be expected to be fairly accurate, however, strictly because that sample differed only randomly from the cases used in the regression analysis. A correlation between the predicted decision (based on research and teaching characteristics in the dossiers) and actual decisions was expected to be slightly less than the .503 reported earlier. The shrinkage would be due to some specific variance picked up in the initial analysis.

In fact, the correlation of predicted with actual decisions for these 79 cases was .71, a surprisingly high value which prompted a re-examination of this sample of cases. This examination revealed an unusually high number of dossiers with both exceptional teaching and exceptional research or both acceptable teaching and acceptable research. The predicted promotion in the former case was .886, in the latter case it was .150. In other words, it was the appearances of many "clear cut cases" which inflated the correlation making the results of this portion of the study equivocal.

CONCLUSIONS

Based on the literature reported, methods used, and sample described the following specific conclusions can be drawn:

1. Medical school deans, under specified conditions, perceive exceptional and adequate research, service, and teaching differently.
2. Medical school deans, under the conditions described here, see excellence in research and excellence in teaching as the primary predictors of promotion from Assistant to Associate Professor with tenure.

3. Unlike results reported elsewhere in the literature, the findings here suggest that research excellence is not valued significantly more than teaching excellence, when comparable evaluative information is available in both areas.

4. Excellence in teaching and excellence in research can be combined to produce a weighted composite which is well correlated with the promotion decision.

5. Medical school deans do not respond differently to different types of information about teaching.

6. Even a partial simulation of a complex decision situation can be engaging and can lead to a productive discussion of important, relevant issues.

Finally, we want to give some emphasis to a reasonably evident point: that there are significant reasons for improving our capacity to evaluate the instructional effectiveness of faculty members, in addition to decision-making for promotion and tenure. Most prominent among these is the overdue need for us to become more accountable to the students we serve, and to the public that supports us. We have long deferred fulfillment of this responsibility on the double grounds that our faculty members would not tolerate having their teaching evaluated (each professor's classroom has traditionally been regarded as sacrosanct), and that we did not have adequate procedures or tools available to enable accumulation of dependable evidence on instructional quality. These two problems are interrelated. People are most resistant to being evaluated when the methods of evaluation are, or are at risk of being, arbitrary, capricious, or unfair. In the domain of faculty evaluation they no longer need to be.

Attached to this paper is a bibliography which documents the extensive effort that has been devoted to studying and improving the process of faculty evaluation and development. We hope you will find these materials useful.

CITATIONS

1. AAMC, *1975-76 AAMC Directory of American Medical Education*, Washington, 1975.
2. Aleamoni, L.M., and M. Yimer, *An Investigation of the Relationship Between Colleague Rating, Student Rating, Research Productivity, and Academic Rank in Rating Instructional Effectiveness*, Office of Instructional Resources, University of Illinois, Urbana, 1972.
3. Astin, A.W. and C.B.T. Lee, "Current Practices in the Evaluation and Training of College Teachers," *Improving College Teaching*, 29: 296-311, 1967.
4. Guilford, J .P., *Fundamental Statistics in Psychology and Education*, McGraw-Hill, New York, 1965.
5. Gustad, J.W., "Policies and Practices in Faculty Evaluation," in Astin, A.W. and C.B.T. Lee (Eds.), *Improving College Teaching*, American Council on Education, Washington, pp. 194-211, 1967.
6. Hayes, J.R., "Research, Teaching, and Faculty Fate," *Science*, 172:227-30, 1971.
7. Kulik, J.A., and W.J. McKeachie, "The Evaluation of Teachers in Higher Education," in Kerlinger, F.N. (Ed.), *Review of Research in Education*, V. 3, 1975.
8. Linsky, A.S., and M.A. Straus, "Student Evaluations, Research Productivity and Eminence of College Faculty," *Journal of Higher Education*, 46:89-102, 1975.
9. Meany, J.O. and Ruetz, F.J., "A Probe into Faculty Evaluation," *Educational Record*, 53:300-07, 1972.
10. Slotnick, H.B., and R.G. Durkovic, "Dimensions of Medical Students Perceptions of Instruction," *Journal of Medical Education*, 50:662-66.
11. Stallings, W.M., and S. Singhal, "Some Observations on the Relationships between Research Productivity and Student Evaluations of Courses and Teaching," *American Sociologist*, 5:141-43, 1970.
12. Voeks, V.W., "Publications and Teaching Effectiveness: A Search for Some Relationship," *Journal of Higher Education*, 33:212-18, 1962.

ANNOTATED BIBLIOGRAPHY

1. Aspy, David N. and Roebuck, Flora N., "From Humane Ideas to Humane Technology and Back Again Many Times," *Education*, Winter, 1975.

The authors assess the effect of the teacher's personality on learning growth in the classroom. Do students who receive high levels of empathy, congruence, and positive regard learn more? Can teachers change to give more of these factors to their students? A review of ten years of systematic research suggests that the answer to both questions is "yes."

2. Centra, John A., *Strategies for Improving College Teaching*. Prepared by ERIC Clearinghouse for Higher Education, One Dupont Circle, Washington, D.C., 1972, 51 pp. (can be ordered from the Publications Department, American Association for Higher Education, One Dupont Circle, Suite 780, Washington, D.C. 20036. Price = \$2.00).

A most important short report surveying the problem of improvement of college teaching. It includes subjects such as an analysis of the teaching and learning process and the elements that go into good teaching practices. Techniques for the improvement of teaching are discussed and include the use of self-analysis, outside teams, peer analysis, and audio-video feedback, development programs, topics for workshops, and finally new technological approaches such as cable TV, computer-assisted instruction and learning kits.

3. Centra, John A., "Self-Ratings of College Teachers: A comparison with Student Ratings," *Journal of Educational Measurement*, V. 10, #4, Winter, 1973.

Teacher self-ratings were compared to ratings given them by students. Discrepancies between self-ratings (or self-descriptions) and those provided by student would underscore the need for student feedback to the instructor as well as highlight specific areas of instruction where feedback is most essential.

4. Cook, J. Marvin, and Neville, Richard F., *The Faculty as Teachers: A Perspective on Evaluation*, Report 13, Washington, D.C.: ERIC Clearinghouse on Higher Education, One Dupont Circle, 1971, 14 pp.

This paper examines the problem of measuring and evaluating teacher performance. Evaluation methods currently in use are reviewed, including the use of student questionnaires. The authors specifically consider the relative merits of measurement based on student performance (direct measurement) and measurement based on teaching activities (indirect measurement) as they relate to the evaluation of faculty.

5. Costin, F., Greenough, W.T. and Menges, R.J., "Student Ratings of College Teaching: Reliability, Validity and Usefulness," *Review of Educational Research*, 1971, 41, 511-535 pp.

This excellent article summarizes the results of research in the field of student evaluation of teaching. It indicates that students' ratings can provide reliable and valid information on the quality of teaching. The authors do emphasize that student ratings should constitute one dimension in the evaluation of teaching effectiveness and other obvious factors such as direction of research, development of new courses or improvement of materials and methods in existing courses, department colloquia, guest lecturers in other courses, all should be considered.

6. Cotsonas, Nicholas, J. and Kaiser, Henry, F., "Student Evaluation of Clinical Teaching," *Journal of Medical Education*, 38, 1963, p. 742-745.

A rating scale was devised to evaluate teachers of a clinical clerkship program whose objective was to better meet the needs of the students.

In analyzing the results, three factors were seen as important in the student-teacher relationship; the attitude factor (person oriented), the teaching factor (task oriented) and the teacher knowledge factor. Students saw the personal relationship with the teacher of greatest importance while faculty ranked teacher knowledge most important.

7. Eble, Kenneth E., *The Recognition and Evaluation of Teaching*, Washington, D.C., American Association of University Professors, 1971.

This monograph is the first report of the two-year project to improve college teaching. Subjects covered include: recognition of teaching, evaluation and improvement of teaching, student evaluation instruments and procedures, impact of student evaluation and faculty review. The appendix contains detailed accounts of course evaluation procedures at Princeton University and the University of Washington, and a case study of classroom visitation at Carnegie-Mellon University. There are examples of questionnaires used and also an extensive bibliography.

8. Hartley, Eugene L. and Hogan, Thomas P. "Some Additional Factors in Student Evaluation of Courses," *Am. Ed. Res. J.*, Vol. IX, No. 2, Spring, 1972, pp. 241-250.

The objective was to change the focus of student evaluation questionnaires from instructor's behavior and course content to the student's estimate of his self-development as a result of the course. Was it possible to identify factors relating to feelings of self-development which were different from factors identified with the traditional type of course evaluation questionnaire? A questionnaire was developed which attempted to accomplish this purpose. Though the results at this time are not definitive, the ideas developed have generated a great deal of interest and additional research.

9. Hildebrand, Milton and Wilson, Robert C., "Effective University Teaching and Its Evaluation," Center for Research Development in Higher Education, University of California, Berkeley, 1970, 39 pp.

The objective of the study was to contribute to the improvement of teaching by characterizing effective performance and providing a satisfactory basis for the evaluation of teaching.

10. Lee, C.B.T. (Ed.), *Improving College Teaching*. Washington, D.C.: American Council on Education 1967. 407 pp.

This book centers around and comments on eight papers prepared for the American Council on Education. The work is essential for those interested in evaluation, particularly Part 4 on "Teaching and Learning," and Part 5 on "The Evaluation of Teaching Performance."

11. Miller, Richard I., *Evaluating Faculty Performance*. San Francisco: Jossey-Bass. 145 pp.

12. Miller, Richard I. *Developing Programs for Faculty Evaluation*. San Francisco: Jossey-Bass, 1974, 248 pp.

If it were necessary to choose only two publications, the two Miller books would be indicated. They provide not only an overview of the field with results of research indicated, but they are a practical, how-to-do-it account of teacher evaluation efforts. Examples of questionnaires and an extensive annotated bibliography are included.

13. Pace, C. Robert (Ed.), *Evaluating Learning and Teaching*. San Francisco: Jossey-Bass, 1973. 110 pp.

A group of experts in the field explores the complexity of the teaching-learning process; its more important aspects, various ways of viewing it, and relationships and interactions about which one should be aware and alert.

Articles of particular interest include: "The New Reporters: Students Rate Instruction," by Robert Menges, pg. 49. "Common and Uncommon Models for Evaluating Teaching," by Leonard Baird, pg. 77.

14. Simpson, Michael A., *Medical Education: A Critical Approach*. Butterworths, London, 1972, 198 pp.

The author is of the opinion that medical education has failed to keep pace with findings of educational research. The author presents factual conclusions of many disciplines relevant to medical education including teaching methods and learning, examinations and assessment techniques, and the reform of medical education.

15. Simpson, R. and Seidman, J.M., *Student Evaluation of Teaching and Learning*. American Association of Colleges for Teacher Education, 1962.

An annotated bibliography plus lists of questions which could be used in formulating a teacher self-evaluation scale.

16. Sockloff, Alan L., Editor, *Proceedings: Faculty Effectiveness as Evaluated by Students*. Measurement and Research Center, 1974, Temple University, Philadelphia, Pennsylvania 19122 (can be ordered directly from the University).

This invitational conference was held in Philadelphia April 25-27, 1973. It concentrated on five broad areas of student evaluation of teaching effectiveness and twelve papers were presented. They ranged from theoretical considerations to actual examples of student evaluation forms. The authors are well-known authorities in the field, and their comments are valuable.

17. Sullivan, A.M., and Skanes, G., "Validity of Student Evaluation of Teaching and the Characteristics of Successful Instructors," *Journal of Educational Psychology*, V. 66, 584-590.

Research disclosed that students are able to provide an accurate estimate of the amount they learn from an instructor, and student evaluations may provide a valid indication of the amount which they have learned.

18. Travers, Robert M.W. (Ed.) *Second Handbook of Research on Teaching*, Chicago, Rand McNally and Co., 1973, 1400 pp.

A major work which is devoted to research in teaching with articles written by leaders in the field. Articles in Part II, Rosenshine's article, "The Use of Direct Observation to Study Teaching," and McNeil's, "The Assessment of Teacher Competence," are of particular interest.

OTHER SUGGESTED REFERENCES

1. Aleamoni, L.M. and Yimer, M., "An Investigation of the Relationship between Colleague Rating, Student Rating, Research Productivity, and Academic Rank in Rating Instructional Effectiveness." *Journal of Educational Psychology*, 1973, 64, 274-277.
2. Aleamoni, L.M., "The Usefulness of Student Evaluations in Improving College Teaching." *Proceedings by Sockloff (Ed.)*, pp. 42-58. Measurement and Research Division, University of Illinois.
3. Batista, Enrique E., "The Place of Colleague Evaluation in the Appraisal of College Teaching: A Review of the Literature." Research Memorandum No. 162, Office of Instructional Resources, University of Illinois, Urbana, Illinois.
4. Bayley, David H., "Making College Teaching a Profession." *Improving College and University Teaching*, Vol. 15, 1967, pp. 115-119.
5. Bejar, Isaac I., *A Survey of Selected Administrative Practices Supporting Student Evaluation of Instruction Programs*. Minneapolis: University of Minnesota.
6. Bergquist, William S. and Phillips, Steven R., "Components of an Effective Faculty Development Program." *Journal of Higher Education*, Vol. XLVI, No. 2, March/April, 1975, pp. 177-211.
7. Centra, John A., "College Teaching: Who Should Evaluate It?" *Findings*, Vol. 1, No. 1, Educational Testing Services, Princeton, N.J., 1974.
8. Centra, John A., "Colleagues as Raters of Classroom Instruction." *Journal of Higher Education*, Vol. XLVI, No. 1, May/June, 1975, pp. 327-337.
9. Doyle, Jr., Kenneth O. and Whiteley, Susan, "Student Ratings as Criteria for Effective Teaching." *American Educational Research Journal*, Summer 1974, Vol. 11, 3, pp. 259-274.
10. Edwards, Scott, "A Modest Proposal for the Evaluation of Teaching." *Liberal Education*, October 1974, pp. 316-326.
11. Feldhusen, John F., et al. "Designing Instruction to Achieve Higher Level Goals and Objectives." *Educational Technology*, Purdue University, October 1974.
12. Frey, Peter, "Student Ratings of Teaching: Validity of Several Rating Factors." *Science*, October 5, 1973, Vol. 18, pp. 83-85.
13. Gessner, Peter K., "Evaluation of Instruction." *Science*, May 11, 1973, pp. 566-569.
14. Ginther, John R., "Transactional Evaluation in a Medical School Setting." AERA Annual Meeting, Chicago, April, 1972.
15. Gray, Charles E., "The Teaching Model and Evaluation of Teaching Performance." *Journal of Higher Education*, 1969, Vol. 40, pp. 636-642.
16. Gromisch, Donald S., et al, "A Comparison of Student and Departmental Chairman Evaluations of Teaching Performance." *Journal of Medical Education*, Vol. 47, April 1972, pp. 281-284.
17. Group for Human Development in Higher Education. "Faculty Development in a Time of Retrenchment." *Change Magazine*, New Rochelle, N.Y., 1974.
18. Grush, J .E. and Costin, F., "The Student as Consumer of the Teaching Process," *American Educational Research Journal*, Winter, 1975, Vol. 12, No. 1, pp. 55-56.
19. Gustad, John W., "Policies and Practices in Faculty Evaluation." *Improving College Teaching by Astin and Lee (Eds.)*, American Council on Education, Washington, D.C., 1967, pp. 194-211.
20. Harvey, J. Noland, "Student Evaluation of Teaching Effectiveness." *Improving College and University Teaching*, 1970, Vol. 18, pp. 275-278.

21. Helfer, Ray A., "Peer Evaluation: Its Potential Usefulness in Medical Education." *British Journal of Medical Education*, 1972, 6, pp. 224-231.
22. Hildebrand, M. "How to Recommend Promotion for a Mediocre Teacher Without Actually Lying." *Journal of Higher Education*, 1972, 43, pp. 43-62.
23. Hoyt, D.P., "Identifying Effective Instructional Procedures." *Improving College and University Teaching*, 1973, pp. 73-76.
24. Jason, Hilliard, "A Study of the Teaching of Medicine and Surgery." *The Canadian Medical Association Journal*, 90, April 4, 1964, pp. 813-819.
25. Jason, Hilliard, "A Study of Medical Teaching Practices." *Journal of Medical Education*, Vol. 37, December 1962, pp. 1258-1284.
26. Jason, Hilliard, "Effective Medical Teachers: Born or Made," *Journal of Medical Education*, Vol. 38, 1963, pp. 46-47.
27. Jason, Hilliard, "Instructional Accountability: Is It Desirable or Feasible." *Journal of Medical Education*, Vol. 49, No. 5, May, 1974, pp. 460-461. (Editorial)
28. Jason, Hilliard, "Some High Priority Needs in Medical Education." *World Medical Journal*, No. 5, 1970.
29. Karman, Thomas A., "Faculty Evaluation." *Liberal Education*, December, 1969, pp. 539-544.
30. Kerlinger, Fred N., "Student Evaluation of University Professors." *School and Society*, October 1971, Vol. 99, pp. 353-356.
31. Kopta, Joseph A., "Evaluating Teaching Methods." *Surgery*, May 1972, Vol. 71, No. 5, pp. 893-894.
32. Kulik, J.A. and McKeachie, W.J., *The Evaluation of Teachers in Higher Education*. Ann Arbor: The University of Michigan, February 1974.
33. Mager, Robert F., *Preparing Instructional Objectives*. Belmont, California: Educational Division, Fearon Publishers, Lear Siegler, Inc., 1962.
34. Maslow, A.H. and Zimmerman, W., "College Teaching Abilities, Scholarly Activity, and Personality." *Journal of Educational Psychology*, 1950, 47, pp. 185-189.
35. McKeachie, Wilbert J., "Student Ratings of Faculty." Reprinted from *AAUP Bulletin*, Winter, 1969.
36. McKeachie, Wilbert J., *Teaching Tips*. Lexington, Mass.: D.C. Heath and Co., 6th Edition, 1969.
37. Menges, Robert J., "The New Reporters: Students Rate Instruction." Pace, Robert C. (Ed.), *Evaluating Learning and Teaching*, Jossey-Bass, Inc., Washington No. 4, Winter 1973, pp. 59-75.
38. Miller, George E., "On Training Medical Teachers." *The Canadian Medical Association Journal*, 92, April 3, 1965, pp. 708-711.
39. Miller, George and Fulop, Tamas (Ed.), "Educational Strategies for the Health Professions." *Public Health Papers*, No. 61, WHO, Geneva, 1974.
40. Musella, Donald. "Improving Teaching: Another Word." *Improving College and University Teaching*, 1968, Vol. 16, pp. 167-169.
41. Rippey, Robert M., *Student Evaluation of Professors: The Art of Crap Detection*. Department of Research in Health Education, University of Connecticut Health Center, November 14, 1974.
42. Rous, Stephen N., et al, "The Improvement of Faculty Teaching through Evaluation: A Preliminary Report." *Journal of Surgical Research*, 11, pp. 311-315.
43. Rous, Stephen N., et al, "The Improvement of Faculty Teaching through Evaluation: A Follow-up Report." *Journal of Surgical Research*, 1972, Vol. 13, pp. 262-266.
44. Rowland, Ray, "Can Teaching Be Measured Objectively?" *Improving College and University Teaching*, 1970, Vol. 18, pp. 153-157.
45. Seldin, Peter and Wakin, Edward, "How Deans Evaluate Teachers." *Change*, Vol. 6, No. 9, November, 1974, pp. 48-49.

46. Silber, John R.; VanWaes, Robert; and Kershaw, Joseph A., "The Debate on Tenure." *Change*, Vol. 6, No. 9, November, 1974, pp. 42-44.
47. Slobin, Dan Y. and Nichols, David G., "Student Rating Teaching." *Improving College and University Teaching*, 1969, Vol. 17, pp. 244-48.
48. Slotnick, Henry B., *Dimensions of Medical Students' Perceptions of Instructors*. Syracuse, New York: Upstate Medical Center, State University of New York. In press, *Journal of Medical Education*.
49. Stritter, Frank T., et al, *Effective Clinical Teaching*. 13th Annual Conference on Research in Medical Education, AAMC/Division of Educational Measurement and Research, pp. 87-92.
50. Stritter, Frank T. and Bowles, L Thompson, "The Teacher as Manager: A Strategy for Medical Education." *Journal of Medical Education*, Feb. 1972, 47, pp. 93-101.
51. Sullivan, Arthur M. and Skanes, Graham, "Validity of Student Evaluation of Teaching and the Characteristics of Successful Instructors." *Journal of Educational Psychology*, 1974, Vol. 66, 4, pp. 584-590.
52. Voeks, Virginia A., "Publications and Teaching Effectiveness." *Journal of Higher Education*, Vol. XXXIII, No. 4, April, 1962, pp. 212-218.
53. Whitely, Susan E. and Doyle, Kenneth O., "Some Contributions from Student Ratings Toward the Development of Criteria for Teaching Effectiveness." University of Minnesota Measurement Services Center, October, 1973.
54. Wilson, Robert C.; Gaff, Jerry G.; Dienst, Evelyn R.; Wood, Lynn; and Bavry, James L., *College Professors and Their Influence on Students*. New York: John Wiley and Sons, 1975.
55. Winthrop, H., "The Worth of a Colleague." *Improving College and University Teaching*, 1966, 14, pp. 262-264.

BIBLIOGRAPHIES

- Educational Resources Information Center (ERIC), One Dupont Circle, Washington, D.C. 20036.
Computer service for educational topics bibliographies.
- DeWolf, Virginia A. *Student Ratings of Instruction in Post Secondary Institutions: A Comprehensive Annotated Bibliography of Research Reported since 1968*, Vol. 1, Washington University, Seattle, Bureau of Testing, June 1974. 86 pp. (Can be ordered from ERIC Document Reproduction Service CMIC, P.O. Box 190, Arlington, Virginia 22210. Document No. EDO93 248. Microfiche \$0.75. Hard cover \$4.20 plus postage.)
- Scott, Craig and Thorne, Gaylord. *Assessing Faculty Performance: A Partially Annotated Bibliography*. Faculty Assessment Project, Teaching Research Division, Oregon State System of Higher Education, January 1974.

SOME THOUGHTS ON MEDICAL EDUCATION

Steven Muller*

Almost exactly one year ago, I was working on an article called "Higher Skilling or Higher Education" that subsequently appeared in the journal *Daedalus*. In that article, I pointed out that higher education involved acquiring knowledge as well as acquiring usable skills. I asserted that we have concentrated so much on skilling that we have begun to ignore general knowledge. We are producing graduates of higher education who are competent in certain skills, but who are often incapable of reading, writing, and even speaking their own language. They are ignorant of a second language. If they are literate in language they may be mathematically illiterate. If the skill they have studied does not involve the sciences, they may lack understanding of basic science. I argued in the article that a better balance between skill and knowledge should be restored. This led some to assume that I am opposed to the teaching of skills altogether which is emphatically not true. My reason for referring to this article is that it seems to me that a version of its argument may be applicable to medical education.

I will make no apology for the fact that I am addressing the subject of medical education without being a physician, but I do have two preliminary apologies to make. One is for the need to generalize. All medical schools are not alike and should not be alike. Today I must generalize, and in doing so I am deeply aware that there is an element of distortion involved. My second apology is for the fact that there may be an excessive Johns Hopkins flavor to what I know and to what I have to say. I can only hope that you are all sufficiently familiar with Hopkins so as to be able to discount appropriately either all or some of my remarks.

PREMEDICAL EDUCATION

Please let me now invite you to examine four major points with me. The first is that the state of premedical education in this country seems to me to be a disgrace. It may be necessary to add that I believe deans of medical schools have unavoidable responsibility with respect to premedical education. To some degree, I would claim that medical admissions is a function of the university as a whole and not exclusively the function of the medical faculty. If there is something wrong with the medical admissions process, then I think neither deans nor medical schools nor university presidents can avoid responsibility.

One aspect of medical admissions that disturbs me and that has a definite impact on premedical education is that we seem to have no effective measurement of motivation or personality. All of us throughout the country are admitting people to medical school in a highly competitive context. It is my admittedly limited observation that medical admissions committees rely primarily on criteria that focus on the ability of the applicant to master complicated technical material. They select those who are intellectually very able, and I would never deny that we all turn out some superb physicians. However, it seems to me obvious that there is no correlation between outstanding intellectual equipment on the one hand and the motivation and personality that produce an outstanding clinician on the other. I do not sense that sufficient attention is being paid in premedical education to motivation or

*President, The Johns Hopkins University, Baltimore, Maryland.

personality, and I tend to blame the medical admissions process itself. I know that applicants to medical schools go through interviews at most institutions, but I am not yet persuaded that such interviews are an adequate device for assessing motivation and personality.

I sense prevailing hypocrisy both in premedical counseling and admissions. That is a very polite way of saying that I think we do some lying. The essential lie is that all of us say that we would like well-rounded undergraduates to enter medical school but that in practice we tend to admit applicants who are intensely specialized, and primarily in basic science. It is, of course, true that it is possible for social science or humanities majors who have taken only a limited number of premedical science courses to enter medical school. However, it also seems to be true that our admissions committees place a premium on that special interest that can be demonstrated by an undergraduate majoring in biology or chemistry and by other specialized commitments such as summer work in hospitals. It is perhaps inevitable that we look for this precisely because we lack a more effective way to assess motivation. The trouble is that the way in which we really choose is well-known, so that applicants tend to make those choices which offer the best chance of being admitted, not necessarily because of spontaneous motivation but because of induced motivation. If we are sincere about using well-rounded applicants, then we ought surely to find a better way to produce them. I may be somewhat unfair to medical admissions committees in what I have said, but I can state as an unassailable fact that undergraduates totally disbelieve that medical schools want anything but highly concentrated applicants.

BASIC SCIENCES

It seems to me probable that the policy of medical admissions -- and consequently premedical education -- may be dominated too heavily by the medical basic sciences. To pick out a single example, my observation is that it is extremely difficult to be admitted to an American medical school if one has less than a grade of A in undergraduate organic chemistry. Now I personally would sincerely wish not to be treated by a physician who either had never taken organic chemistry or who had failed the course. On the other hand, I am not at all sure that someone who got a B in undergraduate organic chemistry might not make an excellent physician. How justified are we in placing so much emphasis on the highest level of academic achievement in undergraduate basic science as the primary criterion for medical school admissions? I used a strong term earlier in calling premedical education a disgrace. That choice of term results from my observation of an undergraduate student body which is hysterical to the point of neurosis in terms of the competition to enter medical school. That is surely not a uniquely personal or a uniquely Johns Hopkins observation. We all know that premedical undergraduates compete fiendishly to attain the highest possible grades in their science courses. I think we must all ask ourselves what can be done in a situation where each year there are young men and women at the age of 21 who believe their lives are blighted or ruined simply because they failed to be admitted to a medical school.

It is troubling to realize that premedical education is such a narrow, single funnel. Is there really no middle alternative between becoming a physician and abandoning all professional involvement with health care? Again, I believe all of us know of undergraduates who are sincerely motivated toward careers in health care and who may have real aptitude but who fail to gain admission to medical school. However, premedical counseling provides virtually no guidance toward alternative health care careers. Therefore, I wonder whether the profession

of medicine has some responsibility at least for cooperating in the structuring of alternative health care careers. Practicing physicians are dependent daily on others involved in hospital administration, in nursing, in clinical laboratories, and in a large variety of related disciplines. It seems to me legitimate to ask medical education to take a broader view of the whole spectrum of health care careers. The faculty in arts and sciences that deals with premedical undergraduates is not as well equipped to provide guidance for students interested in health care careers as a medical faculty would be. I do not see much evidence that medical faculties play much of a role in providing premedical career counseling at the undergraduate level. I cannot claim that I know exactly what should be done, but it is the overspecialized, overcompetitive and overly rigid nature of premedical education which I think is disgraceful.

SPECIALIZATION

The second major point I would like to make relates to specialization within the medical curriculum itself. Let me admit at once that almost every graduate or professional curriculum seems to me to have become overly specialized. When a field has existed for some time, the curriculum increasingly tends to become an accumulation of the prejudice and self-interest of generations of faculty. While this is a natural tendency, it may not be the best or most rational way to build a curriculum. It does imply that the existence of enough accumulated prejudice and self-interest do frustrate significant reform. However, if the need is great enough, reform must, nevertheless, be tried.

My complaint about the medical curriculum in general is that it has ceased to deal effectively with the society in which the physician is going to work and practice. From what I see, the medical curriculum is more crowded each year with intensively technical courses and with highly sharpened specialization. It is sometimes asserted that, of course, the undergraduate college experience should provide the broad base on which this specialized technical education must rest, but part of my first point was the argument that premedical undergraduate education is less and less broad.

Within the medical curriculum there may be more than a need for some courses or at least some time to continue to analyze and understand the larger society. At least some of the time, too much focus may be on disease rather than on the patient and on medicine as a science rather than on healing and care. I am on the same thin ice here that I skated on with respect to skilling as part of education. I do not deny the need to focus on disease nor do I claim that medicine is not a science. It is the balance between disease on the one hand and the patient on the other, and between science on the one hand and healing on the other, that concerns me. I am also struck by the degree to which preventive medicine and public health no longer seem to be an integral part of the medical school curriculum. Separate schools or public health exist and do significant work but that does not necessarily relieve medical education of the responsibility to deal with the prevention of disease. I would argue that there is excessive fragmentation of specialization within the university as a whole and within medical education in particular and that some effort must be made to restore a sense of the whole.

RADICAL REAPPRAISAL

This leads me to my third point, because I do not think the solution, if one wants to reform the medical

curriculum, is simply to scramble the blocks again. Nor do I think the solution is either to shrink the time of the curriculum or to expand it. My third point is that we may have reached a time where a truly radical reappraisal of the purpose of medical education is both possible and necessary.

We could start by an explicit recognition of what we mean when we refer to medical students as medical undergraduates. That was a phrase I initially had some difficulty getting used to, because I thought of undergraduates as people working for the baccalaureate. In the medical school world, however, I find the medical student is referred to as an undergraduate medical student. I concentrate on that because the greatest hope for the future is to realize that a medical school education -- the four years it takes to get an M.D. in most institutions -- is, in fact, nothing more than an undergraduate education in medicine that builds a foundation. It is not the concluding apex of a professional education. We should now talk openly and explicitly about continuing education as a necessary, structured ingredient of medical practice and research. This is the biggest open frontier that we have. It could take much of the pressure off the present medical school curriculum and could give us the opportunity to be more open and flexible. Why cram people with information in four years that will be dated within the next five? That approach almost ruined engineering in this country. The best we can hope to do, no matter how much we cram into people, is to teach them the techniques, the basic principles, and the individual requirements of medical basic and clinical science and then provide them with what we have not yet provided: a mandatory, structured framework for staying in a learning system for the rest of their professional lives.

Every good physician I know has, in effect, been in a lifelong learning situation; and every good medical school has offered continuing education. We are at the point where it is becoming mandatory anyway, but we have not yet drawn the full consequences for the medical school curriculum. We are still trying to push too much into students in four years on the assumption that we will not get them back. We may not, but some other medical school will.

I view medical school as a four-year period threatening to explode from overcompression. We can defuse the explosion by realizing that overcompression is no longer necessary.

I also think that we ought to look at formal, structured, career-long programs of professional continuing education as a new vehicle for underpinning ourselves financially. It will widen the horizons of our faculty. By extending medical education over a much longer period, the whole process will be much more fruitful for everyone concerned with a more humane and more liberal aspect to the four full-time years in medical school.

RESIDENCY PROGRAM

There are some other interesting questions related to this. One also ought to require, as part of the look at the medical school curriculum, a total review of the residency program. There have been arguments about whose responsibility the residents are -- the hospitals' or the medical schools'. Leaving funding aside for the moment, a very good case can be made that there should be less abdication on the part of the medical schools of responsibility for the education of residents. This, of course, would be consonant with the kind of career-long continuum I am describing, and with taking some of the pressure off the first four years.

My fourth and final thought is that the time has probably come when rigorous consideration should be given to the integration of basic science within the universities. There are at least two reasons. Separate and

essentially unrelated basic science departments in medical schools on the one hand and elsewhere in the university on the other are not -- where this situation obtains -- as cost-effective as financial exigencies demand. And probably there are academic and pedagogic virtues as well in the integration of university basic science.

The economies of such integration may not require much discussion. Careful examination should reveal what the facts are and the direction in which they point. On the point of academic virtue, at least two arguments can be made. A reform of undergraduate premedical education -- particularly a reform that includes preparation of undergraduates for a wider spectrum of health careers -- should probably include new offerings at the pre-baccalaureate level of courses in such fields as human anatomy, human biology, physiology, and pharmacology. Such courses surely could best be offered by faculty presently in medical basic science departments who now do not generally teach college undergraduates and who command both the required skills and facilities. At the post-baccalaureate level, integrated basic science departments could offer a richer environment for all of the university's basic science graduate and postdoctoral students and for the faculty itself. At least in some cases the integration may have even broader aspects, because there has been a more than occasional tendency for clinical departments in medical schools to develop their own specialized basic science research components.

These are complex, controversial ideas, too briefly stated, that interrelate closely, at least in my own mind. Let me try to conclude then with a final appeal. I would ask you, as deans of medical schools, to think of yourselves more explicitly as senior educational administrators within the university as a whole. Again at the risk of unfair generalization, it seems to me that -- too often and too much -- medical schools have been a universe apart within the total university and that this is less and less viable. I would argue this not only on academic and economic grounds but also for political reasons. The medical profession and medical education are increasingly subject to public intervention by governmental bodies. The need for funding may make some of this unavoidable, but too much of it will cost more and more in terms of academic integrity. It is my belief that medicine should be a more coherently integrated component of the whole university and that medical education will require the protective envelope of the total university in the effort to resist excessive public intervention. It is no longer feasible for university presidents to think of their medical schools as virtually autonomous enclaves; it is likewise not practical for medical school deans and faculties to keep significant distance between themselves and the rest of the university.

Please let me thank you above all for allowing me to raise -- no matter how inadequately and even misguidedly -- thoughts on medical education with you. I do believe that there is nothing more important for you to consider than that subject. You are, I know, as much preoccupied with the problems of funding, administration, and public policy as I am. Yet those problems are relevant to us only in our roles as educational administrators, and it is my conviction that our future rests more heavily on the quality of the education we offer than on any other factor. Therefore, I can personally think of nothing any of us can do that would be more constructive than to rethink the medical curriculum, and to do so in the explicit context of the entire university.

THE PRODUCT OF OUR MEDICAL SCHOOLS IN PERSPECTIVE

John W. Williamson*

Though a comprehensive, systematic and valid evaluation of medical school programs is technically infeasible at the present time, that goal can, even now, be approached by successive approximation. It is my purpose in this paper to explore two resources for decision-making in this arena that might be helpful: the first -- a developing technology to facilitate decision-making; the second -- a list of important and, in some cases, new information sources of potential value. Their description, together with illustrative data they can provide, constitutes the main focus of this paper.

DECISION-THEORY

The first resource, which has emerged from a relatively new area of research and development in the field of "group judgment technology," is now becoming formalized in schools of business and industrial management. Studies of Delphi, Nominal Group and Decision-Analysis methods are illustrative of this general development. In these days the essential data required for our most important decisions are often unavailable, and the validity of existing information and theories is frequently in question. The studies referred to are designed to assist decision-makers to arrive at more valid and reliable policies under conditions of uncertainty and accelerating change.

To help meet needs created by these conditions, investigators are studying methods of constituting optimal teams for maximum contribution to the policy and decision-making needs of institutions. These methods are based on the theory, now supported by considerable data, that in areas of informational uncertainty it is often impossible to identify, *a priori*, from a group of experts of equal competence, the one person who can provide the "best answer"; rather, the group as a whole, by a series of successive approximations, can arrive at the best response.

Other studies have focused on new methods for eliciting judgments from such formally constituted groups, so as to reduce bias and improve the effectiveness of decision-making. These methods emphasize procedures for making explicit the underlying assumptions regarding the basic value systems operative in, and the factual information which is necessary but unavailable to, the decision-making process. Systematic study is also being undertaken to identify the types of input that will increase the validity of group responses. For example, it has been shown repeatedly that the validity and reliability of a team's second estimate is increased if, after each team member has made an independent initial estimate, the tabulated group results are displayed for group discussion (nominal method) or for private individual consideration (Delphi method). Another finding, not unexpected, is that introduction of any firm measured data also significantly improves the results of group judgment.

To illustrate this first resource -- decision theory -- and to introduce the second -- health services information sources which can contribute significantly in the evaluation of medical school programs -- the Council of Deans was asked to respond to an audience survey about sources of data on specific health care issues. For whatever purpose such information is to be utilized it is important to discuss these sources, to assess the validity

*Professor, Department of Health Care Organization, the Johns Hopkins School of Hygiene and Public Health.

of the data they provide, and to consider whether those data are generalizable, and if so, to what populations.

GROUP DECISION METHODS

Before doing so, however, it is of interest to examine the results of the audience survey as a limited, specific example of one method of group data estimation. The results (Table 1) reveal that, individually, any one respondent would, on the average, agree with the criterion response 33 percent of the time. However, depending on how the group response is defined the group decision agreed with the criterion response 70 percent to 90 percent of the time, clearly a significant improvement. For seven of the 10 items the group median judgment was the criterion response; for an additional two items the group criterion response was within ± 5 percent of the group median estimate, i.e., a very close second choice.* If the three items (numbers 6, 8 and 10) where the criterion response reflected the most personal bias are eliminated, group model judgments would have correctly identified six of seven "best sources," whereas the probability is that any individual in the group would have identified only two of the seven. Similar results have been obtained repeatedly in systematic study of group data estimation.

BASIC ISSUES IN PROGRAM EVALUATION

We can now turn to the major focus of this paper, namely, a discussion of some of the most valid and reliable sources of research data, especially in the field of health services, that might contribute to the content of program evaluation and decision-making. This discussion will necessarily entail a consideration of the relation between medical care processes and health outcomes, and of the limits of validity of medical knowledge, especially in regard to knowledge about the efficacy of therapeutic interventions and the sensitivity-specificity of diagnostic tests, but information about the sensitivity-specificity of combinations of diagnostic methods applied to a variety of different patients is, indeed, limited. These considerations raise many issues regarding the limits of our present judgmental capabilities in clinical medicine. Four basic questions encompass these issues:

Who are the physicians we have produced?

What patients and health problems do they manage in their practice?

How much does the nation spend for the care they provide?

How well are they doing?

The remainder of this paper addresses these questions in terms of the more important sources of information, and reports illustrative data relevant to each.

*However, the reliance placed on two sources, as indicated by responses to items 4 and 6 did reveal a serious discrepancy with the criterion judgment. Specifically, the group median judgment identified the Commission on Professional and Hospital Activities (CPHA), Ann Arbor, Michigan, as the "best source" of data on "major health problems in the U.S. that result in hospital admission." But since data from this source are accumulated from volunteer subscribers, it is not possible to generalize their results to any national population; data from the National Center for Health Statistics would therefore be preferable. In the other case of a serious discrepancy, the group median judgment identified the Bureau of Quality Assurance as the "best source" of evidence "indicating efficacy or effectiveness of medical resources other than drugs." But since the Bureau does not have a clearing-house or bank of controlled clinical trials nor data relating to the limits of accuracy of diagnostic tests, information from that source cannot be utilized for the purpose indicated. In a full trial of group data estimation procedures, specialized information of this type, available from different experts in the group, would be supplied in discussion of the group's initial estimate; a second group estimate reflecting this new knowledge would then be obtained.

Table 1: Audience Survey

FAMILIARITY WITH HEALTH SERVICES INFORMATION SOURCES

Instructions: Below are listed ten information sources followed by ten hypothetical information needs. For each need listed below, indicate the single best information source listed by writing the letter for that source in the appropriate space. The same source may be cited several times; some sources may not be relevant to any of the items listed. Please fill in all items even though it may require a "best-guess" for some.

Sources of Health Services Information and Data

- | | | |
|--|---|---|
| <p>A. Rand Corporation
Santa Monica, California</p> <p>B. Bureau of Quality Assurance
(HEW) Rockville, Maryland</p> <p>C. National Center for Health
Statistics, (HEW)
Rockville, Maryland</p> <p>D. Joint Commission on
Accreditation of Hospitals
(JCAH) Chicago, Illinois</p> | <p>E. Office of Research and Statistics
Social Security Administration
(HEW) Washington, D.C.</p> <p>F. Center for Research and Development
American Medical Association
Chicago, Illinois</p> <p>G. Center for Analysis of Health
Practice, Harvard School of Public
Health, Boston, Massachusetts</p> | <p>H. Commission on Professional
and Hospital Activities
(CPHA) Ann Arbor, Michigan</p> <p>I. Department of Health Care
Organization, The Johns
Hopkins School of Hygiene
and Public Health
Baltimore, Maryland</p> <p>J. Center for Disease
Control (HEW) Atlanta,
Georgia</p> |
|--|---|---|

Potential Information Needs

Fill in Letter Indicating the Single Best Information Source*

Item of information you require:

Percent Correct Responses**

F	1. The 1974 total number of full time U.S. office based physicians, tabulated by medical specialty.	45%
C	2. The health problems that produce the most activity limitation in the U.S. population.	42%
C	3. Frequency distribution of diagnoses and prognostic seriousness by category of patient presenting complaints.	30%
C	4. The major health problems in the U.S. that result in hospital admissions.	16%
C	5. The leading patient complaints that result in ambulatory care in the U.S.	26%
G	6. Compilation of evidence, such as controlled clinical trials, indicating efficacy or effectiveness of medical resources other than drugs.	11%
J	7. National data regarding quality of clinical laboratory performance.	21%
I	8. Literature abstracts regarding assessment of health care processes and/or resulting outcomes.	15%
E	9. Overall U.S. national medical care expenditures by major category.	43%
A	10. Review of methods and data regarding individual and national social values, related to health and quality of life.	10%

* "Best Source" is indicated in box.

** Percent of 83 participants in Council of Deans 1975 Retreat

WHO ARE THE PHYSICIANS PRODUCED?

The Center for Research and Development of the American Medical Association is perhaps the single best source of continually updated demographic information about all U.S. physicians. These data (Figure 1) indicate that among presently active physicians, most (60 percent) are office based; that, of these physicians providing care for ambulatory patients, the surgical specialists constitute the single largest group (36 percent) and that solo practice (62 percent) is still the predominate form of organization.

A recent publication details professional characteristics of living U.S. physicians by medical school and year of graduation. The data covers the graduates of U.S., Canadian, and foreign medical schools by specialty professional activity, geographic location, age, sex, and other variables. This book is based directly on statistical data from the Center for Health Services and Development of the American Medical Association⁸.

WHAT PATIENTS AND HEALTH PROBLEMS DO PHYSICIANS MANAGE?

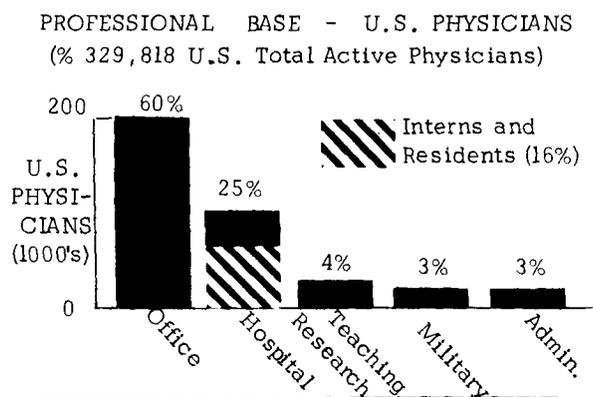
The National Center for Health Statistics (NCHS) is probably the single best source of information on these kinds of issues. The ongoing surveys conducted by NCHS utilize some of the most sophisticated methods known for data collection in this area. For example, in the Health Interview Survey (HIS), a representative sample of households around the nation is interviewed about patient symptoms, complaints, medical care utilization and the like; the Health Examination Surveys (HES) sends teams of experts out in fully equipped vans to perform complete physical examinations on a scientifically drawn sample of patients; the Hospital Discharge Survey (HDS) samples hospital discharges, rather than individual patients, to obtain data on hospital diagnoses, services, and utilization that are generalizable to the entire national hospital population.

A brief description of the procedures employed in the National Ambulatory Medical Care Survey (NAMCS), the most recent of the series, will serve to illustrate the methodology used in the ongoing surveys conducted under the auspices of the National Center. The organizing team, of which I was a member, initiated planning in 1967. Three pilot surveys were conducted to develop valid and reliable data collection methods. At present, this survey uses a multi-stage sampling procedure: the nation is divided into discrete sampling units; a random sample of these units is drawn; all the physicians currently practicing within a selected unit are listed, stratified according to specialty, and then randomly assigned to one of 52 time blocks, each block representing a specified week in the year. Each physician in the sample is asked to complete a special encounter form on a consecutive sample of patients seen during the time block to which he was assigned. Patient complaints, diagnoses, seriousness of the problem, work-up, therapy, disposition and time required for the visit are among the data being collected. Thus the final results represent what happens in the office practice of physicians throughout the entire year and these data are generalizable to the nation.

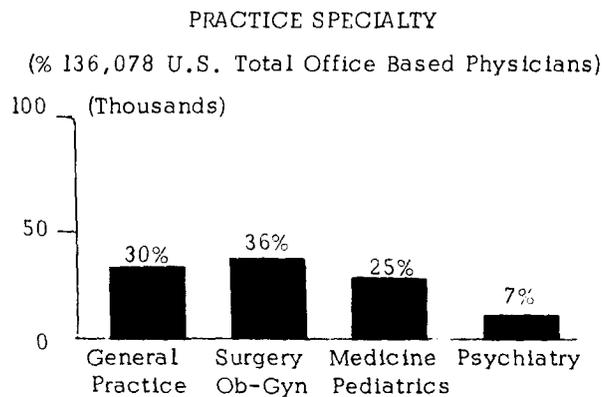
MAJOR HEALTH PROBLEMS

A "major" health problem may be defined as one leading, for example, to death, disability, short-term hospitalization, ambulatory care, and so forth. While the particular problems that surface will differ somewhat depending on the specific criterion chosen, circulatory and respiratory problems are among the top five in incidence on each of the criteria listed above. Neoplasms, accidents and digestive problems are among the other leading causes of

Figure 1: Structure of the Medical Profession

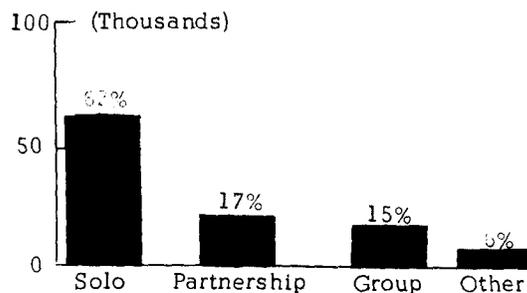


Source: Fisher Stevens 1074*



Source: NCHS-NAMCS 1974

PRACTICE ORGANIZATION
(% 136,078 U.S. Total Office Based Physicians)



Source: NCHS-NAMCS 1974

death, musculoskeletal, neurologic and mental problems among the other leading causes of disability; pregnancy, accidents and digestive problems among the other leading causes of short-stay hospitalization, respiratory and circulatory problems among the other leading reasons for office visits (Figure 2). Interestingly, the category "no illness, asymptomatic patients" encompasses the single most frequent reason people come into the physician's office. This category includes those people who come in for insurance physicals, who want a "check up", or who have some non-medical reason for coming to see the physician; but most come because the physician requested the patient return for a follow-up check after a previous illness or for a regular prenatal or well baby examination.

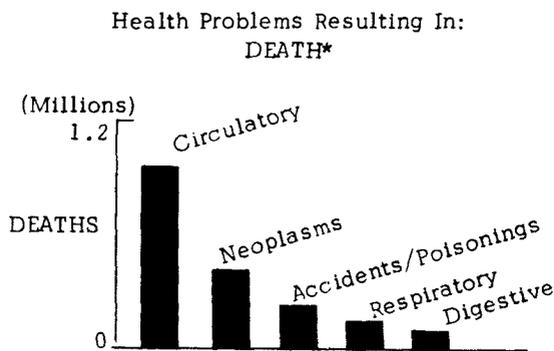
MEDICAL CARE PROVIDED**

From the point of view of sheer volume, vastly more health problems are handled in the office than in the hospital. For example, 645 million office visits were reported in the NAMCS Survey compared with 29 million hospital discharges in the Hospital Discharge Survey. Analysis of these 645 million office visits reveals that

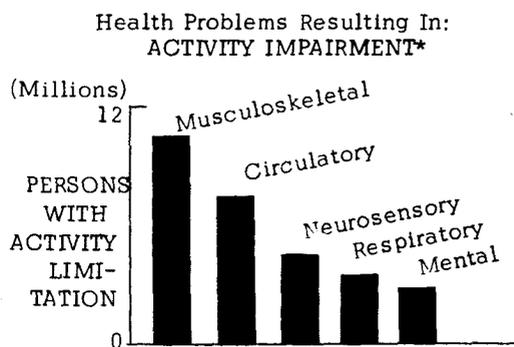
*One of nine franchising houses marketing AMA data. Not recommended for research purposes where data should be obtained directly from AMA.

**All data reported in this section are derived from the NAMCS Survey, 1974, of a scientifically selected net sample of 1,441 physicians, reporting on a random sample of approximately 30,000 patient contacts occurring from May, 1973 through April, 1974. Projecting the results nation-wide they are representative of 645 million visits to 'physicians' offices during the 12-month period.

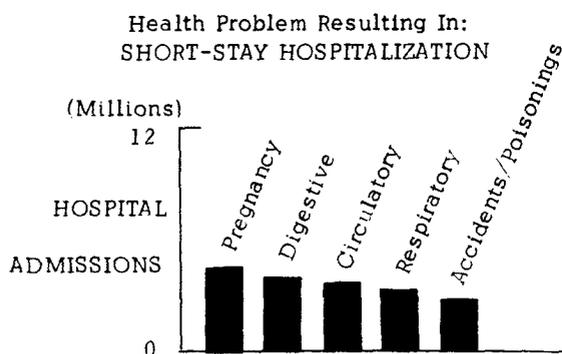
Figure 2: Major Health Problems



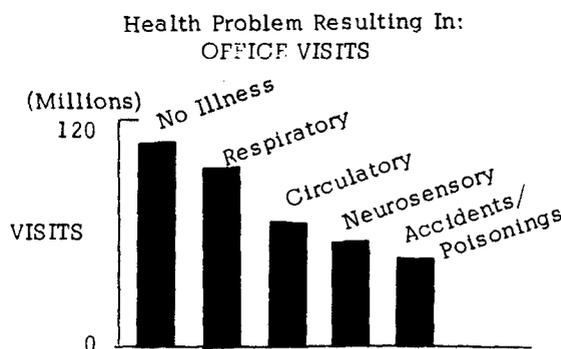
Source: NCHS-NNMS 1971



Source: NCHS-HIS 1968



Source: NCHS-HDS 1971



Source: NCHS-NAMCS 1974

partnerships are most productive in terms of the average number of patients seen per week per physician (102), with group (93) and solo practice (88) somewhat less so. It is also interesting to note that, on the average, the general practitioner sees the most people per week per physician (118) and that among the specialists, psychiatrists, as would be expected, see far fewer (44). Other specialists average between 72-99 patients per physician week. The median time (preliminary tabulation) the physician spent with each patient ranges between 10 minutes for those in general practice to 40 minutes for psychiatrists; that for other specialists (15 minutes) approaches the overall median of 12 minutes per patient.

It is important to look at the nature of the care provided in the context of the presenting complaints and the seriousness of the problem. As revealed in Table 2, about one-third of all office visits were classified "Nonsymptomatic Visits According to Patient's Purpose"; that group plus those who present with leg pain or injury, sore throat, headache or dizziness, arm pain or injury, back pain or injury, cough, "cold" or "flu", or fatigue, account for over half of all office visits. Many of the health problems seen in office practice are self limiting and the responding physicians judged that fewer than one-fifth were related to problems that would be "serious"

*Adapted from Kerr White, M.D.: Life and death and medicine. Scientific American 229(3):24, September 1973.

Document from the collections of the AAMC. Not to be reproduced without permission.

Table 2

LEADING TEN PRESENTING COMPLAINTS (Percent 645 Million Total U.S. Office Visits)		LEADING TEN DIAGNOSES (Percent 645 Million Total U.S. Office Visits)	
• Non-Symptomatic	36.5% ¹	• Well Patient	13.8%
• Leg Pain/Injury	4.0	• Acute Upper Respiratory Infections ²	6.2
• Sore Throat	3.2	• Hypertension	3.5
• Headache/Dizziness	3.1	• Neurosis	2.6
• Arm Pain/Injury	2.9	• Chronic Ischemic Heart Disease	2.4
• Back Pain/Injury	2.9	• Hayfever	1.9
• Cough	2.8	• Otitis Media	1.6
• Abdominal Pain	2.5	• Obesity	1.6
• Cold/Flu	2.4	• Exzema	1.4
• Fatigue	1.8	• Refractive Errors	1.4
		TOTAL	36.4%

Source: NAMCS-1974

¹"Non-symptomatic Visits According to Patient Purposes" NAMCS Symptom code categories 900.0, includes well patient examinations, specific tests or treatments, family planning, advice, and follow-up care.

²Aggregates "acute Upper Respiratory Infection, site unspecified," "acute pharyngitis," and "acute tonsillitis."

(16 percent) or "very serious" (3.2 percent), if untreated.* In terms of diagnoses, "well patient" (approximately 14 percent) plus the nine next most frequent diagnoses account for over one-third of all office visits (Table 2).**

It is therefore especially interesting to note that in the management provided office patients, most received little more than a cursory history and physical; laboratory tests and X-rays were ordered infrequently; drugs and injections were the predominant types of interventions. Most distressing is the fact that the physicians reported that they provided some emotional support or care (psychotherapy/therapeutic listening) in fewer than 5 percent of visits (Table 3).

The profile of the ambulatory patient in the U.S. today as it emerges from analysis of these data is depicted in Table 4: He is a patient who comes in with common complaints or with no symptoms (74 percent of visits); he was seen previously for the same problem (61 percent of visits); he contacts a specialist (59 percent of visits); he is with the physician less than 12 minutes (50 percent of visits); he receives minimal, if any, history/physical or laboratory work-up (70 percent of visits); the problem as diagnoses is slight or not serious (81 percent of visits);

*In this study, no data were obtained from the patient about his views of the seriousness of his problem.

**Data from NAMCS are organized so that it will also be possible to identify the distribution of major diagnoses made in relation to each particular complaint. For example, in the NAMCS 1972 pretest survey (N = 453 ambulatory patient contacts), the major diagnoses in visits were cough was the chief complaint were as follows: bronchitis (37 percent), acute URI (34 percent), allergy (7 percent), influenza (4 percent), pneumonia (4 percent), neoplasm (0.4 percent). These data will be invaluable in analyzing differential diagnoses of physicians.

drugs are prescribed (68 percent of visits); and he is given a return appointment (65% of the time).

HOW MUCH DOES THE NATION SPEND ON HEALTH CARE?

Malpractice insurance rates are but one manifestation of the much larger set of economic issues related to medical care. These issues are becoming increasingly serious and of ever mounting concern to both the public and the profession. However, though vitally needed for planning purposes, unit cost data related to specific health problems are all but nonexistent. Most of the data that are available relate to total health care expenditures¹⁴ (Rice, 1969), and one of the better sources of information on this subject is the Office of Research and Statistics of the Social Security Administration. Figure 3 indicates that over the past 10 years prices associated with various medical services have increased more rapidly than those of other major consumer goods and services. Secondly, over the last two decades the share of the gross national product going into the health industry has steadily increased. Expenditures for health now exceed \$100 billion out of a total GNP in excess of one trillion dollars. Thirdly, the source of funds for health expenditures is shifting from private to public funds, mainly due to increasing federal expenditures.

Table 3

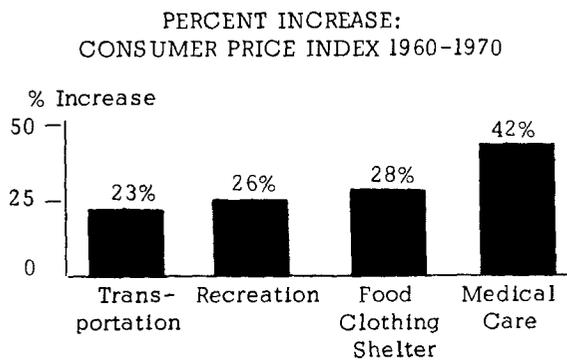
MANAGEMENT PROVIDED	
(Percent 645 Million Total U.S. Office Visits)*	
General Hx/Px	35.9%
Laboratory Tests or X-rays	26.7
Drugs/Injections	68.0
Office Surgery	8.9
Psychotherapy/Therapeutic Listening	4.3

Table 4

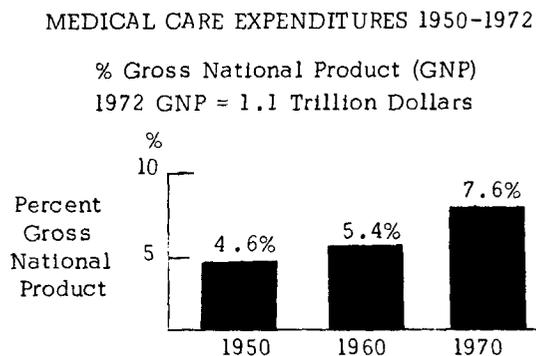
AGGREGATE AMBULATORY PATIENT	
(Percent 645 Million Total U.S. Office Visits)*	
No Symptoms/Common Complaints	74.6%
Seen Previously for Same Problem	61.5
Contacts a Specialist	59.5
With Physician <12 Minutes	50.0
Minimal (if any) Hx, Px, and Laboratory Work-up	70.0
Problem Slight or Not Serious	80.9
Drugs Injected or Prescribed	68.0
Return Appointment Given	65.0

*Source: NAMCS-1974

Figure 3: Health Care Costs

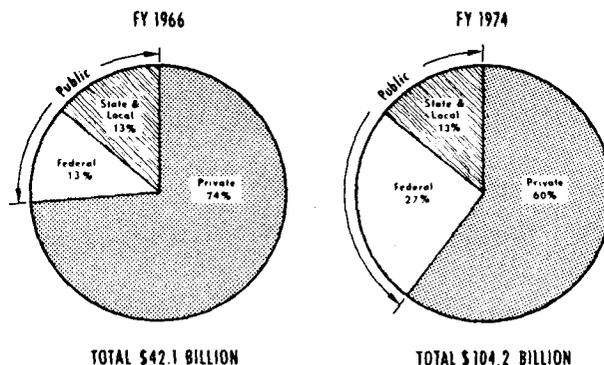


Source: Marshall and Pearson 1972



Source: Saward 1973

DISTRIBUTION OF NATIONAL HEALTH EXPENDITURES
BY SOURCE OF FUNDS, FISCAL YEARS 1966 AND 1974



Source: Social Security Administration 1975

HOW WELL ARE PHYSICIANS DOING IN PRACTICE?

While there is no possible way of obtaining data to answer this question with respect to all aspects of medical care, consideration of three basic issues will help us to approximate an answer to what is clearly the paramount question in the evaluation of the product of our medical schools. To approach this question we must consider: first, what health benefits society demands from health care providers; second, given the present state of our scientific knowledge and medical technology, what health benefits can be achieved; and third, how much of the achievable benefit is not presently being achieved.

WHAT DOES SOCIETY DEMAND?

This may be the most important issue in evaluating present medical care accomplishment. Unfortunately, it has been the focus of the least research effort. One of the earliest attempts was that by The President's Commission on

National Goals²⁰ While the report (1960) of that commission consists mainly of a series of scholarly essays, each by nationally respected specialists in the various aspects of American life, the commission, as a whole, did outline 11 major goals for the nation "at home" and four "abroad." Last among the "Goals at Home" was listed "Health and Welfare." However, the supporting essay contained no mention of any systematic attempt to identify goals or values of our citizens in this regard, and the Commission's recommendations dealt mainly with increased supply of medical personnel, grants for hospital construction, and expanded health insurance to cover medical costs.

In the 1960s little more than isolated exploratory methodologic studies on the technology of value scaling* became available. One such early exploratory study by the Rand Corporation, with major emphasis on "quality of life," indicated that Health ranked very high. Subsequent research on social values suggests that people may not be so much concerned with longevity as they are with the quality of life, and that personal independence may have higher priority than longer life-span.** If society is to authorize and, in most cases, finance research and development in health, perhaps expenditures in this area should more adequately reflect social values identified by systematic and valid measurement methods which are now becoming available.

Table 5: Scaling Human Values

<u>Factor</u>	<u>Median Weight</u>	<u>Factor</u>	<u>Median Weight</u>
Health	20.0	Freedom	8.2
Status	14.0	Security	8.2
Affluence	14.0	Novelty	7.2
Activity	12.2	Aggression	6.1
Sociality	9.8		

Source: Rand Corporation

Canadian experience provides some dramatic lessons in what can happen when social values are neglected. Canada has had National Health Insurance for several years now. Shortly after its introduction the medical profession in Quebec immediately became increasingly embroiled in the issues of mandatory physician participation in the Quebec Health Insurance Plan. This led to a strike of physician specialists, in October 1970, in which over 75 percent of all medical specialists participated. At the same time Quebec legislators were studying the report of the "Commission of Inquiry on Health and Social Services" headed by Claude Castonguay. This report recommended that a strong mandatory system of medical quality assurance be incorporated in the new Health Insurance Plan, and that the government assume responsibility for protecting consumers in their relations with all professionals, not just medical personnel.

The violent reaction of the medical profession seems to have contributed to the reaction of the legislators, who incorporated the above recommendation in Quebec Legislative Bill #250. This bill created a mechanism for direct

*For the most comprehensive coverage of this subject, see the studies of Norman Dalkey and his associates at the Rand Corporation²

**For an authoritative reference on national priorities related to health, see "A New Perspective on the Health of Canadians" by the Minister of National Health and Welfare in Canada²¹

regulation of all professions for purposes of consumer protection. The mechanism that was established partially limits, if not completely rescinds, the self-regulatory rights of some 38 professions (including engineers, lawyers, physicians, teachers, accountants, and notaries). Each profession is now required to set up a corporation of all persons licensed to practice that profession in Quebec. These corporations are to establish a system of quality assurance and to supervise the practice of the profession to "ensure the protection of the public." Finally, a governmental agency, the Quebec Professions Board has been created to establish the requirements of an acceptable quality assurance system and to monitor the performance of each profession. They are many signs that nonresponsiveness to social needs and values by the medical profession is precipitating similar reactions in this country, reactions which underscore the urgent need for systematic data on social values -- a long neglected field.

WHAT HEALTH BENEFITS CAN BE ACHIEVED?

This is another critical area where few data exist and those available are often hopelessly fragmented. What is required is systematic documentation of the efficacy of diagnostic and therapeutic medical interventions. One of the best single sources is the Center for Analysis of Health Practice at the Harvard School of Public Health which, supported by a grant from the Robert Wood Johnson Foundation, is attempting to compile research data in this area.

OBSERVER ERROR

However, in evaluating any data on the capability of present medical resources it is first necessary to determine the limits and validity of clinical information produced under conditions of usual clinical practice. Some fascinating research has been done in this field as described below.

Observer Error in History and Physical Examination Data. One of the early studies was undertaken in 1934 by the American Child Health Association²² In that study a sample of 1,000 11-year-old children was utilized to determine concordance of physician judgment regarding indications for tonsillectomy. On examination it was found that 611 of the sample had already had a T&A. The first of three independent panels of physicians examined the remaining 389 children and judged that one in three required a tonsillectomy. The remaining 116 were then examined by a third panel which judged that roughly a third required a tonsillectomy. Overall, in only 65 cases in this group of 1,000 was a T&A judged not indicated. The investigators commented that it was fortunate for those remaining few that the study had run out of physician teams to provide further independent examinations.

Observer Error in Radiologic Diagnosis. In one of the most thoroughly validated studies of observer error, replicated repeatedly both in Europe and in this country, serious lesions were missed in roughly 25 percent of chest films and from 100 random tuberculosis patients. No relation between the number of errors and the size or character of the lesion could be established. Investigators conclude that there seem to be limits to human performance in reading such films under the conditions and pressures of average practice. To avoid such errors it may be necessary to require dual readings, or to have radiologists read smaller numbers of films at a single sitting. Successful pre-tests have been concluded and a major Efficacy Study sponsored by the American College of Radiology is now underway to obtain systematic data on the validity and efficacy (value to patient) of the most common radiologic examinations.

Observer Error in ECG Interpretation. Similar results have been obtained in analogous studies of cardiologists' interpretations of electrocardiograms. In one study for example, there was major disagreement among independent readings in 20 percent of the tracings from 100 acute coronary suspects. Furthermore, these same physicians disagreed with their own previous readings in 13 percent of tracings fed back to them two weeks later.

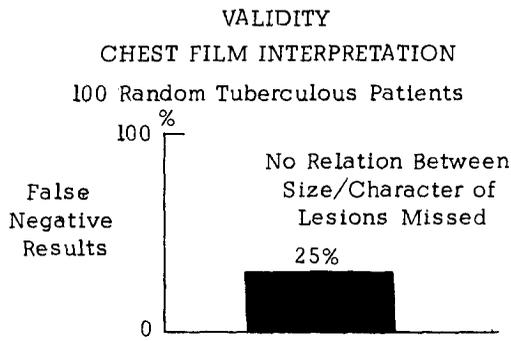
Observer Error in Medical Laboratory Reports. The Center for Disease Control has a division with responsibility for licensing and proficiency testing of all medical laboratories providing interstate service and all others who voluntarily join the proficiency testing program. Two studies are illustrative of preliminary data from this program. The first related to the validity of reports of "Pap" smears from 300 laboratories involving 700 technicians and their supervising pathologists. Each person tested was given a series of 10 prepared smears on glass slides. Among the 10 was a blatant carcinoma which was missed 30 percent of the time overall, and 37 percent of the time by the supervising pathologist. In a second study 24 laboratories were sent blood samples into which a known drug had been introduced (for example, narcotics, barbiturates, or amphetamines) with the request that these be assayed for unknown drug abuse substances. Each laboratory was sent half of a split specimen directly as part of the proficiency testing program; assay results correctly identified the substance in 96 percent of the specimens. The remaining half of each sample was sent to a physician in the community who put a patient's name on it, and sent it in for assay to the same laboratories with the same instructions. The accuracy rate in assays of these specimens dropped to 50 percent. In related studies it has been found that most error rates increase significantly in many laboratories after 5 p.m., when the regular staff leaves, the auto-analyzers are turned off, and older assay methods are employed. Even more significant, proficiency of professional performance reveals no relation to formal certification status. Board certified pathologists, for example, apparently do not do any better than those who are not certified. These findings raise serious questions about the validity of laboratory reports under usual conditions, and suggest that intensive investigation of the whole subject may be required.

Validity of Judgments of Therapeutic Efficacy. Analysis of mortality trends suggests that there are equally serious problems in judging the efficacy of therapeutic interventions. Changes in age-specific mortality rates (Figure 5) indicate that over the past half century dramatic improvement has been achieved only in the first year of life and in the elderly; for the age groups in between, far less improvement has occurred despite intervening developments of immunization programs and "miracle" drugs. For example, the steady and dramatic decline in deaths from tuberculosis and typhoid is more likely due to improved standards of living and public health measures than to immunization or to modern miracle drugs. The main decrease in mortality from all forms of pneumonia occurred before 1940 and thus before the advent of modern antibiotics. Finally, it is interesting to note that with the increasing proportion of elderly patients in the population there is increased mortality due to coronary heart disease, but no increase in mortality due to cerebral vascular disease. These are puzzling findings.

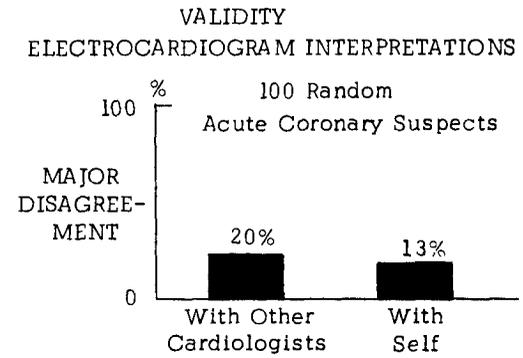
Turning to specific studies of drug efficacy compiled by the Food and Drug Administration (FDA), the most recent tabulation reveals that of the approximately 3,400 products issued between 1938 and 1962,* only approximately 12 percent are effective with respect to all claims and less than half with respect to even one claim. Of the

*Before 1938 most products were approved under a grandfather clause. After 1962 the companies have had to provide evidence that new products are both safe and efficacious. It is this latter group of products that have been the subject of intensive study by national teams of experts under FDA sponsorship in collaboration with the National Research Council of the National Academy of Science.

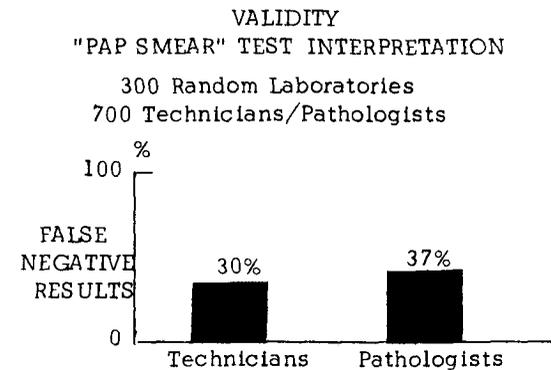
Figure 4: Observer Error in Laboratory Investigation



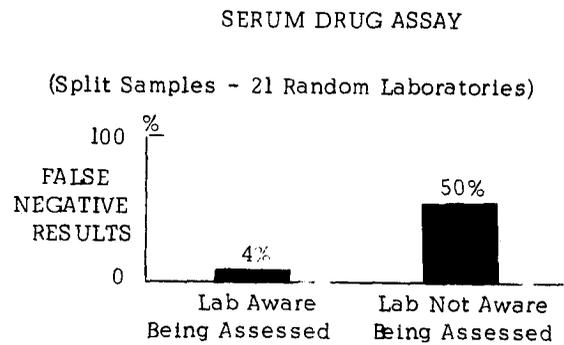
Source: Garland 1949



Source: Davies 1958



Source: CDC - LaMotte 1974



Source: CDC-LaMotte 1974

remainder, more recent, detailed study has shown that approximately four out of five are ineffective and are to be taken off the market. There is little doubt that a very substantial proportion of prescription drugs are being used on the basis of questionable evidence regarding efficacy. Even more serious is the problem of toxicity related to these same products.

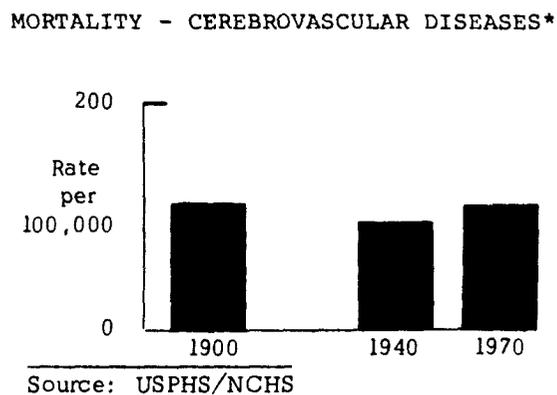
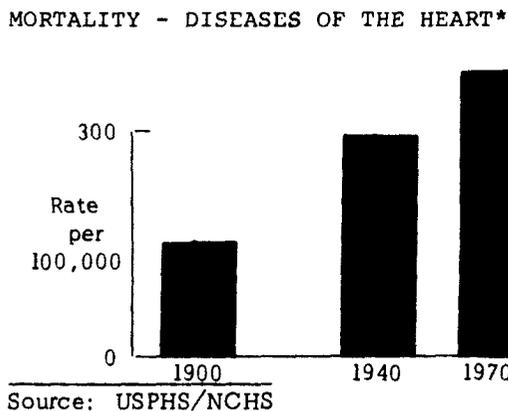
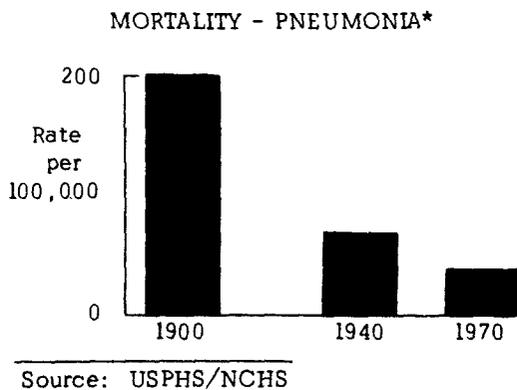
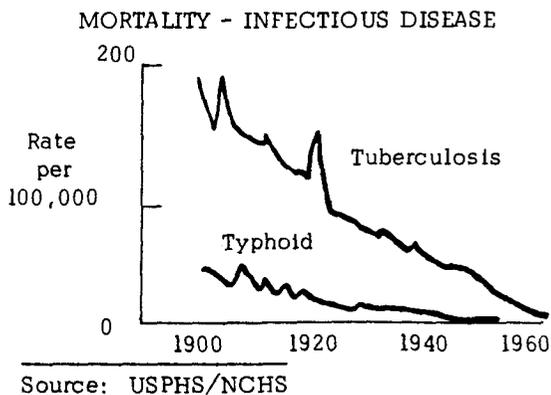
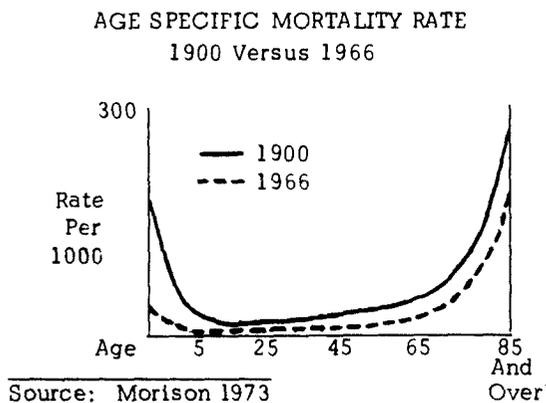
WHAT ACHIEVABLE BENEFIT IS NOT BEING ACHIEVED?

One of the better sources of information on the third major issue related to health services evaluation is to be found in our own work in the Department of Health Care Organization at the Johns Hopkins School of Hygiene and Public Health. We have been working for several years on developing a comprehensive annotated bibliography on the subject of quality assurance. The work encompasses three major divisions: assessing the effectiveness of medical care, assessing efficiency, and effecting improvement of care. Abstracts have now been accumulated for 4,000 articles covering relevant literature from 1900 to 1974. Each abstract has been read and rechecked multiple times to complete the coding according to various content and analytic descriptors.

What is most unique about this bibliography is the development of an analytic index encompassing over 250 descriptors related to the content of the abstracts and over 50 descriptors analyzing the relevance and scientific

merit of the articles for quality assurance applications. For example, for any one article, one code will indicate that patients in the middle socio-economic class were the subject of study; another code might indicate that surgeons provided the care received; another code will indicate a prepaid method of financing care; another might indicate that an outcome study was completed; yet another might indicate that unnecessary surgery was identified. Analytic codes might indicate that the article was highly relevant for assessing effectiveness of care; another analytic code might indicate that the article is also relevant for assessing efficiency of care; a third analytic code might indicate that findings led to an attempt to achieve improvement, that the attempt was evaluated, and that improvement was achieved. Finally, each article has also been coded according to whether it is reporting primary research

Figure 5: Mortality Rates



*Adapted from John H. Dingle, M.D.: The ills of man. *Scientific American* 229(3):82, September 1973.

data, secondary data, or is an editorial, among other such descriptors. By selecting a specific combination of codes for the particular type of article the reader requires, the probability of identifying irrelevant literature should be reduced. This method of coding and retrieval is of particular significance in view of the fact that using conventional search systems, four out of five articles containing important information on medical care quality assurance would probably not be recognized as relevant by the average reader on the basis of information in the title. For example, one of the best validation studies of dental prognostic judgment (essential for establishing outcome standards in assessing dental care) was labeled "Preventive Dentistry." Despite interminable delays, this bibliography should be available in late 1975, and should prove valuable in locating quality assurance literature.

Studies currently available in the area of health care assessment employ two different methodological approaches. That presently stressed by PSRO, entails a focus on "process" data; the other illustrated by our own work, focuses on "outcome" assessment.

PROCESS ASSESSMENT

The carefully designed studies conducted in Hawaii by Beverly Payne and his colleagues represent one of the better examples of process assessment, in that the sampling of both hospital and office practice was designed to permit generalizing to practice patterns in the state as a whole (see Figure 6). For example, he found that deficiencies in hospital care provided in pregnancy -- the single most frequent reason for hospital admission -- were twice as common in abortions as in normal deliveries. Analyses of the charts of 27 randomly selected patients receiving abortions revealed that examination for fetal age and viability was omitted in roughly half the cases and fetal membrane examination was omitted in almost 90 percent. In light of the recent manslaughter conviction of a Boston physician who presumably failed to check fetal viability in performing an abortion, these findings take on added significance.

In this analogous study of office practice in Hawaii, Payne found equally common deficiencies in the care of the well patient -- the single most frequent type of patient seen in the office. For example, over 80 percent of records made no mention of medications taken and even higher percentages had no mention of certain crucial findings elicited in the physical (see Figure 6). Overall performance assessment indicated a very high level of deficiency in all areas studied.

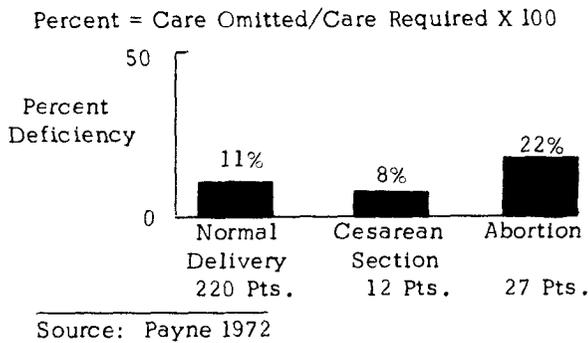
Factors relating to quality of care as measured by this type of process assessment are also interesting: for the examples given above, Payne found hospital care in group practice (Kaiser in Hawaii) significantly better; but he found no relation between quality of care and size of the hospital or board certification, area of specialization, or years in practice of the physician. In this context it is interesting that for 20 of the 23 health problem areas studied, board certification was unrelated to quality of performance; in the other three areas significant difference in favor of the noncertified physicians was found. In the case of the well adult, Payne's data indicate that internists and physicians in group practice (Kaiser) performed best. However, here again quality of care was unrelated to years in practice or board certification. These sample results reported from Payne's study in Hawaii seem to be supported by findings from similar studies elsewhere.

OUTCOME ASSESSMENT

Systematic study of the outcomes of medical care is in its infancy; development of methodology is still

Figure 6: Deficiencies in Care Provided

HOSPITAL CARE: PREGNANCY - HAWAII



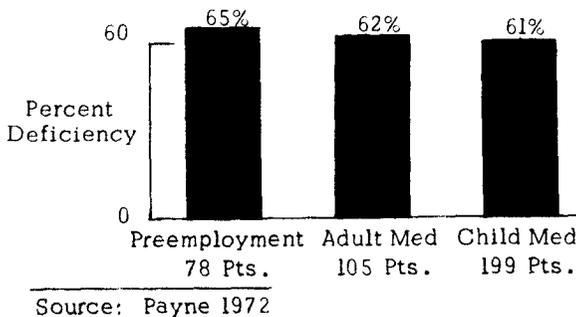
REQUIRED CARE OMITTED: ABORTION

(N = 27 Patients)

	No. Times Omitted
Hematocrit	0
Vaginal Bleeding History	1
Fetal Age/Viability Examination	14
Blood Type (Mother)	16
Fetal Membrane Examination	24

Source: Payne-1972

AMBULATORY CARE: WELL PATIENT - HAWAII



REQUIRED CARE OMITTED: WELL ADULT

(N = 105 Patients)

History and Physical	% Cases Omitted
Chief Complaint	10
Cardiorespiratory	51
Allergy	58
Medication	81
Blood Pressure	5
Chest and Lungs	88
Rectal	93
Breast	94
Fundus	98

Source: Payne-1972

underway. In our own work we have focused on two generic classes of outcomes: diagnostic outcomes and therapeutic or health outcomes. In both, the study team first obtains expert judgment on what constitutes minimum acceptable success rates or maximum acceptable failure rates. Actual outcomes for a specified practice group are compared with the acceptable standard.

Assessment of Diagnostic Outcomes. Results from three specific studies directly related to our research activity will illustrate the procedure (Table 6). The first, in which the study team was headed by Dr. Julius Krevans at Baltimore City Hospital, found that actual diagnosis of heart failure in this Johns Hopkins affiliated hospital fell well within the standards (not more than 5 percent to 8 percent missed) set by the study team. In

Table 6: Assessment of Diagnostic Outcomes

HEART FAILURE MISSED			DEPRESSION MISSED		
	Maximum Acceptable %	Findings %		Maximum Acceptable %	Findings %
University Affiliated Hospital N = 75 Patients	5	3			
University Affiliated Emergency Room N = 100 Patients	8	5	University HMO N = 274 Patients	27	84

Source: Williamson 1968

Source: Schroeder 1975

URINARY TRACT INFECTION MISSED

	Maximum Acceptable %	Findings %
Univ. Medical Clinic N = 18 Patients	15	56
Community Hospital N = 265 Patients	15	71
HMO Clinic N = 126 Patients	5	83

Source: Williamson 1974

the second, however, a study of depression conducted by Dr. Steven Schroeder at George Washington University, measured findings fell far short of acceptable standards. In a third study, or urinary tract infection conducted by Joseph Gonnella at the University of Illinois Hospital, a complete history, physical examination, urinalysis, and urine cultures were obtained by the study team on 133 consecutive new clinic admissions before they were admitted to the clinic system. This work-up was independent of any work-up or care subsequently given in the clinic and the results of the independent work-up were *not* made available to the regular clinic staff. Once admitted to the system the patient was cared for in the standard manner by regular clinic staff. This care was checked by chart review three months later to determine the quality of management provided by the faculty and resident staff of the Department of Medicine. Among the study group of 133 patients were 108 that had one or more major indications (for example, a history of stones, recent pregnancy, or previously treated urinary tract infections) for screening for current urinary tract infection. In only 68 of these 108 cases were the cues identified by the clinic staff. In only 31 of the 68 cases where such a cue was recorded was a repeat urinalysis or a urine culture ordered. Of the 31 on whom this screen was done, 18 patients were identified by the study team as having a current urinary tract infection; however, only eight (44 percent) were so identified by the regular staff of this university outpatient clinic. Since both the study team and the regular staff utilized the same laboratory, no plausible external explanation comes to mind. Indeed, replication of this study in other settings has yielded even higher

percentages of missed diagnoses.

Assessment of Health Outcomes. In order to evaluate health, it is necessary to utilize some kind of health (or health impairment) scale. Table 7 below contains an illustrative six level ordinal scale developed for this purpose. By applying this scale to responses from a group of patients on a relatively simple questionnaire, it is possible to classify each patient in the category that reflects the maximum level of his current impairment. The frequency distribution of these health levels thus represents the overall health impairment for the group as a whole. This measured distribution can then be compared to a health outcome standard expressed in terms of a theoretic distribution of 100 patients whose collective impairment represents the maximum acceptable for a given "patient group-health problem-provider" combination.

Table 7: Health Impairment Levels ¹⁸

HEALTH IMPAIRMENT LEVELS
DEAD
DEPENDENT - For Self Care Activities
RESTRICTED - From Major Life Activity
SYMPTOMATIC - At Major Life Activity
ASYMPTOMATIC - With Detectable Impairment
NO IMPAIRMENT

Table 8 reports the results of studies of four different such combinations; where the problem involves peptic ulcer, pneumonia, diabetes or hypertension, and the care is provided in widely varying settings. In all of these combinations actual health outcomes are significantly below the pre-established standard. There are two possible explanations of such results: one is that the prognostic judgment reflected in the outcome standard was invalid; this explanation would be confirmed if no subsequent effort could effect health improvement. The alternative is that the outcome standard is valid and there are correctable deficiencies in the care such patients received.

For example, in the study of children convalescing from pneumonia, it was found that residual coughing was the major symptom producing the level of impairment measured and little or no significant pathology was found on subsequent investigation; acceptable standards were adjusted downward for future assessment applications. On the other hand, in the studies of insulin dependent diabetics and hypertensive patients, it was apparent that in both cases the disease was not well controlled and that the care provided could be improved. In particular, a one year follow-up of the 87 hypertension patients consecutively sampled from the Baltimore City Hospital emergency room, suggested that three out of four deaths in this group of actively employed blue collar workers (average age - 49), were probably preventable. Subsequent investigation revealed that most were cardio-vascular deaths directly related to hypertension; that many of these patients had often failed to take their medication or even to have their prescriptions filled and/or had failed to return for any type of follow-up care.

Table 8: Health Outcomes for Various "Patient Group - Problem Area - Provider" Combinations* 18

<u>PEPTIC ULCER</u>			<u>PNEUMONIA</u>		
74 Patients Urban Group Clinic (Fee-for-Service) Five Year Follow-up			100 Children Suburban HMO (Prepaid) Two Month Follow-up		
	<u>Maximum Acceptable Percent</u>	<u>Actual Findings Percent</u>		<u>Maximum Acceptable Percent</u>	<u>Actual Findings Percent</u>
DEAD	0	5	DEAD	0	0
DEPENDENT	0	3	DEPENDENT	0	1
RESTRICTED	5	3	RESTRICTED	0	0
SYMPTOMATIC	11	23	SYMPTOMATIC	4	19
ASYMPTOMATIC	84	66	ASYMPTOMATIC	96	80
	P = .005			P = .005	

<u>DIABETES</u>			<u>HYPERTENSION</u>		
121 Insulin Dependent Patients Suburban Group Clinic (Fee-for-Service) One Year Follow-up			87 Patients Baltimore City Hospital One Year Follow-up		
	<u>Maximum Acceptable Percent</u>	<u>Actual Findings Percent</u>		<u>Maximum Acceptable Percent</u>	<u>Actual Findings Percent</u>
DEAD	2	9	DEAD	10	21
DEPENDENT	5	3	DEPENDENT	9	1
RESTRICTED	9	8	RESTRICTED	28	26
SYMPTOMATIC	26	35	SYMPTOMATIC	26	44
ASYMPTOMATIC	58	45	ASYMPTOMATIC	27	8
	P = .05			P = .001	

*Source: Williamson-1974

HEALTH OUTCOME ASSESSMENT AS A MEASURE
OF EDUCATIONAL EFFECTIVENESS

The final set of data to be presented are derived from a study in our present "Health Accounting Project." This study illustrates the general design being applied to improve health outcomes of medical care. In this specific instance related to hypertension, the approach built upon a thesis study at Johns Hopkins by Dr. Thomas Inui, he demonstrated that it is possible to educate physicians to instruct patients in a manner that will achieve increased compliance and thereby improve the health outcome, blood pressure control, in this instance. In our project, a quality assurance outcome study in a Los Angeles area prepaid clinic, the clinic study team set the standard that blood pressure control should be achieved in 95 percent of their cases.

Initial measurement of 248 patients indicated that 36 percent were out of control. To identify more specific

determinants of this unacceptable outcome two questionnaires were developed, based to a large extent, on items from Dr. Inui's study. One questionnaire was directed to the physicians to see how well they were managing these patients and the other directed to the patients to see how well they understood and were coping with this health problem, were administered. Results are shown in Table 9. The most dramatic finding in the patient responses was that 137 of the 148 respondents were unaware that the first symptom of a hypertension patient might experience could be death or stroke. Their behavior reflected their lack of understanding; approximately 80 of the 148 were not obtaining adequate therapy. The physicians' responses indicated that most of them had adequate medical textbook knowledge about hypertension, but very little knowledge of their own patients' control rates or compliance rates; when asked to list ideal care for this health problem many omitted patient education.

At this point, specific educational prescriptions were written for both patients and physicians in the experimental group. For example, the physician was provided with relevant information about each item he had missed on the questionnaire, including, when indicated, a reprint of the findings from the VA hypertension study and information about the compliance rates of his own patients. Each patient was similarly provided with education material in simple language about each question he had missed. Following the educational intervention, significant improvement was achieved in the health outcome criterion for patients involved in the study (Figure 7). Furthermore, the study revealed that significantly less improvement occurred in the group of patients who refused to obtain education from an evaluation assistant, termed a health accountant. Overall, this study illustrates the outcome approach to quality assurance and the manner in which it can be used to focus on those interventions most likely to be associated with improved patient health outcomes.

IMPLICATIONS FOR THE FUTURE

The data presented in this paper were not intended to serve as a basis for generalizing about the quality of medical care in this country. Rather, this paper has been designed to focus attention on important sources of data and kinds of information now becoming increasingly available, that may eventually be of assistance in answering the three major questions raised initially. Hopefully, also, much of what is now available has more immediate application to present decisional needs in educational management. Certainly, it has direct implications for both program evaluation and curriculum development.

Presently available data force consideration of such crucial issues as the validity and limits of specific clinical interventions commonly used: What evidence do we have that present medical care does make any difference to the health of the patient? How many times would the patient have recovered anyway? How many times was the patient made worse because of treatment provided? These are the types of questions and issues that have to be raised. The education of tomorrow's physician must address these issues of such central relevance to clinical practice. Even more important, the medical profession must gain a more accurate awareness of social needs in regard to both health services and medical research expenditures. Improving the quality of life, managing resource maldistribution, increasing the accessibility of and satisfaction with care received, may have far more reaching implications for the future practice of medicine than much present molecular biological knowledge that is dutifully memorized and quickly forgotten.

Finally, this paper has specifically tried to focus on two major resources that may be of special value to

Table 9: Responses to Hypertension Questionnaire

PATIENT KNOWLEDGE

TABLE III. HYPERTENSION QUESTIONNAIRE OF PATIENTS ON KNOWLEDGE ITEMS

	<i>Patients*</i>
Unaware that they had hypertension	23
Unaware that hypertension is a serious risk to health	14
Unaware of the form or dosage of the drug they were receiving to control hypertension	30
Unaware of the possible toxic effects of the drug they were receiving to control hypertension	120
Unaware that the danger of hypertension is not related to symptoms	137
Total with inadequate information†	140

*N = 148 (extrapolated)

†Counting each patient only once

PATIENT BEHAVIOR

TABLE IV. HYPERTENSION QUESTIONNAIRE OF PATIENTS ON BEHAVIOR ITEMS

	<i>Patients*</i>
Not taking any medication for hypertension	39
Taking medication for hypertension sporadically	25
Taking medication for hypertension only if symptoms occur	16
Total obtaining inadequate medication†	80

*N = 148 (extrapolated)

†Counting each patient only once

PHYSICIAN KNOWLEDGE

TABLE I. HYPERTENSION QUESTIONNAIRE OF PHYSICIANS ON KNOWLEDGE ITEMS

	<i>Physicians*</i>
Inadequate drug information	0
Unaware danger of hypertension not related to symptoms	4
Overestimated national control of hypertension and compliance	8
Overestimated control of hypertension or compliance by his own patients	10
Total with inadequate information†	11

*N = 14

†Counting each physician only once.

PHYSICIAN VALUES AND ATTITUDES

TABLE II. HYPERTENSION QUESTIONNAIRE OF PHYSICIANS ON VALUE AND ATTITUDE ITEMS

	<i>Physicians*</i>
Omitted the education of patients among aspects of ideal care for hypertension	11
Questioned the criteria for hypertension determined in a study by the Veterans Administration**	5
Rejected clinical diagnostic standards for evaluating the outcome of treatment	1
Rejected the blood-pressure outcome project or the role of the health accountant‡	1
Total who were possibly deficient	11

*N = 14

educational administrators. The first relates to the developing research on the technology of decision-making, assessing the validity and reliability of both traditional and innovative decisional methods in this era of rapid sociologic and scientific change and under conditions of informational and theoretical uncertainty. The second resource relates to important presently available sources of health services information that provide the most valid and meaningful measured data related to clinical and educational decisional needs. Unfortunately most administrative, as well as most clinical, decisions necessarily must continue to be based on assumptions about casual relations and about the validity of contemporary theories and data. Making these assumptions explicit by use of more advanced group estimation methods, and building on the most solid facts available should help in arriving at decisions and formulating policies that will inevitably have increasing impact on our capacity to meet the health needs of our society.

Figure 7: Uncontrolled Blood Pressure Rates Before and After Educational Intervention*

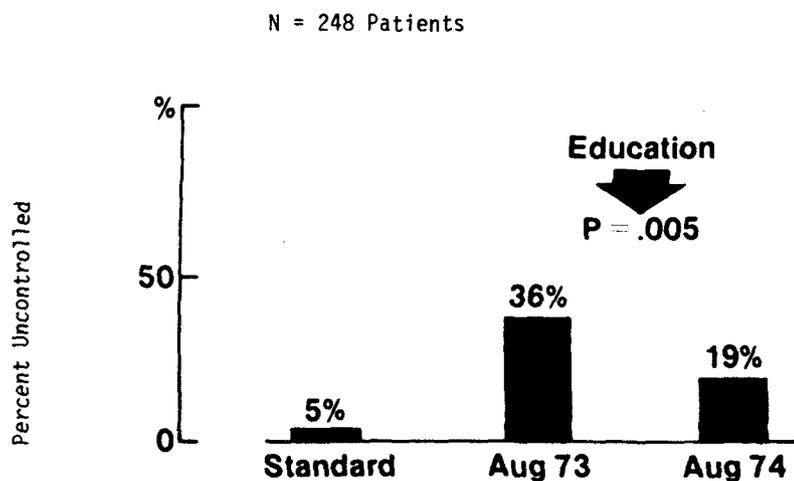


Fig. 6. Uncontrolled blood pressure found in a one-year follow-up of 248 patients participating in the study by HMO International.

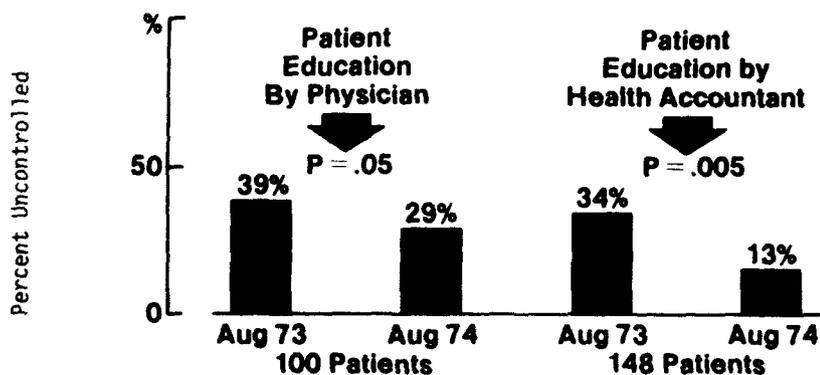


Fig. 7. Uncontrolled blood pressure found in a one-year follow-up recycle of 248 patients participating in the study by HMO International.

*Source: HMD International Study

SUMMARY OF INFORMATION SOURCES

I. WHO are the physicians produced?

Center for Research and Development, American Medical Association, Chicago, Illinois

II. WHAT patient and health problems do they manage in practice?

National Center for Health Statistics, Health Resources Administration - HEW, Rockville, Maryland 20852

- * Health Interview Survey (HIS)
- * Health Examination Survey (HES)
- * Hospital Discharge Survey (HDS)
- * Natality and Mortality Survey (NMS)
- * National Ambulatory Medical Care Survey (NAMCS)

III. HOW MUCH does the nation spend for the care they provide?

Office of Research and Statistics, Social Security Administration - HEW, Washington, D.C. 20201

IV. HOW WELL are they doing?

A. What does society demand?

Rand Corporation, Santa Monica, California 90406

B. What benefit is achievable?

Harvard School of Public Health, Center for Analysis of Health Practice, Boston, Massachusetts 02115

C. What achievable benefit is NOT being achieved?

The Johns Hopkins School of Hygiene and Public Health, Department of Health Care Organization, Baltimore, Maryland 21205

REFERENCES

1. American Medical Association Center for Health Services Research and Development: *U.S. Distribution of Physicians by Patient Care and Other Professional Activities*. Clifton, N.J., Fisher-Stevens, Inc., April-May 1975.
2. Dalkey, N.C., Rourke, D.L., Lewis, R., et al: *Studies in the Quality of Life: Delphi and Decision-Making*. Lexington, D.C. Heath and Co., 1972.
3. Davies, L.G.: Observer variation in reports on electrocardiograms. *British Heart Journal*, 20:153-161, April 1958.
4. Dingle, J.H.: The ills of man. *Scientific American*, 229(3):77-84, September 1973.
5. Garland, L.H.: On the scientific evaluation of diagnostic procedures. *Radiology*, 52:309-328, March 1949.
6. LaMotte, L.: personal communication, 1974.
7. Marshall, C.L., Pearson, D.: *Dynamics of Health and Disease*. Englewood Cliffs, N.J., Prentice-Hall, Inc., 1972.
8. Martin, B.C.: *Medical School Alumni*. Rockville, Md., Aspen Systems Corporation, 1975.
9. Morison, R.S.: Dying. *Scientific American*, 229(3):55-62, September 1973.
10. National Center for Health Statistics: Inpatient Utilization of Short Stay Hospitals by Diagnosis - United States - 1971. *Vital Health Statistics*. 13(16), July 1974.
11. National Center for Health Statistics: Limitation of Activity Due to Chronic Conditions - United States - 1969 and 1970. *Vital Health Statistics*. 10(80), April 1973.
12. National Center for Health Statistics: Mortality Trends for Leading Causes of Death - United States - 1950-69. *Vital Health Statistics*, 20(16):59, March 1974.

13. National Center for Health Statistics: Unpublished data from the National Ambulatory Medical Care Survey, May 1973-April 1974.
14. Office of Research and Statistics: *The Size and Shape of the Medical Care Dollar - Chart Book/1975*. U.S. Department of Health, Education, and Welfare, Social Security Administration, 1975.
15. Payne, B.C., Lyons, T.F.: *Method of Evaluating and Improving Personal Medical Care Quality. Vol. I. Episode of Illness Study*. Ann Arbor, University of Michigan School of Medicine, February 1972.
16. Payne, B.C., Lyons, T.F.: *Method of Evaluating and Improving Personal Medical Care Quality. Vol. II. Office Care Study*. Ann Arbor, University of Michigan School of Medicine, February 1972.
17. Schroeder, S.: The feasibility of an outcome approach to quality assurance - A report from one HMO, to be published in *Medical Care*, January 1976.
18. Williamson, J.W.: *Medical Care Outcomes: Assessment and Improvement*. To be published.
19. Williamson, J.W., Aronovitch, S., Simonson, L., et al: Health Accounting: An outcome-based system of quality assurance: Illustrative application to hypertension. *Bulletin of the New York Academy of Medicine*. 51(6):727-738, June 1975.
20. *President's Commission on National Goals: Goals for Americans: The Report of the President's Commission on National Goals*, New York, The American Assembly, Columbia University, 1960.
21. Lalonde, M.: *A New Perspective on the Health of Canadians: A Working Document*. Ottawa, Government of Canada, Minister of National Health and Welfare, 1974.
22. American Child Health Association: *Physical Defects: A Pathway to Correction*. New York, American Child Health Association, 1934, pp. 80-96.