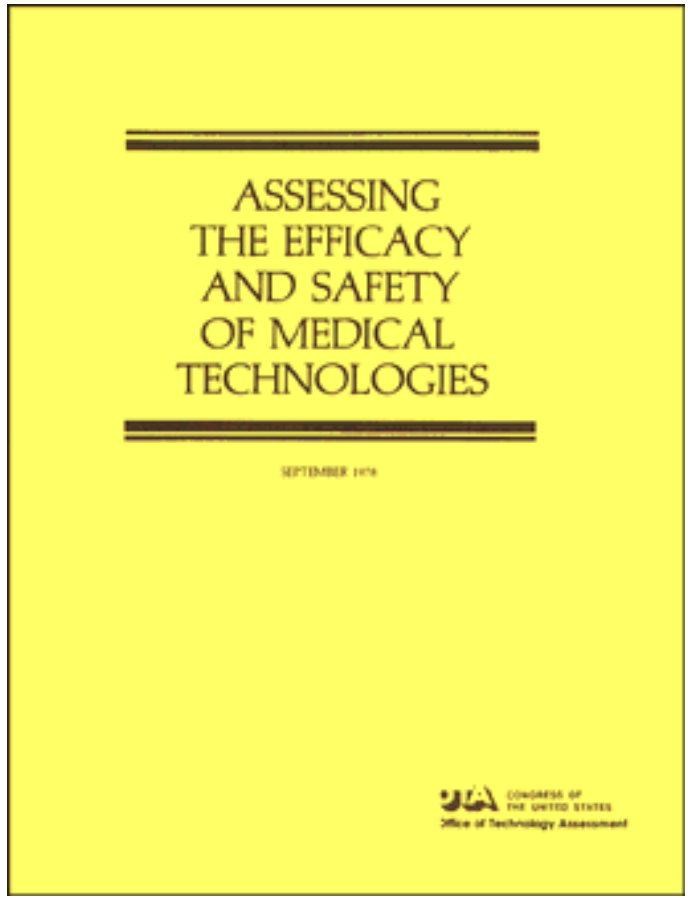
Assessing the Efficacy and Safety of Medical Technologies

September 1978

NTIS order #PB-286929



Library of Congress Catalog Card Number 78-600117

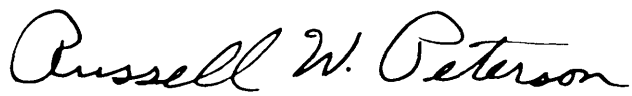
For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 Stock No. 052-003 -00593-0

FOREWORD

The Office of Technology Assessment (OTA) was requested by the Senate Committee on Human Resources ". . .to examine current Federal policies and current medical practices to determine whether a reasonable amount of justification should be provided before costly new medical technologies and procedures are put into general use. " This area of study was approved by the OTA Board in April of 1975.

Recognizing the range and complexity of issues relating to medical technologies and their use, the OTA Health Advisory Committee recommended dividing the subject into a series of discrete studies. The first report, *Development of Medical Technology: Opportunities for Assessment,* focused on assessment of the societal impacts of medical technologies. That report was published in August 1976. The second, Policy Implications *of the Computed Tomography (CT) Scanner,* examined the effects of public and private policies on the development, diffusion, use, and reimbursement of CT scanners. That study, published in August of 1978, was also requested by the Senate Committee on Finance. This report, *Assessing the Efficacy and Safety of Medical Technologies,* examines the importance and the current status of information on efficacy and safety as well as techniques and programs for generating that information.

The study was conducted by staff of the OTA Health Program with the assistance of an advisory panel chaired by Dr. Lester Breslow. It was reviewed by the OTA Health Advisory Committee, chaired by Dr. Frederick C. Robbins, and by a large number of individuals from a variety of backgrounds. The resulting report is a synthesis and does not necessarily represent the position of any individual.



RUSSELL W. PETERSON
Director
Office of Technology Assessment

OTA HEALTH PROGRAM STAFF

H. David Banta, Study *Director* (until December 1977)

Clyde J. Behney, *Study Director* (from December 1977)

Dennis P. Andrulis, Research *Associate*

Editorial Staff

Page S. Gardner Ellen Harwood

Administrative Staff

William S. Burnett M. Margaret Puglisi

Carole Stevenson Cheryl Sullivan

H. David Banta, *Program Manager*

OTA PUBLISHING STAFF

John C. Holmes, *Publishing Officer*

Kathie S. Boss Joanne Heming

ADVISORY PANEL ON EFFICACY AND SAFETY

Lester Breslow, *Chairman*
Dean, School of Public Health, University of California at Los Angeles

Bernard Barber
Professor
Department of *Sociology*
Barnard College
Columbia University

Philip A. Brunell
Chairman
Department of Pediatrics
The University of Texas
Health Science Center of
San Antonio

John C. Fletcher
Assistant for Bioethics to
the Director
Clinical Center
National Institutes of *Health*

Robert P. Goldman
President
RPG Productions, Inc.

William P. Longmire, Jr.
Professor of *Surgery*
University of California
School of *Medicine*

C. Frederick Mosteller
Professor and Chairman
Department of *Biostatistics*
Harvard University
School of Public Health

Jack D. Myers
University Professor of Medicine
University of *Pittsburgh*

Daniel W. Pettengill
Vice President
Group Division
Aetna Life and Casualty

Judith P. Swazey
Associate Professor
Department of Socio-Medical
Sciences and Community Medicine
Boston University Medical School

Kenneth E. Warner
Assistant Professor
Department of Health Planning
and Administration
University of Michigan
School of Public Health

Richard N. Watkins
Staff Physician
Medical Staff *Research Committee*
Group Health Cooperative
Seattle, Washington

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

OTA HEALTH PROGRAM ADVISORY COMMITTEE

Frederick C. Robbins, *Chairman*
Dean, School of Medicine, Case Western Reserve University

Stuart H. Altman
Dean
Florence Heller School
Brandeis University

Robert M. Ball
Senior Scholar
Institute of Medicine
National Academy of Sciences

Bernard Barber
Professor
Department of Sociology
Barnard College
Columbia University

Rashi Fein
Professor of the Economics of Medicine
Center for Community Health and Medical Care
Harvard Medical School

Melvin A. Glasser
Director
Social Security Department
United Auto Workers

Sidney S. Lee
Associate Dean
Community Medicine
McGill University

C. Frederick Mosteller
Professor and Chairman
Department of Biostatistics
Harvard University
School of *Public Health*

Helen Ewing Nelson
Director
Center for Consumer Affairs
University of Wisconsin—Extension

Anthony Robbins
Executive Director
Department of *Health*
State of Colorado

Charles A. Sanders
General Director
Massachusetts General Hospital

Kerr L. White
Chairman
U.S. National Committee on Vital and Health Statistics

ACKNOWLEDGMENTS

The Office of Technology Assessment (OTA) was assisted in the preparation of this report by a large number of individuals. OTA would especially like to thank the following people: John Abbott, Department of Health, Education, and Welfare; James W. Day, University of Texas School of Public Health; Steve Gordon, National Institutes of Health; Michael Gough, National Institutes of Health; Jack Paradise, Children's Hospital of Pittsburgh; Thomas Preston, U.S. Public Health Service; Samuel Putnam, University of North Carolina School of Medicine; and Chariklia Spiegel, Robert Wood Johnson Clinical Scholar. Joseph Mohbat and Kerry Britten Kemp should be acknowledged for their editorial assistance. The Health Program staff would also like to thank Carl A. Taylor for his contributions and guidance during the first portions of this study.

CONTENTS

Contents—continued

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF TERMS

Controlled clinical trial—An experimental research method by which human or animal subjects are assigned, in accordance with predetermined rules, either to an *experimental* group in which subjects receive technology or dosage levels of uncertain efficacy or safety or to a *control* group in which subjects receive some other technology or dosage level, usually the standard one or a placebo. If the predetermined rules specify that the subjects are assigned to groups randomly, the result is a *randomized controlled clinical trial.* The vast majority of randomized clinical trials are also controlled trials.

Device—Any physical item, excluding drugs, used in medical care (including instruments, apparatus, machines, implants, and reagents).

Drug—Any chemical or biological substance that may be applied to, ingested by, or injected into humans in order to prevent, treat, or diagnose disease or other medical conditions.

Effectiveness—Same as efficacy (see below) except that it refers to ". . average conditions of use. "

Efficacy—The probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use.

Epidemiology—The study of the frequency, distribution, and determinants of diseases and disabilities in human populations and the impact of interventions on them.

Medical technology—The drugs, devices, and medical and surgical procedures used in medical care, and the organizational and supportive systems within which such care is provided.

Morbidity—Illness, injury, impairment, or disability in an individual.

Mortality—The death of an individual; often used in epidemiological studies where mortality rates for a population for a certain disease or injury are calculated.

Placebo—An inactive substance or procedure that is often used in controlled clinical trials to evaluate efficacy. It is also used in medical practice to satisfy a symbolic need for therapy.

Procedure—A medical technology involving any combination of drugs, devices, and provider skills and abilities. Appendectomy, for example, may involve at least drugs (for anesthesia), monitoring devices, surgical devices, and physicians', nurses', and support staffs' skilled actions.

Reliability—The extent to which an experiment, test, or measurement yields the same results on repeated trials.

Risk-A measure of the probability of an adverse or untoward outcome occurring and the severity of the resultant harm to health of individuals in a defined population associated with use of a medical technology applied for a given medical problem under specified conditions of use.

Safety—A judgment of the acceptability of relative *risk* in a specified situation.

Validity—The extent to which the measures used to assess efficacy and safety accurately reflect the performance of the technology under study.

GLOSSARY OF ACRONYMS

AAMI —Association for the Advancement of Medical Instrumentation
ACS —American Cancer Society
ADAMHA--Alcohol, Drug Abuse, and Mental Health Administration
ANSI —American National Standards Institute
ASTM —American Society for Testing and Materials
BCDDP —Breast Cancer Detection Demonstration Project
CDC —Center for Disease Control
CON —Certificate of Need
CT —Computed Tomography, or Computerized Axial Tomography
DOD —Department of Defense
ECMO —Extracorporeal Membrane Oxygenator
EFM —Electronic Fetal Monitoring
EKG — Electrocardiogram
ESRD —End Stage Renal Disease
FDA —Food and Drug Administration
GMP —Good Manufacturing Practice
GNP —Gross National Product
HCFA —Health Care Financing Administration
HEW —Department of Health, Education, and Welfare
HIP —Health Insurance Plan of Greater New York
HMO —Health Maintenance Organization
HRA —Health Resources Administration
HSA —Health Services Administration
HSQB —Health Standards and Quality Bureau
IND —Notice of Claimed Investigational Exemption for a New Drug
IPPB —Intermittent Positive Pressure Breathing
MRFIT —Multiple Risk Factor Intervention Trial
NASA —National Aeronautics and Space Administration
NCHS —National Center for Health Statistics
NCHSR –National Center for Health Services Research
NCI —National Cancer Institute
NDA —New Drug Application
NEI —National Eye Institute
NHLBI —National Heart, Lung, and Blood Institute
NIAAA —National Institute on Alcohol Abuse and Alcoholism
NIAID —National Institute of Allergy and Infectious Disease
NIAMDD--National Institute of Arthritis, Metabolism, and Digestive Diseases
NICHHD –National Institute of Child Health and Human Development
NIDA —National Institute on Drug Abuse
NIDR —National Institute of Dental Research
NIGMS —National Institute of General Medical Sciences
NIH —National Institutes of Health
NIMH —National Institute of Mental Health
NINCDS —National Institute of Neurological and Communicative Disorders and Stroke
NSF —National Science Foundation
OHPA —Office of Health Practice Assessment
OMB —Office of Management and Budget
OTA —Office of Technology Assessment
PHS —Public Health Service
PSRO —Professional Standards Review Organization
SSA —Social Security Administration
TAR —Treatment Assessment Research
VA —Veterans Administration
VZ —Varicella-Zoster

INTRODUCTION AND SUMMARY

1 •

INTRODUCTION AND SUMMARY

The role of science in medicine has expanded rapidly in the past decades. As a result, the practice of medicine today is heavily, and increasingly, dependent on technology. Each year, hundreds, perhaps thousands, of new technologies enter the medical care system. New preventive, diagnostic, and therapeutic tools currently are available, and many infectious diseases can now be prevented. Innovations such as antibiotics have provided efficacious treatments for a number of conditions. Many of these technologies, and others, have undoubtedly contributed to the past century's substantial improvement in the health status of the American people. Additionally, relief of pain, amelioration of symptoms, and rehabilitation now have become possible for many patients with diseases that cannot be successfully prevented or treated.

However, concerns have arisen about the possible negative effects of the pervasive use of technology in medical care. The costs of medical care, which have escalated sharply, often are viewed as a significant societal problem. Currently, expenditures for medical care consume close to 9 percent of the gross national product (GNP); in 1960, health care costs represented 5.2 percent of the GNP. Third-party payers exacerbate the rise in health care cost because they put few constraints on expenditures. Prevailing methods of reimbursement encourage both inefficient utilization and increased provision of services, often without evidence of commensurate benefit to the patient.

Because of the lack of a direct and explicit relationship between the sharp cost increases of health care, the expanded use of medical technologies and improved health, questions have been raised about the efficiency of our health care delivery system. Additional concerns have been raised regarding the fact that many people or population groups have only limited access to medical care and its technologies. The increased role of science and technology in medicine also has led to ethical concerns regarding both the use of certain technologies, such as amniocentesis or renal dialysis, and the use of human subjects during research on medical technologies. Critics of the increased use of technology charge that medicine is being dehumanized by the use of machines and scientific methods. Some of the criticisms and concerns mentioned above may be unfair, some incorrect, and others fully accurate. Determining their validity is beyond the scope of this report. Health policy makers, though, must consider these and many more issues both comprehensively and individually. Consequently, the Office of Technology Assessment (OTA) has examined the individual issue of efficacy and safety because it is one of the prime keys to understanding many other health care concerns.

Efficacy and safety, or the direct medical benefit and risk of a technology, are the basic starting points in evaluating the overall utility of a technology. For example, ethical issues would not have been raised regarding amniocentesis if it had been demonstrated as inefficacious or clearly unsafe. In addition, efficacy and safety data are required in evaluations of the cost-effectiveness, cost-benefit, or social impacts of technologies. Well-informed decisions concerning modifications in the systems for reimbursement and the dif-

33-496 O - 78 - 2

fusion and use of medical technologies also require efficacy and safety information (153,196,204,226,254,299,340) .

Evidence indicates that many technologies are not adequately assessed before they enjoy widespread use (52,72,124,369). For example, the computed tomography (CT) scanner (355), the electronic fetal monitor (see chapter 3, case 7), and mammography (see chapter 3, case 4) are used frequently despite the lack of adequate information demonstrating their efficacy and safety. Many technologies which have been used extensively have later been shown to be of limited usefulness.

Information obtained from assessments of the efficacy and safety of new and existing medical technologies might serve three important purposes:

- To ensure that technologies demonstrated to have potential benefits with acceptable risks are made available rapidly in the private and public sectors; administrators of existing Government regulatory and financing programs could make sounder and faster decisions regarding the use of medical technologies with such information;
- To constrain the diffusion and use of technologies which either lack efficacy or cause excessive harm;
- To guide appropriate use of all technologies because they are rarely completely inefficacious or completely unsafe.

The Federal Government is concerned with questions of efficacy and safety because of its general role as protector of the public and its specific role as developer and user of medical technology. Because public funds pay more than 40 percent of the national health expenditure, concerns have naturally arisen about the benefits of medical care. Such questions seem certain to lead to increasing scrutiny of medical care expenditures and accelerated efforts to generate information on the benefits derived from the use of medical technologies. Indeed, a variety of Federal programs are hampered in carrying out their mandated tasks by lack of such information.

A state of total information on the efficacy and safety of medical technologies perhaps can never be attained because they are so numerous, complex, and varied. The task of evaluating all technologies would be overwhelming and, to OTA's knowledge, no health care expert has advocated such an undertaking. Therefore, the task of identifying and selecting technologies for assessment becomes critical.

SUMMARY

Efficacy and safety are complex measurements of actions or results that are best expressed in probabilistic terms. Efficacy is the probability of benefit from the use of a medical technology. When possible, this benefit should be expressed in terms of four factors: the type and probability of benefit, the medical problem giving rise to use of the technology, the population affected, and the conditions of use under which the technology is applied. Specifying the conditions of use serves to distinguish the terms efficacy and effectiveness. For efficacy, the conditions of use are considered to be ideal, or, as a substitute, experimental research settings. Effectiveness refers to average conditions of use.

Safety is a judgment of the acceptability of the risks posed by the use of a technology. Risk is parallel to efficacy in that it is a probabilistic measurement, and the four factors mentioned above are also part of its specification. Risk and safety can apply to either ideal or average conditions of use, but when the term "efficacy and safety" is used, it refers to safety and risk under ideal conditions.

Efficacy and safety should be considered together to be relevant to clinical or policy decisionmaking. Judgments must be made to determine whether the benefits justify the risks associated with the use of a technology in particular circumstances.

Case Studies: Issues Related to the Assessment of Efficacy and Safety

Seventeen short case histories illustrate the diverse nature of medical technologies, the difficulties in assessing their efficacy and safety, and Federal involvement in medical technology development, diffusion, and use. They also illustrate the fact that social impacts, such as economic and ethical problems, influence assessments of safety and efficacy. The cases do not exemplify all points concerning efficacy and safety; however, they do demonstrate many of the complexities that must be recognized and considered if medical technologies are to be evaluated for efficacy and safety.

Techniques for Estimating Efficacy and Safety

Techniques used in estimating efficacy and safety may take many forms. Traditionally, clinical experience, based on informal estimation techniques, has been the most important. Other techniques, such as epidemiological studies, formal consensus development, and randomized controlled clinical trials, however, are being used increasingly. The last technique, especially, has gained prominence (in the past 20 years) as a tool for assessing efficacy and safety.

No technique is universally applicable. Depending on the situation and technology, less complex methods may be more appropriate than the use of statistically sophisticated controlled trials. Frequently, combinations of various techniques are used because technology has its own strengths and weaknesses. For example, informal assessment techniques are based upon the valuable clinical experience of physicians; however, they are subject to strong biases and frequently are based on very small numbers of observations. Controlled clinical trials can draw upon larger numbers of observations and use complex statistical techniques to eliminate or reduce bias. Yet, difficulties also exist in conducting such trials. For example, trials often raise ethical concerns regarding the denial of a "promising" but unevaluated new technology to the control group members. Also, the design of the trial and the interpretation of the results are often subject both to value judgments and measurement problems. Nonetheless, all these techniques, especially controlled trials, remain powerful tools for gathering evidence on efficacy and safety.

Current Assessment Programs

Certain programs for evaluating efficacy and safety are required by Federal law. The Food and Drug Administration (FDA) administers regulatory programs which are

limited to medical products—drugs and devices. Manufacturers of these products are legally required to conduct efficacy and safety tests using the FDA guidelines. In addition, products must be licensed for marketing by FDA.

Other Federal agencies, such as the National Institutes of Health (NIH) and the Veterans Administration (VA), have no explicit mandate to assess efficacy or safety but do conduct clinical trials and other tests of efficacy and safety as part of their general mission. These trials test drugs, devices, and procedures.

The private sector supports numerous efforts designed to assess the efficacy and safety of medical technologies. In addition, the private sector often supplies the personnel and institutional resources for Federal assessment programs. However, private sector activities, particularly regarding medical and surgical procedures, are fragmented and uncoordinated. Individual physicians (either hospital or individual practice-based) are the source of many innovative procedures and much of the efficacy and safety testing done on procedures.

In summary, demonstrating the efficacy and safety of drugs and devices is required by Federal law prior to marketing. There is no corresponding requirement for procedures; however, some procedures are being tested by various Federal and private groups.

Implications and Status of Efficacy and Safety Information

Often, it is difficult or impossible to obtain information regarding the probable benefits and risks of technologies when used under actual or average conditions. Determining the efficacy and safety of a particular technology in controlled settings, therefore, represents the starting point in the effort to evaluate its potential benefit and risk. Consequently, efficacy and safety serve as the prime and critical criteria for judging the possible technical effects of medical technologies.

Any person or organization using or directly affecting the use of medical technology is a user of information on efficacy and safety. Patients, physicians, other health care professionals, biomedical researchers, and personnel in Government regulatory and reimbursement programs, public and private health planning agencies or quality assurance programs, other Federal and State health agencies, and medical schools are the prime examples of such users. Because of the large numbers of people who use efficacy and safety information, the development and dissemination of well-validated, timely, and relevant information is particularly critical.

Optimally, the processes of developing and disseminating safety and efficacy information should be coherent, coordinated, and the clear responsibility of one or several agencies or groups in the public or private sector. These processes, though very complex, can be perceived in terms of four basic elements: identification of technologies to be studied, testing through use of various techniques to generate information on efficacy and safety, synthesis of the finding of testing data and of any other relevant information—which often results in judgments or recommendations, and *dissemination* of the synthesized information to appropriate parties, including decisionmakers.

When current activities and programs for assessing efficacy and safety are compared to the optimal model described above, shortcomings are evident.

- There is no formal or well-coordinated overall system,

- Identification of technologies to be studied is a very informal, usually agency-specific process.
- Existing technologies are identified much less frequently for study than are new and developing technologies; thus, they are studied much less frequently.
- Medical drugs and devices are subject to a more rigorous process of assessment than medical procedures.
- Preventive technologies receive far less attention than therapeutic ones.
- Serious questions have been raised concerning the adequacy of funding for clinical trials.
- Synthesis activities are still too modest despite their recent expansion.
- The quality and appropriateness of medical literature, the primary source of synthesized information, has been criticized.
- Synthesis activities cannot be adequate when there is a critical lack of information regarding efficacy and safety.
- Federal agencies have not assigned a high priority to disseminating information.

These and other shortcomings may have contributed to the status of information on efficacy and safety, which may be inadequate to allow the rational and objective utilization of medical technologies. It has been estimated that only 10 to 20 percent of all procedures currently used in medical practice have been shown to be efficacious by controlled trial. Given the shortcomings in current assessment systems, the examples of technologies that entered widespread use and were shown later to be inefficacious or unsafe, and the large numbers of inadequately assessed current and emerging technologies, improvements are critically needed in the information base regarding safety and efficacy and the processes for its generation.

Policy Alternatives

Policy alternatives presented in this report are grouped into five sections outlined in chapter 7. The first section discusses alternatives to current Federal assessment activities both in terms of their expansion or change and the extent of that potential expansion. The other four sections correspond to the four steps in the assessment model. Each of these sections presents a number of options concerning the organizational location of the four functions of assessment. Following is a brief outline of these options:

Section One: Congressional Alternatives

Alternative A-1: Changes or expansions in the development of information concerning the safety and efficacy of medical technologies could occur solely in the private sector. This alternative would give the Federal Government the role of stimulating the private sector and monitoring its activities.

A-2: The Federal Government could expand its activities relating to the development of information on efficacy and safety of medical technologies. In this alternative, legislation could mandate the performance of certain activities.

A-3: Some combination of Alternatives A-1 and A-2 could be pursued.

Section Two: Identifying Technologies That Need Assessment

Alternative B-1. A new commission
B-2. Institute of Medicine
B-3. National Institutes of Health
B-4. Agencies involved in technology development
B-5. Food and Drug Administration
B-6. A new Federal office or agency, or the Office of Health Technology

Section Three: Requiring, Stimulating, Conducting, or Funding Studies

Alternative C-1. National Institutes of Health
C-2. Other Federal agencies
C-3. Food and Drug Administration
C-4. A new Federal office or agency, or the Office of Health Technology

Section Four: Synthesizing Information

Alternative D-1. A new commission
D-2. Institute of Medicine
D-3. National Institutes of Health
D-4. Agencies involved in technology development
D-5. Food and Drug Administration
D-6. Office of Health Practice Assessment
D-7. A new Federal office or agency, or the Office of Health Technology

Section Five: Disseminating Information

Alternative E-1. National Institutes of Health
E-2. Other Federal agencies
E-3. A new Federal office or agency, or the Office of Health Technology
E-4. A new office in the Department of Health, Education, and Welfare

SCOPE OF THE REPORT

This report discusses various possibilities for assessing the efficacy and safety of medical technologies systematically, thoroughly, and scientifically. It therefore focuses on efficacy and safety. Although efficiency, effectiveness, ethical, and other social concerns are related to efficacy and are also very important to the medical care system, these are not discussed at length.

Medical technologies are used for six different purposes: prevention, diagnosis, treatment, rehabilitation, patient support, and administration. The latter two classes of technology are not discussed in this report. Similarly, the report does not preview the safety and efficacy of technologies used in the psychosocial medicines, such as psychotherapy, counseling, and behavior modification. Rather, this report considers only the products of traditional biomedical research.

This report highlights both the critical need for data pertaining to safety and efficacy and the current and potential systems for obtaining such information. Once developed, this information could affect many alternative variables not discussed in this report, such

as the organization of medical care. The report discusses biomedical research and technology development in broad terms and presents a general framework for assessment; however, the policy alternatives refer primarily to options which can be implemented by Federal agencies.

ORGANIZATION OF THE REPORT

The report is organized into six remaining chapters. Chapter 2 presents the concepts of efficacy and safety, describes their characteristics, and develops working definitions of both terms. Chapter 3 outlines a short history of interest in the assessment of efficacy and safety and includes 17 brief case studies of medical technologies. The case studies are designed to illustrate various aspects of efficacy, safety, and their assessment. Particular attention is paid to highlighting policy issues raised by the use of certain technologies. Analytical techniques used to assess efficacy and safety are described in chapter 4. Chapter 5 reviews both Federal and private sector agencies and programs that engage in assessment activities. Chapter 6 discusses various aspects and the implications of the current assessment systems and programs. Building upon those implications, the last chapter presents a range of policy alternatives resulting from the analysis.

2.

THE CONCEPTS OF EFFICACY AND SAFETY

2.

THE CONCEPTS OF EFFICACY AND SAFETY

The concepts of efficacy and safety have not been suddenly discovered or created. They have always existed in medical thought. In an intuitive sense, an efficacious and safe medical technology is one that "works" and causes no undue harm. That statement may sound naive to individuals working in the field of health today. However, for a major portion of the history of medicine, efficacy and safety were measured by that intuitive standard. Furthermore, that intuitive standard still lies at the heart of medical practice, but the meaning and measurement of those concepts have evolved with increased sophistication of scientific methods in medicine.

This chapter introduces the concepts of efficacy and safety. It begins with a brief discussion of the nature of efficacy and safety knowledge, presents the characteristics and concept of efficacy, then of safety, and finally, discusses efficacy and safety in relation to each another.

THE NATURE OF EFFICACY AND SAFETY KNOWLEDGE

Measurement of efficacy and safety is in essence an examination of interventions in the processes by which various phenomena affect health and disease. Neither these phenomena (whether they be biological, psychological, or social) nor the interventions (often, technologies) need be thought of as having a fully predictable mechanistic effect. A probabilistic view of effects—that is, when an event occurs, there is a range of possibilities that other events will occur—is more useful. The concept of probability is used to summarize the effects of causal variables which are unknown or not taken into account. Thus, we can speak of estimating or evaluating efficacy and safety, but not exactly determining them. Specific technologies have certain probabilities of effects; therefore, efficacy and safety information is normally expressed in terms of probabilities.

EFFICACY

There is no shortage of definitions for efficacy; nor is there a lack of confusion relating to distinctions between terms such as efficacy, effectiveness, benefit, and efficiency. Table 1 on the following page lists several definitions of efficacy.

Despite the sometimes substantial differences among the various interpretations of efficacy, one can isolate four critical factors that, taken together, form a comprehensive view of the concept.

Table I.—Selected Definitions of "Efficacy"

Source	Term defined	Definition	Relation to four factors (See below)
Federal Food, Drug, and Cosmetic Act (363)	Effectiveness, Efficacy (interchangeable)	A drug is effective if it has "the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the proposed labeling thereof"	Benefit: Explicit Population affected: Implied Medical problem: Explicit Condition of use: Not included
A. Cochrane (72)	Efficacy (interchangeable with effectiveness)	"The effect of a particular medical action in altering the natural history of a particular disease for the better"	Benefit: Explicit Population affected: Not included Medical problem: Explicit Conditions of use: Not included
World Health Organization (435)	Efficacy	Benefit or utility to the individual of the service, treatment regimen, drug, preventive or control measure advocated or applied	Benefit: Explicit Population affected: Explicit Medical problem: Explicit Conditions of use: Not included
Discursive Dictionary of Health Care (347)	Efficacy (as a variant of effectiveness)	"The degree to which diagnostic, preventive, therapeutic, or other action or actions (undertaken under ideal circumstances) achieves the desired result"	Benefit: Explicit Population affected: Not included Medical problem: Not included Conditions of use: Explicit
Office of Technology Assessment, in this report	Efficacy	The probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use	Benefit: Explicit Population affected: Explicit Medical problem: Explicit Conditions of use: Explicit

The factors are:

1. Benefit to be achieved,
2. Medical problem giving rise to use of the technology,
3. Population affected, and
4. Conditions of use under which the technology is applied.

1. Benefit: The fact that a technology's efficacy depends heavily on its benefit to the recipient seems a simple concept. Yet the question of what outcomes represent benefits is not so simply answered. Outcome criteria have usually been restricted to measurement of mortality and morbidity; less consideration has been given to life expectancy (longevity) or psychosocial and functional factors (40,41). The definition of benefit to be used will vary depending on the goals of the investigator and the type of technology being assessed.

A range of relevant outcomes can be considered in regard to a particular technology (227). A curative technology, for example, is efficacious only if it has a direct causal relationship to a positive patient outcome. In other cases, however, the consideration of intermediate criteria may be appropriate. For example, the benefit resulting from use of diagnostic technologies can be examined at five levels (116):

1) Technical capability-Does the device perform reliably and deliver accurate information?
2) Diagnostic accuracy—Does use of the device permit accurate diagnoses?
3) Diagnostic impact—Does use of the device replace other diagnostic procedures, including surgical exploration and biopsy?
4) Therapeutic impact—Do results obtained from the device affect planning and delivery of therapy?
5) *Patient* outcome—Does use of the device contribute to improved health of the patient?

If it is assumed that the function of a diagnostic technology, such as skull X-ray, is to perform accurate diagnoses of individuals' illnesses, the evaluation of benefit concentrates on the second level. If the diagnostic technology is expected to affect therapy or eventual patient outcome, then the fourth and fifth levels would be examined. Studies at the fourth and fifth levels may be difficult to conduct because long-term followup is required. As a result of this difficulty and the emphasis on diagnostic accuracy, evaluations in terms of therapeutic planning and patient outcome are infrequently performed.

The specification of benefit is often difficult for other classes of technologies as well. For example, is the efficacy of coronary bypass surgery to be evaluated in terms of its ability to give relief from symptoms (e. g., pain) or in terms of increased longevity for the patient? Thus, two different measures of benefit may possibly yield two different statements of efficacy for the same technology. This concept is illustrated by case study 8 on coronary bypass surgery in chapter 3.

2. Medical Problem: A technology's efficacy can be evaluated only in relation to the diseases or medical conditions for which it is applied. Obviously, one would not spend much time evaluating the efficacy of plaster cast applications for controlling hypertension. In general, however, the specification of medical problems is complex and can lead to controversy regarding the evaluation of the efficacy of a particular technology. For example, hysterectomies have been performed for a variety of medical conditions: premalignant states and localized cancers, descent or prolapse of the uterus, and obstetric catastrophes such as septic abortion (see chapter 3, case 11). They may also be performed as prophylaxis to avoid possible later cancer or pregnancy. If the efficacy of hysterectomy has been estimated for one of these diseases or medical conditions, it cannot be assumed automatically that the procedure will have similar efficacy for the others.

3. Population Affected: The effect of a medical technology varies depending on the individual treated. Sometimes, however, enough uniformity of effect exists to permit careful generalizations (163). These generalizations, or extrapolations, apply to the specific population type within which the original observations were made and should be supported by valid and reliable statistical techniques. For example, in the late 1960's the Veterans Administration (VA) conducted a multi-institutional controlled clinical trial of treatment for hypertension using the drugs hydrochlorothiazide, reserpine, and hydralazine (399) (see chapter 3, case 12). The treatment was shown to be efficacious for patients with diastolic blood pressure above 105 mm mercury. But, all the patients in the trial were males. Thus, the treatment could be considered to be efficacious (based on that trial and other evidence) for the population studied, males, but no automatic assumptions can be made concerning its efficacy for females.

Assumptions cannot be made because there are physiological and other differences among various population types. Children under certain ages, for example, may be affected by the same drug quite differently than adults. Therefore, the population undergoing treatment needs to be specified when the efficacy of a medical technology is discussed.

4. Conditions of Use: The outcome of the application of a medical technology is partially determined by the skills, knowledge, and abilities of physicians, nurses, and other health personnel, and by the quality of the drugs, equipment, institutional settings, and by support systems used by those personnel during the application. Cardiac surgery, as a commonly cited example, may result in a better outcome when conducted by skillful, well-trained surgeons who frequently perform such operations than when conducted by surgeons who rarely use that technology. Similarly, a drug's benefit may be greater if correct dosages are administered at the correct times. Also, the interaction of a drug with other drugs may affect the benefit. A situation where the physician is skillful and experienced, medication is administered carefully, and the patient receives the best care possible must be described as ideal. By definition, not all physicians are the most skillful, and not all conditions of use are of the highest possible quality. Average conditions of use inherently contain a great many variables, such as physician skill, that may differ from one hospital to another, and from one application of a technology to another. Thus, it is valuable to have an outcome measure that is not dependent on the differing variables inherent in average conditions of use. Efficacy is this measure. By defining efficacy as benefit under *ideal* conditions of use, a reasonably consistent measure for that factor is introduced. No conditions of use are absolutely ideal, but, for most purposes, carefully controlled research settings can serve as a substitute for ideal circumstances. These carefully controlled situations are frequently found in research hospital settings. For example, the efficacy of ambulatory maternal care can be studied in clinics, home situations, or hospitals. The essential criterion is "best possible control of conditions. "

When the four factors described above are specified for the application of a specific medical technology, a relatively comprehensive statement has been made as to that technology's efficacy, Because a definition is merely a description of the properties of an entity, these four variables or factors can serve to define the concept of efficacy. This report uses the following definition of efficacy, not because it is necessarily more "correct" than others, but because it can be useful for discussion. It explicitly declares several key variables that, together, describe the potential usefulness of a medical technology.

> Efficacy: The probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use.

This report differentiates efficacy from effectiveness. Effectiveness is concerned with the benefit of a technology under average conditions of use. An effective technology has positive benefits for those people who are treated with the technology in a typical medical setting. Although the efficacy of a drug, for example, may be evaluated for individuals in a research setting, its effectiveness in an average setting may be influenced by variables such as those mentioned above. These variables, such as proper administration of a drug, are more rigorously controlled in a research setting. Thus, the efficacy and the effectiveness of a drug may differ.

Though they can be viewed as distinct, efficacy and effectiveness are closely related concepts. The effectiveness of a technology is estimated by methods similar to those used to estimate its efficacy; however, estimating effectiveness is often more difficult because

of the absence of rigorously controlled settings. To be desirable, of course, a medical technology should be both efficacious and effective (as is, for example, polio vaccine). A medical technology can be efficacious and of limited effectiveness (e. g., a technology that benefits individuals but can be applied by only a few highly trained physicians). But a technology that is not efficacious cannot be effective.

SAFETY

Safety, like efficacy, is a relative concept: no technology is ever completely safe, or completely efficacious. In the beginning of this chapter, a safe technology was described intuitively as one that "causes no undue harm. " Despite the apparent simplicity of that informal definition, it reflects a critical property of the concept of safety: that safety represents a value judgment of the acceptability of risk. Risk can be thought of as "a measure of the probability and severity of harm to human health" (218). This definition of risk implies that investigators and policymakers should be concerned with both the nature of the risk and the probability of its occurrence. For example, a low but measurable probability of death can be more significant than a high probability of experiencing pain, discomfort, or other minor impairments.

Thus, if the risks of using a medical technology are acceptable (to the patient, physician, society, or other appropriate decisionmaker), the technology may be considered "safe" in that instance. Safety can then be defined as a judgment of the acceptability of the risk associated with a medical technology (90). That definition is useful to organizations, such as the Food and Drug Administration (FDA), which need both to consider the risk of a technology, such as a drug, and to decide whether and under what circumstances that risk may be considered acceptable. If FDA decides that a technology has certain risks which are likely to be acceptable to a sufficient number of decisionmakers, and if it is efficacious, the agency will approve that technology for marketing.

As with efficacy, several factors must be specified when risk and safety are discussed. The *medical problem* for which the technology being evaluated is applied must be specified, not only because the medical problem or condition of the patient will often affect the action of the technology and thus the associated risks, but also because the judgment of acceptable risk depends on the type and severity of the medical problem. For example, technologies used to treat Hodgkin's disease* have types of risks that can at times be severe, although the probability of their occurring may be relatively low. A second malignancy may develop as a result of using radiotherapy and chemotherapy. Also, treatment may cause bone marrow suppression, pneumonitis, or several other deleterious effects (273). These risks, however, must be compared to the benefits of a normal life span, which is very often the direct result of treatment. Given these alternatives, the patients may regard the treatment as acceptably safe; that is, the risks are acceptable under the specific circumstances.

The *population affected* is also an important factor to be specified for reasons similar to those given regarding efficacy. For example, persons above a certain age or below a certain age may be especially susceptible to undesirable side effects of a drug, or they may be less able than most adults to withstand the rigors of a prolonged surgical procedure. Thus, the risks to those persons would be greater and more severe.

● Hodgkin's disease is a form of cancer that affects the lymphatic system.

The risk associated with a particular medical technology also depends on the *conditions of use* under which the technology is applied. The reasoning for the inclusion of this factor parallels that presented in the previous section, *Efficacy*.

For the purposes of this report, then, risk maybe defined as follows:

> Risk: A measure of the probability of an adverse or untoward outcome occurring and the severity of the resultant harm to health of individuals in a defined population associated with use of a medical technology applied for a given medical problem under specified conditions of use.

This definition covers risk under ideal (research) settings, under average or typical settings, and under conditions where quality is below average. This coverage is afforded by the specification of "conditions of use." Normally, when "efficacy and safety" judgments are being discussed, risk is assumed to be measured under ideal conditions of use.

Given this definition of risk, safety can be specified.

> Safety: A judgment of the acceptability of risk in a specified situation.

EFFICACY AND SAFETY

Efficacy and safety are separate concepts; they can be measured and discussed as distinct properties of a medical technology. Efficacy is defined in terms of a benefit; safety, in terms of a risk. There are, though, many similarities between the two concepts. Neither efficacy nor safety is absolute. Both are discussed in terms of *probability* and *magnitude* of benefit or harm. Also, both are specified by several common factors: medical problem, population affected, and conditions of use. Most importantly, however, each can be fully evaluated only in terms of the other. A technology may provide benefits, but the *value* of those benefits depends on the risks involved in using the technology.

The controversy surrounding the use of mammography illustrates the interdependency of these concepts (see chapter 3, case 4). The benefits of reduced or delayed mortality due to using mammography for detection of breast cancer must be balanced against the risk of developing cancer from radiation emitted by the mammography device. The benefits and the risks are estimated separately, but the value of the technology depends on a comparison of the two estimates. In the case of mammography, for example, the Breast Cancer Screening Consensus Development Panel, assembled by the National Institutes of Health, "found no convincing justification for routine mammographic screening for women under the age of 50" (385). Efficacy and safety evaluation, then, is one specialized form of benefit-risk analysis.

Although efficacy assessments and safety assessments are for the most part symmetrical, at least four factors differ:

1. Ranges of effects,
2. Number of people affected,
3. Whether effects are known or expected, and
4. Time period of effects.

1. Range of Effects: In assessing efficacy, a limited number of specific benefits are usually sought. A certain drug, for example, may be tested for its ability to reduce blood pressure to a safe level in hypertensive individuals. The researcher often does not expect that drug to cure or ameliorate other disease conditions. Assessing safety, however, involves consideration of the broadest range of risks that can be assessed within practical limitations.

2. Number of People Affected: An efficacious medical technology results in benefits for patients with a given medical condition. Preferably, the technology will benefit a high proportion of people having the condition. Measurement and assessment of risk, however, consider the negative health effects of a technology for even a small proportion of patients. For example, although a technology may be beneficial for many patients, FDA may judge it unsafe if only a small proportion of those benefiting suffer significant, unacceptable, negative effects. The trade-off between benefits and risks will depend on the perceived magnitude and value of both benefits and risks for the proportions of people affected in each case.

3. Known or Expected Benefits vs. Unknown or Unexpected Risks: When the efficacy of a new technology is tested, a specific type of benefit is, in general, expected. Other benefits are usually ancillary to the outcome sought. In assessing risk, however, the negative outcomes are often unknown or unexpected. And, unlike the ancillary benefits, the significance of these effects must be considered to the extent practicable before a technology is deemed of acceptable risk. When thalidomide was tested as a sleeping pill, no major negative effects were discovered. Its effects upon the fetus were not tested, and thalidomide was marketed as a safe drug. The birth defects that resulted vividly demonstrate the need to consider risks from many perspectives. However, even with extensive examination of possible risks, we cannot expect absolute safety.

4. Time Period of Effects: The benefits derived from the use of a technology often may be observed sooner than the adverse effects. The time difference in the manifestation of adverse and beneficial effects is particularly characteristic of therapeutic technologies. Some deleterious effects, such as surgical complications or certain adverse drug reactions, can be observed almost immediately; others may not occur for years after the treatment. In some cases, the offspring may suffer more harm than the patient.

33-496 0 - 78 - 3

3.

EFFICACY AND SAFETY ASSESSMENT: HISTORY AND CASE STUDIES

3.

EFFICACY AND SAFETY ASSESSMENT: HISTORY AND CASE STUDIES

The purpose of this chapter is to provide an initial perspective on the development and current state of issues relating to efficacy and safety. To provide this perspective, the chapter briefly traces the evolution of interest in estimating the efficacy and safety of medical technologies. Seventeen brief case studies are presented to illustrate many of the issues relating to such estimation.

EVOLUTION OF INTEREST IN EFFICACY AND SAFETY ESTIMATION

> Taking the expression *clinical trial* in its widest possible sense—that is, to cover the test of any therapeutic procedure applied to a sick person—it is obvious that the clinical trial must be as old as medicine itself. Even the witch-doctor trying out for the first time a new and nauseating compound must surely, like Alice nibbling at the mushroom in Wonderland, have murmured to himself 'which way?' —though he would no doubt have concealed his anxiety from his patient with the customary bedside manner. Such personal observations of a handful of patients, acutely made and accurately recorded by the masters of clinical medicine, have been, and will continue to be, fundamental to the progress of medicine (165).

As Bradford Hill's comment above indicates, the development of statistical techniques for evaluating efficacy and safety does not lessen the-historical importance of clinical judgment and individual decisionmaking: modern evaluation techniques should complement the traditional,

Today's techniques and the willingness to use them did not come about overnight; nor did they come about because physicians today are more concerned than their predecessors about the outcomes of medical practice.

As early as the 18th century, statistics and probability techniques were used, though rarely, in support of medicine and public health. Cotton Mather, the American clergyman, reported in 1721 that in the Boston smallpox epidemic of that year more than 1 in 6 persons who were not inoculated against the disease died, but that only about 1 in 60 who were inoculated did so. Though his mathematics were crude by today's standards and his "experimental design" certainly weak, this effort represents one of the very early statistical tests of the benefit of a medical technology (316). In 1759, Benjamin Franklin published an account of the success of vaccination in Boston (122). His report* contains mathematical analyses of the results of vaccination, but it is also an early example of a medical "review article" and (in many respects) of a policy analysis.

*"Some account of the success of inoculation for the smallpox in England and America. Together with plain instructions, by which any person may be enabled to perform the operation, and conduct the patient through the distemper. "

In later years of that century, mathematicians, such as Bernoulli, studied the efficacy of various vaccines. The relatively widespread introduction of vaccination after 1800, with the associated dramatic and clearly observable improvements in mortality and morbidity from the target diseases, greatly diminished the interest in statistical studies of inoculation and its relation to the prevention of disease.

Toward the end of the 18th century, though, interest arose in the use of statistics to study the effectiveness of treatments. For example, in **1793,** when Benjamin Rush announced that he had discovered a definite cure for yellow fever, William Cobbett, an English politician visiting Philadelphia, inquired as to Rush's proof that the treatment, largely bleeding and purging, was effective. Rush, like most physicians of the time, did not keep complete case records, so Cobbett assembled statistics from "bills of mortality" and discovered a positive correlation between the numbers being treated and mortality rates. Though his calculations omitted many important variables, Cobbett's claim of harm rather than cure did seem to have some basis. As Shryock states:

> Here, at any rate, was an appeal to statistical evidence against a particular therapeutic procedure—a rather unique appeal for the period. So unique, was it, indeed, that it received small attention from either doctors or laymen. Cobbett was eventually convicted of slander, fined, and practically driven out of town. Yet he actually suggested the use of statistics in therapeutic research. What one man saw, in the heat of controversy, others would realize sooner or later in the course of calm investigation(316).

In 1810, Laplace's classic study of the calculus of probabilities included a strong statement of the potential of that technique in medical research. By the middle of the 19th century a trend toward increased interest in the use of statistics in medicine became evident (203). Paralleling this interest was the increasing reliance of medicine on scientific methods and on discoveries of the natural sciences. By combining the new emphasis on linking symptomatology to treatment to pathology with the techniques of mathematics, Pierre Louis of France was able to study the effectiveness of various therapies.

His statistical techniques were simply a sophisticated method of extending the experience and quantifying the impressions of physicians (316). These techniques put forth by Louis and others in the 1800's were resisted, sometimes for logical reasons, but rapidly became an integral part of clinical investigation.

Because so many of the therapies in vogue in the first half of the 19th century were not efficacious, the increasing use of statistical techniques began to reduce substantially the number of accepted treatments. This "therapeutic nihilism" was not countered by the development of efficacious therapies to replace those discredited.

Thus, paradoxically, the advance in medical science represented by submitting therapies to quantitative evaluation was one of the contributions to the fairly widespread loss of public confidence in medicine during the last half of the 19th century (316). In this period, however, the study of microorganisms and their role in disease was beginning to produce a base for later prevention and treatment.

These and other developments in medical science led to a number of striking advances, beginning roughly at the turn of the century. Mortality and morbidity declined rapidly—perhaps substantially as a result of improvements in the environment and personal habits, but also often because of medical preventive and therapeutic measures. The use of statistics and the scientific assessment of efficacy and safety grew slowly and were not generally regarded as a critical aspect of medicine. The reasons were numerous: the successes of the first third of the 20th century seemed evident, public confidence in medicine was high, and few legal requirements to demonstrate efficacy and safety existed.

Increasingly, however, medicine has been directed toward the chronic and degenerative diseases. Because it is more difficult to study the effects of treatments for such conditions, the state of medicine's ability to evaluate efficacy and safety became more critical. This situation, added to the growing interest of a small number of individuals (such as Bradford Hill) and improvements in the mathematical techniques available, led to debate as to the appropriate form, role, and magnitude of efficacy and safety evaluation. Miller states:

> The conquest of most of the acute and chronic infections in the developed world has left medicine now preoccupied with a large number of diseases of multiple etiology and long duration, where the assessment of therapeutic results presents real difficulties(242).

Barnes agrees th0 - scientific studies are essential, but he contrasts the relatively unsophisticated techniques of the early 20th century to today's techniques:

> Possibly the most critical and central defect in these cited studies of (late 19th, early 20th century) innovative surgical therapy is the lack of control experience, The concept of controls appeared to be totally unknown to the surgeons of this period. . . . (23)

The reluctance of physicians to embrace a statistical approach to effectiveness continued into the 20th century (163). In 1921, a writer in the *Lancet* asked whether the quantitative method was an "important stage in the development of (medicine)" or a "trivial and time-wasting ingenuity as some hold" (164). By 1971, Hill was able to report the medical community's answer, which was a "remarkable and increasing acceptance of the method. " In 1938, the Federal Food, Drug, and Cosmetic Act was passed, requiring th0t the safety of new drugs be demonstrated by scientific investigation before marketing was allowed. Cochrane believes th0t a "critical step forward" in the use of experimental methods in clinical medicine took place in 1952, when Daniels and Hill published their study of the efficacy of chemotherapy for pulmonary tuberculosis (72). The 1962 Kefauver-Harris Amendments to the Federal Food, Drug, and Cosmetic Act added the requirement th0 - efficacy as well as safety be demonstrated for drugs. Since 1976, certain medical devices have been required to be demonstrated as safe and effective.

The 1970's present a contrast. The techniques available to estimate efficacy and safety are more sophisticated th0n ever, and 0 - the same time concern is increasing about too little, too *much*, and *inappropriately timed* evaluation of efficacy and safety. These issues are discussed in chapter 6, but the next section of this chapter illustrates many of them by presenting 17 brief case studies.

CASES ILLUSTRATING EFFICACY AND SAFETY ISSUES

As mentioned above, medical technology has transformed medical practice in the past several decades by making new preventive, diagnostic, and therapeutic tools available to the medical care system. On the other hand, the accelerating pace of technological development has raised a number of troubling issues. Questions are being raised about whether current research and development efforts are directed at developing the most desirable technologies, whether new technologies are adequately assessed for safety and efficacy before they come into widespread use, and whether valuable technologies come into general use as rapidly as they might.

One way to address these issues is to assess the efficacy and safety of new medical technologies prior to diffusion and, when possible, existing medical technologies where serious doubt exists as to their effects. The nature, aims, current status, cost, and policy

implications of medical technologies will all influence evaluation of their safety and efficacy. The 17 case studies presented here illustrate a variety of points concerning the efficacy and safety of medical technologies. They do not, however, illustrate all the possible issues or concerns, nor are they intended to be complete reviews of the efficacy and safety of the particular technologies used. The cases may sometimes touch on effectiveness or cost-effectiveness. Although this report is not concerned directly with these issues, the relationships between efficacy, effectiveness, and cost-effectiveness are important. The cases can often serve as an introduction to the interworkings of these concepts.

Cases of accepted efficacy are included, as are cases of uncertain efficacy. Expensive technologies are represented, as well as relatively inexpensive ones (at least on a per-unit basis). Some of the technologies have already been widely diffused; others are only beginning to diffuse. Several cases show considerable Federal Government involvement; others, relatively little. Taken together, they are intended to demonstrate some of the complexities that must be recognized if medical technologies are to be evaluated for efficacy and safety.

Case 1: Pap Smear for Cervical Cancer*

The Pap smear test is an analysis of cells taken from the uterine cervix (neck of the uterus) to screen for cancer of the cervix. These cells are usually and most effectively obtained by scraping the cervix. The smear may be taken in a doctor's office, clinic, or hospital. The procedure is quick and simple but may cause some discomfort. Its safety has never been questioned.

In 1973, cervical cancer accounted for 6,000 deaths and ranked fifth among cancers for women as a cause of death (U.S. Vital Statistics, 1975). The death rate from, and incidence of, cancer of the cervix have been declining in the United States since before screening began. The American Cancer Society (ACS) estimates 20,000 new cases annually. The disease is more prevalent among women of lower socioeconomic classes, women who begin sexual intercourse at an early age, and women who have many sexual partners.

The Pap smear has been widely promoted for annual use in the United States. In 1973, 75 percent of U.S. women over the age of 17 had had a Pap smear at least once and nearly half had had one in the year prior to the survey (National Center for Health Statistics, 1975). No other country in the world has achieved this level of screening.

The average cost of examining a Pap smear by a cytological laboratory is about $5, with a range of $3 to $10. The actual costs of screening are, in fact, higher, as they should include the cost of the gynecologist or clinic unless the visit is for other purposes. One must also count the costs of followup for definitive diagnosis for those women who have abnormal Pap smear results but do not have any disease (false positives).

The test was generally accepted when it was introduced in 1943. In recent years, however, health professionals, particularly epidemiologists, have disagreed over the efficacy of the Pap smear as a screening device. The controversy has centered on three issues: the natural course of the disease, the accuracy of the test, and the efficacy of screening in lowering cervical cancer mortality rates.

The test results of the Pap smear are usually reported in five classes: I—normal, II—atypical, III—suspicious (dysplasia), IV—carcinoma in situ, and V—invasive carcinoma.

● This case was adapted from a paper prepared for OTA by Anne-Marie Foltz, Yale University School of Medicine.

Some laboratories use as many as seven classifications and the names may differ. The first two classes are considered normal. Dysplasia and carcinoma in situ are considered "precancerous," while the last class is malignant.

For those with tests in the last three classes, the usual diagnostic procedure in the past and in areas without colposcopically trained physicians is a diagnostic-cone biopsy (removal of a section of the cervix). This is an in-hospital surgical procedure and carries a risk. Today, the usual procedure is a colposcopic evaluation (essentially, looking at the cervix with 15 x magnification) and biopsies (tissue samples) if there appear to be abnormalities on the cervix. Colposcopy, a relatively recent technique, requires some training.

Cervical cancer is widely believed to pass through three stages: dysplasia, carcinoma in situ and invasive carcinoma, with the process taking up to **35** years (62). This progression is supported by the evidence that the peak incidence of these conditions occurs at progressively higher ages, with the peak of invasive carcinoma at the ages of 60 to 64 (68).

There is also some evidence to the contrary. Invasive cancer has been found in women regularly and recently screened (Sandmire et al., 1976). The explanation may be that there are slow-growing tumors that pass through the three phases over 20 to 30 years, while others become invasive within a year. These findings are consistent with findings for lung and breast cancer (Charlson and Feinstein, 1974; Wells and Feinstein, 1977). Dysplasia has been found to regress, though probably not permanently (Stern and Neely, 1963); and in the few cases where carcinoma in situ has gone untreated, it has not necessarily progressed to invasive cancer (Spriggs, 1971).

This uncertainty about the natural history of the disease affects the efficacy of the test. It is difficult to evaluate efficacy if one cannot be certain what is being prevented.

The issue of the accuracy of the Pap smear test did not receive much attention when the test was disseminated. The accuracy of the test has been stated to be about 95 percent (ACS, 1975; Dickinson, 1972). However, this statement is misleading. In any condition with a low prevalence, such as cancer of the cervix, this statistic can hide a proportion of missed lesions (false negatives rate).

Recent studies have shown that cytologists read test results differently, particularly regarding carcinomas (Seybolt and Johnson, 1971; Lambourne and Lederer, 1973; Kern and Zivolich, 1977). Some of this variance occurs because the different classes are not clearly defined. Because of this variability, the number of lesions that are missed (false negatives) can vary from 2.4 to 40 percent (Husain, 1976; Coppelson and Brown, 1974). This variability also occurs among the pathologists who read the followup biopsies. Such misreading may lead to unnecessary hysterectomies, as this is the usual treatment for invasive carcinoma in situ in women who are not interested in future childbearing (Brudnell, 1973). However, hysterectomies also carry risks (see case 11).

Finally, false positive test results (those women with abnormal test results but no disease) are rarely reported in the literature, although these women may be subject to repeated test, biopsies, and perhaps hysterectomies with their concomitant personal and social costs.

The efficacy of the Pap test was not carefully studied before its wide diffusion. An efficacy experiment would have compared the rate of invasive cervical cancer and resulting death rates in a screened and treated population with those in an unscreened population. By the end of the 1950's, however, professional consensus endorsed the posi-

tive benefit of the screening procedure. A controlled clinical trial is difficult to carry out once a procedure is generally accepted as efficacious. Denying the supposed benefits to a segment of the population was believed to be unethical. However, it can be argued that if one can demonstrate substantial doubt of efficacy, it would be unethical not to study the technology.

In the absence of such a controlled study, other epidemiological methods have been used to estimate the value of Pap smear screening in cancer control. The two long-term screening projects on large populations have been those of Boyes in British Columbia and Christophersen in Louisville, Ky. The latter project has been supported by the National Cancer Institute (NCI). Reports from both studies, which have been operating more than 20 years, indicate that screening has led to a decline in mortality from cervical cancer. Christophersen has stated:

> That a decrease in death rates of the magnitude observed here is not to a major extent due to mass screening must be proved by a demonstration of a comparable decrease in an unscreened population. Such evidence has not been presented to date(68).

The decline in mortality was found to be significantly correlated with the intensity of screening in each State in the United States, but this may have been an artifact of intervening variables (Cramer, 1974). A more cautious analysis of screening and mortality, using Canadian data and controlling for socioeconomic variables, concluded that, at least for the age group 30 to 64, over the period 1960-62 to 1970-72, the intensity of screening had a significant effect on reduction of mortality (Miller et al., 1976).

It seems safe to say that screening seems, in some cases, to have had some effect on mortality. Proponents of an annual or frequent screening program cite the preventability of invasive cancer, the low cost of the test, the relation of screening rates to a fall in mortality, the need for frequent screening to catch fast-growing tumors, and the fact that any death from cervical cancer is preventable and therefore all women should be screened frequently (Guzik, 1977).

Opponents cite the low prevalence of the disease, the uncertainty of its natural history, and the accuracy of the test. Opponents may concede that screening has lowered mortality rates, but they point out that this seems to have occurred in areas such as Abderdeen, Scotland, where the screening intervals are not annual, but every 5 years (MacGregor, 1976).

In Canada in 1976, the Conference of Deputy Ministers of Health appointed a task force to evaluate the effects of screening. After **a** careful review of the scientific literature and in light of the costs of the program, they recommended in June 1976 that screening should be undertaken at the following intervals:

> A woman should have her first smear at age 18 if she is sexually active. If the initial smear is satisfactory, a second smear should be taken a year later.
>
> After that, further smears should be taken at approximately 3-year intervals until the age 35 and thereafter at 5-year intervals until age 60.
>
> Women at continuing high risk should be screened annually (62).

Because the Pap smear is a screening procedure, its efficacy and safety are not regulated by the Federal Government. The Center for Disease Control (CDC) and other Federal and State programs do regulate the quality of clinical laboratories that perform the cytological analyses, and Federal funds have been available to train cytologists. The Cancer Control Division of NCI in the past has supported research and screening pro-

grams through grants. Since 1974, in response to specifications in the 1974 Amendments of the Cancer Act, NCI's division of Cancer Control and Rehabilitation has supported the use of the Pap smear in 38 States through contracts with State health departments. These programs, which focus on reaching women who have never had a Pap smear, are being phased out. The Health Services Administration (HSA) of the Department of Health, Education, and Welfare (HEW) supports Maternal and Child Health Clinics and migrant health programs, which both offer the Pap smear. HEW also supports 4,500 family planning service sites that are required to provide Pap smears to all women using their services. The Pap smear is not covered as a benefit by the Medicare program, and its coverage by private insurance programs varies.

In summary, the Pap smear was widely diffused for 30 years without demonstration of its efficacy through controlled trial. Since then, its use has not been questioned, but its accuracy and the frequency of necessary screenings have been. Once the Pap smear was in widespread use, the very extent of use and professional consensus of its efficacy argued against carrying out a controlled trial. As the risks to women whose tests were found falsely positive by the Pap smear have never been seriously documented, it is possible that a controlled trial to examine that question may be of value. As case 1 illustrates, it is important that some method exist for bringing questions about the efficacy or safety of techniques technologies to the attention of investigators and public or private research policymakers.

Case 2: Amniocentesis*

Amniocentesis (from the Greek "amnion," the membrane surrounding the fetus within the uterus, and "centesis, " puncture) can be performed at various times during pregnancy for a variety of reasons. But it has come chiefly to refer to the most widely employed form of prenatal diagnosis. In this role, it is a method for obtaining a sample of the fluid that surrounds the fetus by inserting a hypodermic syringe through the abdominal wall into the uterus, generally at about 16 weeks gestation.

The procedure has been in existence for some time. Its use for discovering fetal sex was first reported in 1956 (I), but it did not come into wider use as a diagnostic technique until the early 1970's. The delay was partly for technical reasons: it was necessary to develop ways of examining constituents of the amniotic fluid that would reveal a disease or defect in the fetus. Development of amniocentesis also depended, however, on an important political change taking place about then: the loosening of legal restrictions on abortion, culminating in the 1973 Supreme Court decision (Roe v. Wade, 410 U.S. 113) that made abortion before 24 weeks gestation a matter to be decided between a pregnant woman and her physician, without State interference. That is because the goal of amniocentesis is information about the fetal state—information that will lead to the prevention of many kinds of birth defects by preventing the birth of those afflicted by them, via abortion.

An additional reason for the somewhat cautious early development of amniocentesis was concern about its safety, as it involved direct invasion of the uterus, and the risk to either mother or fetus was unknown. The National Institute of Child Health and Human Development (NICHHD) coordinated a study that pooled data on more than 1,000 cases of amniocentesis from nine major medical centers that were pioneering the technique; the results of that study were announced in the fall of 1975 and published a year later (2). The findings, since confirmed by other studies elsewhere in the world, were that amniocentesis was both safe and accurate. The difference between the rate of spontaneous

*This case is adapted from a paper prepared for OTA by Tabitha M. Powledge of the Hasting Center.

abortion among women who had undergone amniocentesis and the control group women who had not was not statistically significant, and maternal complications—vaginal bleeding, for instance—were minor. Followups of babies born after amniocentesis, at birth and at the age of 1, reveal no differences between them and other babies; longer term followups will, of course, have to await the passage of time. At this point, however, the technique (assuming it is done by qualified people) appears quite safe for both mother and baby—an important consideration, because in about 96 percent of amniocentesis cases, the tests will reveal no abnormality and the pregnancy will therefore be brought to term. The study also revealed that the error rate in diagnosis was substantially below 1 percent, Like safety, accuracy is an exceptionally important consideration in amniocentesis, as a wrong diagnosis will usually lead either to the birth of an unwanted, afflicted child, or the abortion of a wanted, unafflicted one.

A variety of tests can be performed on amniotic fluid, and others will probably be developed. Fetal cells obtained from the fluid can be laboratory-cultured and karyotypes (pictures of the fetal chromosomes) prepared from them; this procedure takes several weeks. The cells can also be examined for a variety of very rare biochemical abnormalities. Other constituents of amniotic fluid, such as hormones, can also give information about the fetal state. One expanding area of amniocentesis is the assessment of amniotic fluid α=fetoprotein, which is diagnostic of several kinds of birth defects, particularly the neural tube defects anencephaly and spina bifida.

Candidates for amniocentesis are drawn from groups of women thought to be at higher-than-average risk for bearing a child with a birth defect. This can sometimes be (and usually is, in the case of the rare biochemical abnormalities or sex-linked disorders) because she has previously borne an afflicted child. But the largest number of amniocenteses is performed on women over 35 (or, in some places, over 37 or 40), who are statistically at higher risk than younger women for bearing a child with a chromosome abnormality, particularly Down's syndrome, the most important single cause of severe mental retardation. Amniotic fluid α = fetoprotein assessments are becoming an increasingly important part of amniocentesis, particularly because of pilot programs such as the one currently going on in Nassau County, N. Y., where the assessments are used to confirm the less reliable diagnoses of neural tube defects obtained via assessment of α= fetoprotein levels in the blood of pregnant women (3).

The procedure appears safe (except, of course, for the affected fetus), but is it efficacious? Diagnostic accuracy alone satisfies only one of the possible standards of efficacy (see chapter 2). Though in almost all cases the results of the tap will be negative and therefore provide prospective parents with months of relief from anxiety, a small preliminary study has revealed a high rate of depression among mothers *and* fathers in cases where an abortion followed a positive diagnosis (6). The parents under study, however, did declare that, despite its psychological effects, they would certainly repeat the procedure rather than bear a defective infant. The situation will probably be worse for those parents who are opposed to, or ambivalent about, abortion.

Another problem is knowing which is the "defined population" in which the efficacy of amniocentesis will be judged. The mother alone? The fetus? The entire family, whose resources may be spared by prevention of the birth of an affected child? Society, whose resources are also at stake when care for the chronically ill or retarded is involved? This latter point is important because, while amniocentesis is usually justified as a needed service to individuals, a strong second line of argument has been that it can relieve some burdens on society. A proposal emanating from the Columbia School of Public Health for a gradual four-stage program that would eventually reach all pregnant women at-

tempted to demonstrate that even such a massive program would provide huge monetary savings over the cost of institutionalizing those with Down's syndrome (8). On the other hand, that same money might be more efficiently spent on improvements in prenatal nutrition or delivery procedures that might reduce the amount of mild mental retardation that is much more widespread in the population and may, on balance, constitute more of a burden to society than Down's syndrome.

Some other considerations that either bear on the question of what constitutes usefulness (broadly defined) or demonstrate that "efficacy" (narrowly defined) can be only a partial measure of the usefulness of medical technologies are:

. The cost of amniocentesis is not trivial. It currently ranges between $300 and $500, depending on the amount of laboratory work involved. Increasingly, that cost is being borne by third parties, either insurance companies or the State.

. Widespread use of amniocentesis will require a large and expensive personnel training program; most laboratories doing this work are already operating at capacity. The labs will also have to be monitored. The Federal Government seems the logical focus of both training and monitoring, but that, once again, means the cost will be borne by society rather than the individual.

• Amniocentesis provides a nearly foolproof way of finding out the sex of a fetus. Though not often employed in this way in the past (except for diagnoses of sex-linked diseases), its use for the purpose of picking the sex of children is likely to increase as facilities expand and more private physicians are trained in the technique. This in turn may require an array of public policy decisions, at least on the question of whether or not such use of amniocentesis should be subsidized.

In summary, amniocentesis appears to be safe and efficacious. Complex ethical and legal issues surrounding the use of this technology must be taken into account in an evaluation of its societal usefulness. Amniocentesis is peculiar because it depends in part for its effectiveness upon the wide availability of abortion. The fate of amniocentesis is therefore tied to the abortion debate in this country. This case demonstrates the problem of viewing efficacy and safety as the sole determinants of appropriate use.

Case 3: Chicken Pox Vaccine

A successful vaccine produces, without harm to the recipient, a degree of protection that approaches the immunity that follows a disabling attack of chicken pox itself. A vaccine is a preparation of bacterial or viral material that has been inactivated or weakened. This material can stimulate the body's immune system and prepare it to attack the agents of the corresponding disease should they invade the body.

By preventing disease, rather than treating them or their symptoms, vaccines have averted suffering and saved lives. Immunization programs have reduced financial as well as human costs. Vaccines such as those against smallpox, measles, and tetanus have prevented a variety of infectious diseases.

Chicken pox is an infectious disease caused by the Varicella-Zoster (VZ) virus. Chicken pox is a common, usually mild, childhood illness. The United States recorded 154,248 cases of chicken pox in 1975, an increase of almost 13,000 from the previous year (388). In 1974, only 106 deaths were reported from chicken pox.

Early in 1977, Japanese investigators reported the development of a live, attenuated (weakened) virus vaccine against VZ. Twenty-six children who had been exposed to

chicken pox were given the vaccine after exposure. None of them developed clinical chicken pox. A control group of 19 exposed children was left unvaccinated, and all developed typical chicken pox. Blood tests demonstrated development of immunity to chicken pox by those given the vaccine. On the basis of preliminary evidence, the Japanese vaccine can be considered to produce immunity to VZ virus and to prevent the development of chicken pox in those inoculated (12,13).

A vaccine against chicken pox, which appears to be a definite possibility, might prevent thousands of cases and more than 100 deaths per year. Most of those deaths, however, probably occur in individuals at high risk, such as those with leukemia. The danger of chicken pox in high-risk individuals could be reduced by a well-organized program of passive immunization with gamma globulin, which contains antibodies against chicken pox produced by other infected individuals (48). *

The risks of the chicken pox vaccine are unknown. Some normal children could react adversely. Because high-risk individuals would be likely to have variable reactions to the vaccine, it could be expected to cause a certain level of morbidity and mortality in this group (48).

The most worrisome risk, however, is the possible effect of the vaccine on the rates of zoster, another disease caused by the VZ virus. Zoster (shingles) is usually a disease of adulthood. It occurs in persons who have recovered from chicken pox many years after their recovery. The percentage of individuals who develop latent infection** is unknown, although the rate of zoster in high-risk populations, such as those on chemotherapy for cancer, has been reported to be as high as 50 percent (6). Why latent virus causes disease years later is not known, and the relationship between chicken pox and zoster is not well understood (134).

The vaccine could postpone infection from childhood, when it is a mild illness, to adulthood, when it may be quite severe. This could occur if immunity produced by immunization of infants and children waned in adulthood. Although reimmunization might prevent this problem, persuading adults that they need a vaccine against chicken pox might be difficult. In addition, the attenuated virus might itself become latent and cause infection years later. Because the latent period for zoster can be 10 to 30 years, results of vaccination would need to be studied for decades in order to establish the health benefit (48,388).

Furthermore, viruses related to the VZ virus have been shown to cause cancer in animals and have been related to some cancers in humans. This suggests some risk of producing cancer with such a vaccine (48).

Weighing these possible benefits and risks, Brunell has stated that “the mortality and morbidity produced by varicella (chicken pox) in normal children could hardly justify a major effort to eradicate varicella” (48). The National Institute of Allergy and Infectious Disease (NIAID) agrees regarding the use of live vaccine in normal children. Carefully controlled trials in children with cancer or leukemia under conditions of isolation may merit consideration. A killed or inactivated vaccine, such as a subunit vaccine free of nucleic acid, is also a possibility for varicella as for other herpes viruses (388).

‘*Passive* immunization refers to injection into the patient’s body of antibodies derived from another source, human or animal. *Active* immunization occurs when the antibodies are produced by the patient due to injection of a vaccine. *Passive* immunization of gamma globulin, therefore is not a full substitute for vaccination.

* *The state of continuing viral infection without clinical illness is referred to as “latency.”

NIAID has shown that a drug, adenine arabinoside, reduces the mortality of herpes encephalitis, thus offering promise that a drug efficacious against VZ virus may be developed. A natural antiviral substance, interferon, has been shown to limit the severity of VZ infections manifested as herpes zoster.

The live chicken pox vaccine is a case where efficacy can be predicted, but long-term risks in normal children are unpredictable without studies that would take decades. The benefits, although positive, are relatively small, while potential risks are large. This concern about potential risks has led NIAID to decide not to test the vaccine in normal children. A live chicken pox vaccine for general use seems to be a technology that will work, but that will not be developed in this country unless further research makes it possible to minimize risk.

Case 4: Mammography

Mammography is a special X-ray examination of the breast with a machine designed for that purpose. It is used both as a screening procedure on apparently healthy females and as a diagnostic procedure in clinical situations to detect breast cancer and to aid in diagnosis of the disease. The recent controversy and studies concerning mammography relate to its use in screening, not in clinical diagnosis. This case, therefore, examines the use of this procedure only for screening.

Breast cancer, the most common cancer among women in the United States, represents 27.2 percent of all cancers in women. It is diagnosed in about 90,000 women annually, and every year about 34,000 women die of breast cancer (8). It is the leading cause of death among women 40 to 44 years of age. Its mortality and incidence rates increase with age. Its incidence in the United States has increased since the mid-1940's (8,76).

Studies carried out before current treatments for breast cancer were available estimated mean survival from onset of symptoms at about 39 months, Only 18 percent of affected women survived 5 years without therapy (304). Today the overall 5-year survival rate is approximately 60 percent. (The two percentages cannot be directly compared, because breast cancer is being diagnosed earlier, and in the earlier study, survival was dated from onset of symptoms.)

Efforts to improve survival rates have emphasized early diagnosis and treatment. If the cancer can be found and surgically removed before it metastasizes (spreads) to other organs, the survival rate is good: a 5-year survival rate of more than 80 percent. * This has led to recommendations for periodic breast examination by a physician, for monthly self-examination of the breasts, and for periodic mammography.

Mammography was first used on patients in 1913, began to have more clinical use in the 1930's, was considerably improved by Egan in the 1950's, and was widely used starting in the 1960's. In the early 1960's, M. D. Anderson Hospital in Houston, with support from the National Institutes of Health (NIH), carried out a clinical trial of Egan's technique of mammography for the diagnosis of breast cancer, X-ray findings were correlated with pathological diagnoses of breast cancer on normal breasts in 1,580 patients. The technique was found to be reasonably accurate, with a false positive rate of 7 percent and a false negative rate of 6 percent. These findings encouraged use of mammography in screening for breast cancer (70).

● Information supplied by the National Center Institute.

In the mid-1960's, Shapiro, Strax, and Venet conducted a controlled clinical trial to see whether annual screening for 4 years with clinical examination and mammography affected mortality from breast cancer (313). More than 60,000 women were divided into a study group and a control group (313). The study (usually referred to as the HIP (Health Insurance Plan of Greater New York) Study) found that "repetitive screening with clinical examination and mammography leads to at least a short-term reduction in mortality from breast cancer. Over the 7-year period of observation for which data are available, there were 70 deaths due to breast cancer in the total study group as compared with 108 breast cancer deaths in the control group" (312). In a recently completed 9-year followup, Shapiro found 90 breast cancer deaths in the study group and 128 in the control (310).

Early findings from the Shapiro study, which remain valid after 9 years of followup, led NCI and ACS to support and promote, beginning in 1973, a Breast Cancer Diagnosis Demonstration Project (BCDDP). This program has involved some 270,000 women, ranging in age from 35 to 74, from 29 participating screening centers at a total FY 1977 cost of $9.5 million (386).

The screening program consists of instruction in breast self-examination, an initial clinical history and physical examination, mammogram (for certain age groups), and thermogram. All are repeated annually for 5 years, with a 5-year observation period after completion of screening. The estimated cost for the a complete individual examination is $35 per year. By 1976, about 1,800 cases of breast cancer had been found, at an approximate cost of $11,000 per case.

Recently, however, the safety of mammography has been questioned. Radiation dose from mammography can be as high as 6.5 rads per examination. Bailar stated that the risk to symptom-free women of getting cancer from high exposures might equal or exceed the benefit of finding a cancer early that could not be found by physical examination (16).

This potential risk led NCI and ACS to appoint three* expert committees in 1976 to assess the risks and benefits of screening, particularly with mammography and physical examination. Breslow chaired a group that reanalyzed the HIP data and affirmed the lack of benefit for women under the age of 50 and substantial efficacy for those over 50 (383). Although the Breslow Committee noted also that radiation dosage from mammography had decreased because of technological improvements since the HIP study, the committee was unable to ignore findings of the HIP Study because it was the only controlled study that examined the important parameter: overall reduction, from screening, in the number of deaths from breast cancer.

Another group, chaired by Upton, reviewed evidence concerning health hazards to those screened. It focused on whether radiation to the breast can cause cancer of the breast. The committee concluded that it can. It argued that even small doses of radiation to the breast are risky. The committee postulated a l-percent increase in risk with a dose of 1 rad to the breast. Assuming this dose, application of mammography to the entire population would thus add six avoidable cases of breast cancer per rad per 1 million women per year, after a 10-year latent period (383). However, Upton also noted that new equipment such as that used in the BCDDP delivers a radiation dose under 1 rad, perhaps as little as 100 mrad (0.1 rad).

● The third group studied the pathology of breast cancer and will not be considered here.

The three groups together made a series of recommendations (383):

1. That rigorous attempts be made to keep radiation dose under 1 rad per screening examination;
2. That mammography for routine screening of women under 50 years of age be discontinued; and
3. That NCI support a clinical trial of mammography to furnish more conclusive evidence of its usefulness.

On August 23, 1976, NCI and ACS, in a letter to directors and coordinators of demonstration projects for breast cancer detection, quoted findings from a preliminary report of the Breslow and Upton groups and concluded: "We cannot recommend the routine use of mammography in screening asymptomatic women ages 35 to 50 in the NCI/ACS BCDDP at this time. However, in the face of a very small presumed risk for any individual woman, we do not recommend withholding mammography from a woman age 35 to 50 years if she and the physician agree that it is in her best immediate interest" (117). This recommendation was intensified in May 1977, when BCDDPs were told that mammography was to be used on women with personal or family histories of breast cancer.

Because of the controversy surrounding mammography, NIH held a 3-day conference on breast cancer screening in September 1977. Sixteen leading scientists, epidemiologists, physicians, and lay persons reviewed technical information, ethical issues, and other information (on the development of the BCDDP project) and heard testimony from a variety of groups. They also considered the report of a special group set up by NCI to review the BCDDP data (384), The panel concluded that "the only sound scientific evidence which demonstrates favorable benefit in breast cancer screening is derived from the HIP Study. " Because of this and because of the radiation risk, the panel recommended continuation of the screening program in women age 50 and over, but recommended limitations for younger women. For women 40 through 49 enrolled in the BCDDP, the panel recommended mammography for women having personal histories of breast cancer or whose mothers or sisters have such histories. This was consistent with existing NCI guidelines initiated in September 1976. For women below 40, mammography was recommended only for those with personal histories of breast cancer (385). This recommendation has been incorporated into recent BCDDP guidelines.

The Bureau of Radiological Health of the Food and Drug Administration (FDA) has the responsibility to regulate X-ray machines used in mammography. Most observers agree that the radiation dose from use of mammography in the general community is too high. FDA and NCI, under an interagency agreement, are attempting to decrease the exposures used in State-operated, community-level programs of mammographic screening.

Third-party payment programs, including Medicare and Medicaid, cover mammography for diagnostic purposes. Screening is not consistently covered.

In summary, mammography is a screening tool for early detection of breast cancer that has been widely used as a result of studies in the 1960's. Questions about its safety have recently been raised, and it has become a controversial technology. Many believe that technological improvements make it efficacious and safe for all women, but there is no scientific information derived through controlled studies to support such a view. Most, including the NCI panel, believe it has been shown to be efficacious for women over the age of 50 and should be used routinely for that group. Because existing information did not adequately answer the question of net health benefit, NIH collected all available information and conducted the exercise described above. Not all controversy

33-496 0 - 78 - 4

was settled; therefore, future studies will probably be necessary, especially concerning the question of benefit in the 40 to 49 age group using the modern mammographic equipment. Mammography is an example of a medical technology that was being widely diffused before questions about its safety began to countervails conclusions about its efficacy, leading to a scientific controversy that may yet strike the proper balance for society.

Case 5: Prophylactic Oral Antibiotics in Elective Colon Surgery

Prophylactic use of antibiotics, the routine administration of such drugs prior to surgery to prevent postoperative infection, is very common in surgery on the abdomen. Each year approximately 217,000 persons undergo surgery on the intestines for such conditions as cancer of the colon (large intestine), polyps, and chronic ulcerative colitis (374).

A common complication of bowel surgery is contamination of the incision (wound) by bacteria normally found in the gut. Such contamination can lead to abscess formation, generalized sepsis (infection) with serious morbidity, and even death. Antibiotics began to be used to prevent postoperative infection shortly after their introduction in the late 1940's. Specific antibiotics have been found to destroy certain types of bacteria; thus, knowledge of the types of bacteria in the gut and the types of bacteria that cause wound infections permit identification of antibiotics that might be efficacious in preventing infection. Such antibiotics can be administered in several ways: intravenously (injected into the bloodstream) during, before, or shortly after surgery; orally a few days prior to surgery; applied locally following surgery; or combinations of these methods.

In the early 1960's, the Ultraviolet Light Commission found that patients who received prophylactic antibiotics had a higher incidence of wound infections than those who did not receive antibiotics (251). Since then, numerous studies of this technology's efficacy and safety have been conducted. Stone found, however, that "there are approximately 50 poorly founded and retrospectively reviewed 'testimonials' for every one controlled and statistically significant study" (326).

Clinical studies have serious problems themselves. Everett, for example, found no change in the incidence of wound infection when he used only neomycin (114). Because of the bacteria that neomycin inhibits, however, it could be expected to be only partially effective. Studies headed by Rosenberg (291) and Sellwood (307) with partially effective antibiotics found a significant decrease in the rate of infection, but the rate of infection in their controls was so high that their results must be viewed with caution. Barker used a combination of antibiotics now believed to be an ineffective dose (22). Nichol's study found no wound infections in patients given oral antibiotics, but his group was small and he did not use a double-blind experimental design (258).

The most rigorous study was Washington's, a prospective, randomized, double-blind study (409). A single surgeon performed the surgery in the study. The study found that a rational combination of oral antibiotics does reduce the rate of postoperative wound infections. Moreover, the treated group did not have serious postoperative complications because of the use of antibiotics.

The prophylactic use of antibiotics, in certain combinations and under controlled conditions, has thus been shown by one study to be efficacious and safe. The question of efficacy and safety, however, is rarely "settled for all time. " There still has been no widely accepted demonstration that systematic use of antibiotics prevents the complications

of elective colon surgery (397). Washington's study needs to be replicated and various combinations of antibiotics and methods of administration need testing.

FDA certifies antibiotics for both prophylaxis and for treatment. Thus, any approved antibiotic can be used prophylactically. To guide future decisions, the Veterans Administration (VA) is beginning a study to compare oral antibiotics with those given by injection. Use of antibiotics is covered under all Government medical care programs, including Medicare and Medicaid, and providers are reimbursed for using prophylactic antibiotics in bowel surgery. Reimbursement policies have not changed over the years, despite questions about such use.

This case study illustrates a technology whose use has been based not on testing but on surmise. After one study raised questions about the usefulness of such prophylactic antibiotics, however, a number of clinical trials were carried out, some with Federal support. Most of these trials have been inconclusive, because of methodological problems. One recent study allows a tentative conclusion that prophylactic antibiotics are useful in colon surgery. However, the many variables involved in the situation lessen the impact of any single study and also make complete assessment very difficult.

Case 6: Skull X-Ray

X-ray of the skull is a standard diagnostic procedure widely used in the United States for a variety of conditions. Approximately 17 million skull films were taken in this country in 1970, in the course of about 4.2 million skull examinations (362) (each skull examination includes multiple skull X-rays). In 1977, an estimated 5.7 million skull examinations were carried out. * The major corrective treatment for abnormal conditions within the skull is surgery; about 70,000 intracranial operations are done per year. * * The validity and reliability of skull X-rays have been studied extensively, but, according to Weinstein, Alfidi, and Duchesneau, their use produces an "extremely low yield of meaningful information that will contribute to the potential diagnosis or alter the course of therapy" (414).

Skull X-ray is used widely (in conjunction with physical examination, history, etc.) as a screening tool, especially in case trauma to the head, to determine if injury has taken place. An estimated 20 to 30 percent (0.8 million to 1.3 million skull examinations) of these examinations each year are done to evaluate head injury (28). Bell and Loop studied the use of skull X-rays for trauma by two hospitals in 1969 and 1970. They reported that 93 fractures were found in 1,500 skull examinations, or 1 in every 16 skull series. They found that the physician's evaluation of the patient was relatively accurate, especially with more severe injuries. Furthermore, only 28 of the 93 patients with skull fracture (30 percent) had therapy altered because of the demonstrated fracture; in those cases, skull fracture led either to prophylactic antibiotics or to surgery. Bell and Loop stated that "unsuspected fractures may be associated with less trauma and less disability, and perhaps seldom need to be demonstrated. " They also found that 20 percent of examinations were done for "trivial injury" and that another 34 percent were done to protect against possible malpractice suits (28).

Other findings also suggest excessive use of skull X-rays. Lusted and his coworkers found that about 16 percent of skull X-rays were ordered even when the physician reported certainty about the diagnosis (220). Jergens, Morgan, and McElroy studied a large emergency room and found the same situation as reported by Bell and Loop. Less

*Information furnished by the Bureau of Radiological Health, HEW.

• ● Information furnished by the National Center for Health Statistics, HEW.

than 1 percent of skull X-rays were positive and 19 percent were ordered for medico-legal reasons. They also noted that many examinations were done at the request or the demand of the patient.

Skull X-rays have little direct impact on therapy because the underlying brain damage, not fracture, is the critical variable for treatment—and brain damage does not appear in X-rays. Roberts and Shopfner state, "physicians can instruct patients and lawyers that head trauma causes injury to all cranial structures, including the brain, blood vessels, bone, and scalp, but that bone fracture almost never has any bearing on the patient's need for treatment and hospitalization" (288).

The apparently limited benefit from skull X-ray also needs to be weighed against the risk of exposure of a large population to radiation. One skull X-ray causes about 330 milliroentgens exposure; with an average of four X-ray exposures per skull examination or series, the average exposure to the individual is about 1.3 roentgens (362). Although no specific risk can be assigned this amount of radiation (2), risks of radiation should be minimized whenever possible.

The costs of skull X-rays also need to be considered. In 1970, when a skull series cost about $30, the aggregate cost was about $120 million (28). By 1977 the cost for a skull series has risen only to $39, yet the aggregate cost was $221 million. * Weinstein, Alfide, and Duchesneau comment: "We do not wish to imply that all skull roentgenograms are contraindicated. However, millions of dollars could be saved annually if skull roentgenograms were obtained only when indicated" (414). Bell and Loop developed a list of indications for skull X-ray in trauma and found that 29 percent of all those given skull X-rays did not meet any of their criteria (28).

The Federal Government is not supporting any clinical trials on skull X-rays. However, the FDA's Bureau of Radiological Health has supported the development of criteria for appropriate use of skull X-ray. Phillips (274) developed such criteria, based on the work of Bell and Loop (28), in 1973. Beginning in 1975, the high-yield criteria were applied in the emergency room of the University of Washington Hospital to all cases of head trauma. Despite a compliance rate of only 55 percent by physicians, the number of skull examinations for trauma decreased 39 percent from the previous year.

This result was encouraging enough that the Bureau of Radiological Health has supported an extension of use of the criteria to 5,000 patients in Washington State, working through the Washington State Professional Standards Review Organization (PSRO) program. If the project is successful, it might be extended to all PSRO programs. The National PSRO office is following the experiment with great interest.

The X-ray machines used for skull X-rays are regulated by the Bureau of Radiological Health to minimize population exposure to ionizing radiation. Skull X-rays as ordered by a physician are provided under all Federal programs for medical care and reimbursement. PSROs do not generally review skull X-rays.

In summary, skull X-ray is a technology with recognizable risks and a large financial cost. Whether the technology can be regarded as efficacious depends on the level of diagnostic efficacy at which it is being evaluated (see chapter 2). For example, is it efficacious in terms of accurate diagnoses? Its effect on diagnosis and patient outcome appears to be limited; thus, it is of low efficacy by those criteria. This case, therefore, points out the importance of specifying which level of diagnostic efficacy is being used in evaluating the usefulness of a diagnostic technology. Careful studies of indications for use

*Information furnished by the Bureau of Radiological Health, HEW.

could improve the application of the technology. At present, skull X-ray appears to be overused. If this is the case, then aggressive policies to decrease such use, especially in trauma cases, could decrease wasted expenditures and prevent unnecessary radiation exposure.

Case 7: Electronic Fetal Monitoring

Fetal monitoring is the continuous observation and recording of biological variables considered to be reliable indicators of a fetus' condition. In practice, fetal monitoring is done during labor, and has traditionally involved monitoring of the fetal heart rate by a nurse using a stethoscope (auscultation). An electronic device for fetal monitoring is a recent innovation, and its use has been growing. In "indirect" (267) or "noninvasive" (336) monitoring, the fetal heart rate and uterine contractions are monitored by sensors placed on the woman's abdomen. In "direct" or "invasive" monitoring, an EKG (electrocardiogram) electrode is attached to the head of the fetus through the vagina (336). In direct monitoring, a small needle is often inserted into the fetal scalp to sample fetal blood (336). In addition a catheter is usually passed into the uterus to obtain information about the frequency, duration, and intensity of uterine contractions (63).

The rationale behind fetal monitoring, in general, and electronic monitoring, specifically, is that the condition of the fetus can deteriorate rapidly during labor (56). So-called "fetal distress" can lead to mental retardation or even death. About 7,500 infants annually die during labor in the United States (63). Another 44,000 individuals are born mentally retarded each year (281). If fetal distress is discovered by changes in the fetal heart rate or in the acid-base balance of fetal blood, Cesarean section might save the life of the fetus or prevent brain damage.

Obstetricians and neonatologists (those specializing in medicine concerned with the newborn) believe that electronic fetal monitoring (EFM) is markedly superior to monitoring done with a stethoscope (56,169,182,281,327). Some propose its use in all deliveries. Several experts have suggested that monitoring by stethoscope is essentially useless (63). A report from the Pan American Health Organization states that "the appraisal of fetal condition by cardiac auscultation and palpation of the uterus is a less accurate, not continuous, time-consuming, and fatiguing method. In the majority of cases it does not enable the early detection of fetal distress" (56).

The belief of obstetricians in the efficacy of EFM is largely based on falling newborn mortality rates in institutions where EFM has been introduced (27,110,168,173). A typical experience is that reported by Quilligan and Paul, who found that the neonatal death rate at their institution fell after introduction of electronic monitoring (281). However, other changes were occurring in obstetrical practice besides EFM during the same period (209,269), and important changes were taking place in the general health of pregnant women. Better nutrition has been provided to pregnant women, and widespread contraception and abortion have changed the age at, and conditions under, which many give birth, leading to a better outcome (145,262,338). Wennberg also analyzed this question by examining hospitals in Vermont (420). He found a 30-percent decline in neonatal mortality rates from 1969 to 1974 at university hospitals where EFM had come into use—and a similar decrease in death rates in other hospitals in Vermont without any changes in obstetrical practices.

A few reports from institutions have analyzed their results by birthweight. Several investigators have found that low-birthweight newborns account for more than half of neonatal mortality (168,269). When results are analyzed by birthweight, much of the

change in perinatal mortality is in this group. Beard (1975) found a striking decline in neonatal deaths in premature infants with electronic monitoring. Wennberg also examined the factor of birthweights in Vermont, and found that perinatal death rates have fallen markedly in newborns under 2,500 grams over the past decade, while the death rates in normal-sized newborns have remained unchanged. The implication is that modern obstetrics has been of value to the small-birthweight infant, but of little benefit to the normal-sized infant.

The question of efficacy of EFM could be studied by controlled clinical trials, which would report on the results of long-term followup of children born with and without fetal monitoring. The three clinical trials that have been done so far have looked only at short-term outcomes, all in high-risk women. Two trials carried out in Denver found no benefit when EFM was compared to nurse monitoring (159,160). Efficacy of monitoring was measured by infant outcome on a variety of measures, including neonatal death and neonatal nursery morbidity. In fact, outcomes were the same in the two groups, but it was observed that EFM intruded on the process of birth and that it depersonalized care. The flashing lights of the monitors adjacent to the bed and the sound of each fetal heartbeat disturbed the mothers. Haverkamp also noted that "very close physical contact with the patient was necessary for the nurse to auscultate fetal heart tones adequately. This was not true to the same degree with the monitored group. Nursing attention to the gravida (pregnant woman) with respect to maternal comfort, emotional support, and 'laying on of hands' could have a significant impact on the fetus. " However, the trial excluded low-birthweight babies. A similar trial, from Australia, found substantial benefit but included low-birthweight infants (283). All three trials had methodological problems, particularly, the failure to use a research design that would minimize the influence of investigator bias on the results. The findings of the three trials are also consistent with a small degree of benefit.

However, Neutra and his coworkers used the. data from several years' experience at a large hospital to develop a statistical model of monitoring, and did find a modest benefit. In their model, monitoring 27 percent of labors with demonstrable risk factors would avert 80 percent of the potentially preventable neonatal deaths. Thus, clinical trials have not demonstrated clinical benefit, but clinical experience does suggest benefit for low-birthweight infants. It has also been claimed that monitoring prevents fetal brain damage (281), but there is no evidence of such benefit (145).

Electronic monitoring has its risks. Scalp abscesses and lacerations of the fetal scalp and perforations of the uterus can occur (63,238,267). Uterine infection can occur from the catheter (135,208). Also, practices associated with use of fetal monitors may induce the very fetal distress they are meant to detect. * Before an internal monitor is inserted, the amnionic sac must be ruptured, which may cause abnormally strong contractions that increase fetal stress. Effective use of the external monitor, on the other hand, requires that the woman remain still, which may have the effect of prolonging labor (studies show that frequent changes in position, and upright positions, speed labor). Furthermore, if a woman lies on her back to avoid disrupting an external monitor, the weight of the fetus in the uterus may constrict circulation in the aorta and vena cava and cause depression of the fetus, and maternal blood pressure, or both ("vena cava syndrome").

The most important risk to both mother and child, however, is Cesarean section and its risks. The Cesarean section rate has risen in the United States from 5.5 percent of deliveries in 1965 to 12.5 percent in 1976. There seems little question that this rise is

*Information furnished by the Food and Drug Administration.

associated with electronic monitoring. Many institutions report higher Cesarean section rates in monitored than unmonitored patients or increased Cesarean sections after introduction of EFM (69,130,131,269,314,327,420). In the first Denver controlled clinical trial (160), the Cesarean section rate was 6.6 percent in the nurse-monitored group and 16.5 in the EFM-monitored group. The rise seems to be associated with an increase in the diagnosis of fetal distress that follows monitoring, although other changes in obstetrical practice also contribute somewhat to the rise (11,185,344).

The use of monitoring has been increasing rapidly. By the end of 1972, an estimated 1,000 fetal monitoring systems were in use in the United States (267). It is probable that all obstetrical services soon will have monitoring capability and that more than half of the approximately 3 million deliveries a year could be monitored electronically.

Sales of monitoring equipment reached $25 million in 1976 and may reach $40 million (in today's dollars) by 1986 (219). Estimates of the added cost per delivery of EFM range from $35 to $50 (281) to $75 (157). Thus, if electronic monitoring were used in every delivery, it could cost society $200 million or more.

Delivery by Cesarean section increases the cost of delivery from $700 to $3,000 (157). Thus, if half of the increased number of Cesareans are attributable to normal fetal stress that is interpreted as fetal distress, $175 million has been added to the national health bill from Cesarean section associated with use of electronic fetal monitoring. This estimate does not include the cost of death and morbidity of mother and child from monitoring and Cesarean section.

Legal issues complicate the use of electronic fetal monitoring. The risks raise the possibility of malpractice suits. On the other hand, with strong professional support for electronic monitoring, physicians who do not use it may also face malpractice suits.

NICHHD of NIH is funding one study of electronic fetal monitoring. The Office of Maternal and Child Health of HSA is supporting a study by Haverkamp and his coworkers comparing nurse monitoring, electronic monitoring, and electronic monitoring with fetal scalp sampling.

Electronic fetal monitors are regulated by FDA under the Medical Device Amendments of 1976. Several local and State governments have taken steps toward requiring that all hospitals with maternity care units provide electronic and biochemical fetal monitoring as well as trained personnel to carry out the monitoring. The Health Department of New York City has already made such a recommendation (29).

Electronic monitoring is usually covered under third-party reimbursement programs, including Medicaid. Other Federal programs also provide it. For example, the Office of Maternal and Child Health of HSA distributes formula grants to States to support maternal and child health clinics whose intensive care units for women and infants considered to be in "high-risk" categories provide electronic fetal monitoring.

In summary, although many believe that electronic fetal monitoring is useful, its relative efficacy and benefit have not been established. Two controlled studies indicate that monitoring by nurses may be equally efficacious and provide additional benefits; a third finds EFM to be of some relative benefit. Moreover, fetal monitoring may be associated with considerable risks and financial costs. It is a technology that may well have been diffused prematurely. It is an example of a technology for which guidelines on appropriate indications for use might be needed. Guidelines could suggest what types of patients and delivery situations would result in benefits exceeding the possible risks.

Case 8: Surgery for Coronary Artery Disease

Coronary artery disease is caused by narrowing and blocking of the arteries that supply blood to the heart. The blockage results from arteriosclerosis (hardening of the arteries). The most common manifestations of coronary artery disease are myocardial infarction (heart attack or coronary), angina pectoris (severe temporary chest pain), and sudden death.

Coronary heart disease is the number one cause of death in the United States. In 1975, it was responsible for 642,719 deaths. The same year an estimated 4,120,000 Americans reported a history of heart attack and/or angina pectoris. Arteriosclerotic heart disease was the most frequent condition diagnosed for patients at the time of discharge from hospitals in this country in 1968 (210).

For more than half a century, surgeons have believed that an efficacious surgical approach to coronary artery disease is possible. Prior to the modern bypass operation, five different operations were developed and advocated enthusiastically (279). Although all five operations were ultimately abandoned as of no value, initially they were alleged to be efficacious, with reports in the medical literature claiming "objective" evidence of benefit. These operations were accepted and diffused by many members of the medical profession on the basis of experiential evidence. Other physicians usually preferred careful medical management and sound advice on how to conduct one's life, with surgery as a second line of defense.

For example, in the 1950's a surgical operation called internal mammary artery ligation was widely advocated by a small number of surgeons for improving blood supply to the heart. In retrospect, this procedure has little scientific rationale. The mammary artery is tied surgically. Because this artery is near the heart, surgeons hoped that this action would force blood to flow through other arteries in the vicinity, including coronary arteries.

In 1958 and 1959, two randomized, controlled clinical trials were conducted by, respectively, Cobb and Diamond (51). Patients were assigned randomly to control or operative groups, and the control group was given a sham operation, * in which the internal mammary artery was surgically exposed, but was not ligated. Both groups of patients reported relief from anginal pain and increased tolerance of exercise. As a result of these trials, the operation was largely abandoned. That both groups benefited suggests a strong placebo effect in the treatment of angina.

The experience with prior surgical operations for coronary artery disease points out that: (1) initial enthusiasm for, or belief in, an operation, based on current medical concepts, did not assure or predict results; (2) experiential evidence (anecdotal) led physicians to the false conclusions that the operations were successful; (3) with the exception of the internal mammary artery ligation operation, no truly objective (scientific) assessments of efficacy were made; (4) the operations were diffused without prior testing of efficacy or evaluation of safety; and (5) physicians reported dramatic relief of symptoms (angina) for all operations, demonstrating that a double-blind study is often necessary for evaluation of symptomatic response to technological intervention.

Coronary bypass surgery was introduced in the early 1970's. In this procedure, a graft is put on the coronary artery to bypass the constricted portion of the artery. This procedure has become the primary surgical approach to treatment of coronary artery dis-

*Sham surgery in a clinical trial would most likely not be possible today because of ethical considerations.

ease (51). Approximately 25,000 operations were performed in 1973 and at least 70,000 in 1977. Yet the benefits of coronary bypass surgery have not been clearly demonstrated. Claims that the operation prevents death remain largely unproven (73). Nonetheless, one proponent was quoted as saying that the United States should prepare to do 80,000 coronary arteriograms a day to screen for coronary disease. Coronary arteriogram is a special X-ray examination of the coronary arteries that is used to gather information useful in deciding whether to perform the bypass surgery. Such a widespread diagnostic program would itself cost more than $10 billion (162).

Coronary bypass surgery seems to give excellent symptomatic relief from angina pectoris. It is reported that 70 percent of patients evaluated 1 to 60 months after surgery are initially completely relieved of angina (210), but the improvement diminishes with time. However, the placebo effect mentioned above needs to be kept in mind because: 1) the initial results are similar to previous operations; 2) nonsurgical treatment also produces good results; and 3) the methods of evaluation of symptomatic relief are experiential.

Several important clinical trials of coronary bypass surgery have been conducted. From 1970 to 1974, VA conducted a randomized prospective cooperative trial that compared the efficacy of medical to surgical therapy for patients with stable angina pectoris (398). Of the 1,015 patients in this study, 113 were found to have a significant narrowing to the left main coronary artery. On followup of this group, those treated by surgery had a better survival rate. In the main study group, however, there was no difference in survival between medically and surgically treated patients. Surgery appears to have little effect on mortality except in a small group of patients.

The National Heart, Lung, and Blood Institute (NHLBI) is sponsoring two trials of coronary artery surgery. One compares medical to surgical therapy for patients with unstable angina. To date, the mortality rate is low and comparable for both groups, but the surgically treated group has had an incidence rate of myocardial infarction higher than that of the medically treated group. The second trial will resemble the VA study. Its results are not yet available. Three other randomized controlled trials in this country show no difference between surgical and nonsurgical groups (197,235,306).

Many advocates, convinced of the efficacy of the surgery, have declined to participate in clinical trials. The same advocates argue that the results of clinical trials may not be valid because some of the most skillful surgeons have declined to participate in the trials (200).

The risks of coronary bypass surgery are similar to those of any major surgery. The hospital mortality rate for patients undergoing such surgery is reported between 0.3 and 8 percent, with a usual range of 1 to 4 percent. However, only good results are published, generally, and the operative mortality rate derived from a large number of hospitals providing comparable data was 4 percent in 1976. * Other complications include myocardial infarction during surgery, in about 7 percent of patients (210).

The total cost of a coronary bypass procedure averages $15,000, so that aggregate costs in 1977 were more than $1 billion. Most of this amount was paid by third parties. Medicare and Medicaid programs reimburse for such surgery when considered by a physician to be medically necessary, On a per capita basis, Health Maintenance Organizations (HMOs) use the operation at less than one-half the national rate, and in Western Europe the rate is about 7 percent of the rate in the United States.

*Source: Commission on Professional and Hospital Activities, Ann Arbor, Mich.

Coronary artery bypass surgery is based on a scientific rationale and may be of measurable benefit to some patients. It is usually performed for angina pectoris and appears to give substantial relief from symptoms, but the extent to which this relief is an effect of surgery is not known. Limited studies suggest that coronary bypass surgery improves life expectancy significantly for only a small number of patients, with a particular type of coronary artery disease. Controlled studies have shown no improvement in life expectancy for patients studied.

Case 9: Tonsillectomy

Tonsillectomy is surgical removal of the tonsils, small bodies of lymphoid tissue in the throat. Tonsillectomy is the third most common operative procedure performed in hospitals in the United States. Approximately 884,000 tonsillectomies were performed in 1973 (374), and about 680,000 in 1976. Removal of the tonsils is by far the most frequent surgical procedure performed in hospitals for patients under the age of 15.

Tonsillectomy has been done throughout recorded history, with attempts at removal dating at least as far back **as** 600 B.C. (265). Before **antibiotics, it was** probably medicine's only weapon against serious complications of throat infections (tonsillitis). After 1900, refinement of surgical technique encouraged its wide application. The popularity of tonsillectomy peaked in the 1930's, and its use has gradually declined since then.

Despite its long history, tonsillectomy has not been well evaluated for efficacy. The inadequate design of published studies makes credible conclusions about its relative benefits impossible (37). Paradise has summarized problems of experimental design (264):

1. The selection of patients for surgery was not random.
2. Severity of tonsillitis varies within and between operated and control groups.
3. Indications for surgery were not stringent, so that many children with mild or no disease were subjected to operation.
4. Because of ethical considerations, children who appeared to the investigators most in need of surgery were excluded from studies and given the operation.
5. Postoperative evaluation was based not on direct examination of the children but only on information obtained from parents.

In part because of the lack of experimental knowledge, the attitudes of pediatricians and surgeons toward tonsillectomy vary greatly (264,328). Some believe it to be a useless procedure and routinely refuse to perform or recommend it. Others, impressed by cases of children whom tonsillectomy appears to have helped dramatically, continue to recommend it. Paradise, et al., have stated, "Differences among authorities aside, a history of recurrent throat infection remains the indication for tonsillectomy most commonly advanced by parents and invoked by physicians, and constitutes a principal criterion in current quality-of-care standards for the reasonableness of tonsillectomy" (266). Tonsillectomy is uniquely indicated when the tonsils are large enough to obstruct breathing or swallowing. Even accepting these indications for tonsillectomy, a significant number of physicians believe that many unnecessary tonsillectomies are performed (264,328).

It has been estimated that 30 to 40 deaths a year result from tonsillectomy (434). Other estimates run as high as 300 deaths per year. * Postoperative hemorrhage, either

*Personal communication, J. Paradise, M.D.

immediate or delayed, can contribute to the morbidity attributable to tonsillectomy. Psychological risks, although difficult to document, certainly exist. Some speculate that serious problems such as Hodgkin's disease can result years after tonsillectomy (265), but no long-term ill effects have been demonstrated convincingly.

The rates for tonsillectomy vary considerably. For example, one study found that rates of tonsillectomy varied from 20 per thousand to 5.6 per thousand depending on area of the country (348). Tonsillectomy is covered by most if not all third-party insurance plans, including Medicaid. A study of 22 States, encompassing more than 6 million Medicaid eligibles, showed markedly different rates of tonsillectomy by area of the country, varying from a high of 1,709 per 100,000 people in Nevada and 1,324 per 100,000 in Maine, to a low of 179 per 100,000 in Arkansas (348). The total cost of tonsillectomy in the United States is estimated at up to $500 million per year (434).

NIH funded a controlled clinical trial of tonsillectomy and adenoidectomy at the Children's Hospital of Pittsburgh in 1973. The Pittsburgh group has made a concerted effort to define carefully the group that would be admitted to surgery and to ensure that this group did in fact have repeated episodes of tonsillitis. A preliminary report from the study shows the importance of doing so, since most patients with histories of recurrent infections that were not well documented proved to develop relatively few episodes when followed closely (266). For patients actually admitted to the randomized clinical trial, careful followup of both the operated and control groups is being done. In March 1978, NIH funded the study for 3 more years.

In 1974, NIH sponsored a Workshop on Tonsillectomy and Adenoidectomy. Its participants concluded that a nationwide, collaborative, controlled clinical trial of tonsillectomy was indicated, modeled after the Pittsburgh study. More recently, NIH funded a group to assess the feasibility of such a multicenter trial. The findings of that group were presented at the July 1978 meeting of the Ad Hoc Advisory Panel on Tonsillectomy and Adenoidectomy. The panel did not reach unanimous agreement with the group's recommendation to go ahead with the multicenter trial.

In summary, tonsillectomy is a surgical procedure that has long held a place in medical practice, but its efficacy and indications for use are inadequately understood. Reliable and valid data are not available, and the practicing community has reached no consensus on its value. Available evidence seems to indicate that many unjustified tonsillectomies are performed, especially in some areas of the country. The major well-controlled study currently in progress in Pittsburgh may provide better data on the efficacy of tonsillectomy and its indications. However, developing better information is only the first step. After that, the cooperation of the practicing medical community will be necessary to bring medical care more in line with the new information.

Case 10: Appendectomy*

Appendectomy is surgical removal of the appendix, a small tubular extension of the intestine ordinarily located in the lower extension of the intestine in the lower portion of the abdomen. It is usually performed as treatment for appendicitis, inflammation of the appendix. Without treatment, some inflamed appendices perforate and release bacteria into the abdominal cavity. Such perforation can cause peritonitis, a generalized infection of the abdominal cavity that can threaten life.

*This case is adapted from material prepared for OTA by Richard Watkins, M. D., a member of the advisory panel for the study.

In 1973, approximately 350,000 appendectomies were performed in the United States (374), and 1,060 deaths from appendicitis were reported (436). Although physicians and the public believe in the efficacy of appendectomy (35,52), no controlled clinical trials have been carried out. A study in China of 955 cases of appendicitis treated without surgery reported two deaths (417,1). One trial of nonsurgical treatment in the Western World reported 471 cases and one death (75). Although one cannot generalize from these trials because of their small size and other factors, the reported appendicitis death rates from the trials are lower than the 1973 U.S. death rate for appendicitis (436). The number of deaths attributable to appendectomy itself is not known. If the risk of death is estimated to be between 0.01 and 0.1 percent, deaths from appendectomy in the United States would be between 35 and 350 per year.

Examination of the mortality rate from appendicitis over time raises questions about the effectiveness of appendectomy. Appendectomy was widely adopted after 1900. Appendectomies were performed at rates of about 400 appendectomies per 100,000 population in 1920, about 600 in 1930, and 800 in 1938 (80). The reported appendicitis death rate rose from about 10 deaths per 100,000 population in 1900 to 13 in 1920 and 15 in the early 1930's (80,211,430). Increasing mortality over the early decades of appendectomy has also been noted for Australia (120) and the United Kingdom (44).

In the 1930's and 1940's other therapies for appendicitis came into use, notably intravenous fluids, relief of abdominal distension by a tube passed into the stomach, and antibiotics. Several writers have attributed the subsequent fall in rates of mortality to those innovations (120,337). Mortality began to decline from its high of 15 per 100,000 in the mid-1930's to 10 deaths per ,100,000 in 1940 (75), two deaths per 100,000 in 1950 (389), and one death per 100,000 in 1960 (389). The appendectomy rate also fell from about 700 per 100,000 in 1940 (80) to 200 per 100,000 in 1965 (374).

The beneficial effects of antibiotics and other technologies might have obscured any effect of surgery on mortality in the 1940's and 1950's. Assuming that appendectomy generally prevents death, rates of death and rates of appendectomy should be inversely correlated (255). Both rates, however, have continued to drop, the mortality rate falling to 0.9 deaths per 100,000 in 1965 (390) and 0.5 deaths per 100,000 in 1973 (374), and the appendectomy rate falling to 160 appendectomies per 100,000 population in 1973 (374,436).

The rates of appendectomy for regional U.S. populations for 1965-73 vary from 100 to 620 per 100,000 (100,214,421,422). Rates among Federal employees using different health care systems contrast sharply. In 1968, Federal employees who received medical care from 14 prepaid group practice plans underwent appendectomy at the rate of 110 per 100,000, while Federal employees enrolled in Blue Shield underwent appendectomy at the rate of 210 per 100,000 (272).

The Group Health Cooperative of Puget Sound, a large prepaid group practice with an age and sex composition similar to that of the United States as a whole, had an appendectomy rate of 105 per 100,000 population from 1970 to 1976, and an appendicitis mortality for the same period of 0.24 deaths per 100,000 population. These rates may be compared to an appendectomy rate of 160 per 100,000 and 0.5 deaths from appendicitis per 100,000 for the United States as a whole in 1973 (374,436). Group Health Cooperative physicians tend to observe the patient when the diagnosis of appendicitis is dubious (411). Possible, mild inflammation of the appendix subsides during observation, and surgery is avoided. Recently a group of surgeons at Johns Hopkins University found that observation in dubious cases reduced their overall appendectomy rate by almost one-third without an increase in perforation (423). The use of more discriminating criteria for appendectomy appears likely.

The cost of appendectomies in the United States is estimated at more than $350 million annually (28). Much of this cost is covered by third-party payers, both public and private. Appendectomy is a standard benefit of almost all health insurance programs, including Medicare and Medicaid.

Thus, appendectomy is a costly technology with the standard risks associated with surgery. The relative benefits and risks of treating appendicitis through surgery or other treatment have not been fully evaluated. For example, there is strong evidence suggesting that appendicitis may be treated with substantially fewer appendectomies without increased loss of life. Thus, a controlled clinical trial of the nonsurgical or delayed-surgical approach to treatment of certain categories of patients with evidence of appendicitis might be warranted.

Case 11: Hysterectomy

Hysterectomy is surgical removal of the uterus. It can be performed by either gynecological or general surgeons; indeed, legally, by any physician. The National Center for Health Statistics (NCHS) estimates that 678,000 hysterectomies were performed in the United States in 1976. At a rate of 622.2 hysterectomies per 100,000 females per year, this major operation is performed at a higher rate than any other. If such a rate continued into the future, more than half of U.S. females would have had their uteruses removed by age 65 (49). Moreover, the rate increased approximately 25 percent from 1965 to 1976 (348). In the late 1960's the hysterectomy rate in the United States was more than twice as high that of England and Wales (50).

These facts helped lead to allegations that hysterectomies are carried out unnecessarily in many patients. However, there is no clear-cut definition of what is necessary; nor are the indications known for those hysterectomies that were performed.

Hysterectomy is performed for a variety of conditions, including premalignant states and localized cancers (see case 1), descent or prolapse of the uterus, and obstetric catastrophes, including bleeding and septic abortion. Recently, indications for the operation seem to have been broadened beyond those traditionally accepted. Functional problems and conception control have become common indications. Cole argues that the differences in national rates and the increase in the rate of hysterectomy in the United States are a result of "prophylaxis, " that is, to prevent later cancer or pregnancy. The reasoning is "based on the rationale that if a woman is 30 or 40 years old and has an organ that is disease-prone and of little or no further use, it might as well be removed" (77).

Hysterectomy has risks. Cole and Berlin estimate a mortality rate of 0.06 percent, or 600 deaths per 1 million women operated on (78). Operative morbidity, although difficult to quantify, also exists. About 30 percent of women have postoperative fever and 15 percent require transfusions, which introduce some risk of hepatitis. Other potentially important health losses are less obvious. Hysterectomy appears to affect ovarian function, even when the ovaries are left intact. It has been postulated that if estrogen (female hormone) levels are affected by hysterectomy, higher rates of coronary artery disease could result (78). Even a l-percent increase in death rates from coronary disease would offset any possible gain from preventing cancer (77). The psychological response to hysterectomy may be another major problem. Several studies have found psychiatric disturbance, including severe depression, in women after hysterectomy. Despite methodological problems, these studies seem to indicate a significant amount of disturbance. Notman believes it may be difficult for a woman to adjust to the loss of reproductive poten-

tial, but emphasizes the need for well-controlled studies of the emotional consequences of hysterectomy (259).

Cole has analyzed the benefits that could be derived from carrying out hysterectomies on 1 million women at age 35. Assuming a conservative 600 deaths from the operations, the million women would overall have a slightly longer life expectancy as a result of surgery. Only the 1.3 percent of women who would have died from cancer of the cervix and uterus would benefit, with an average of 14.3 years of life each (77). These calculations assume a constant rate of occurrences of cancer of the cervix and uterus.

In economic terms, Cole estimated that 1 million hysterectomies would cost $2.9 billion, and would result in savings of $1.4 billion, including 35,000 cases of cancer. He concludes on the basis of his analysis that the benefits of prophylactic hysterectomy are not worth the costs (77).

Other benefits are more difficult to assess, such as the value of hysterectomy for contraception, reduction of the fear of cancer, or the elimination of unpredictable bleeding. There are no data on how many women believe hysterectomy either improved or lowered the quality of life. Even if such data were available, however, decisions about routine hysterectomy would be difficult to make. Bunker and Brown studied physicians' wives on the assumption that they would be knowledgeable consumers of medical care and found a higher rate of hysterectomy in this group than in the general population (52).

Despite these questions, the Office of Technology Assessment (OTA) has been unable to identify any clinical trial of hysterectomy underway in this country. Hysterectomy is accepted as a standard surgical procedure and reimbursed by both Medicare and Medicaid. Rates of hysterectomy vary in the United States and are associated with such factors as geographic location and type of insurance coverage. In the Medicaid program, for example, the annual rate of hysterectomy among 6,609,684 eligibles in 22 States was 303 per 100,000 population, with a range from a low of 34 per 100,000 in Mississippi to a high of 2,488 in Nevada and 1,277 in North Carolina (348).

In summary, hysterectomy is a surgical procedure that is efficacious for some conditions. But some consider it to be overused. It illustrates the difficulty of determining indications for use and of defining desirable outcomes and expected risks. Physicians and consumers appear to consider the procedure valuable. Even with the best studies, it will be difficult to make decisions concerning hysterectomy and its use (including whether Federal reimbursement programs should pay for surgery for contraceptive purposes) on fully objective bases.

Case 12: Drug Treatment for Hypertension

Hypertension, or high blood pressure, is the most common chronic disease in the United States (232). The heart generates pressure as it pumps blood to all parts of the body. Average resting blood pressure is about 120 mm of mercury systolic and 80 mm of mercury diastolic; that is, 120/80. For largely unknown reasons, this pressure can become elevated. People with high blood pressure are more likely to have strokes, heart disease, and kidney failure than people with normal blood pressure.

NHLBI estimates that 54 million people have blood pressures of 140/90 or above and require further evaluation and monitoring. At least 26 million persons have blood pressures of at least 160/95, and many of these might profit from drug therapy. At least 6.1 million persons have diastolic blood pressure above 105 mm, and all of these require drug therapy (405).

Hypertension can be effectively treated. In the late 1960's, VA carried out a multi-institutional controlled clinical trial of treatment of males for high blood pressure with the drugs hydrochlorothiazide, reserpine, and hydralazine. The control group, which was randomly selected, was given placebos. The treatment was demonstrated to be remarkably effective for men with diastolic blood pressures above 105 mm mercury. Strokes, for example, were reduced by a ratio of 4 to 1, and congestive heart failure, renal failure, and dissecting aneurysm occurred only in the control group (399). Benefits were not as clear for those with diastolic blood pressure levels below 105 mm. VA carried out an additional pilot study to collect more data on male patients with mild hypertension. NHLBI is also sponsoring further trials of men and women with all levels of hypertension, including diastolic pressure less than 105 mm mercury.

The side effects of the treatment, although seldom dangerous, are annoying. They may include dizziness, impotence, and general malaise. VA investigators state that these side effects can be minimized by careful prescription and the monitoring of treatment. Long-term use of the drugs may have side effects that are not known (60), although many of the drugs have already been in use for years.

Other questions remain unanswered. The VA study involved only relatively young male patients: does it apply equally to females; does it apply to those over age 65; what about those individuals with blood pressures under 105 mm diastolic? (126)

Furthermore, diagnosing hypertension is not easy. Validity and reliability of the measurements can be questioned for various reasons, including both systematic and random errors in reading the pressure of patients (261). Transient elevations of blood pressure are common, and care must be taken to ensure that the patient actually has hypertension (285). Many instruments for automatically determining blood pressure have been marketed; often they have not been adequately tested in the field (261).

Data obtained from national surveys based on probability samples from the early 1960's and the early 1970's indicated little change in the status of hypertension control. Approximately half of those persons with hypertension were unaware that they had elevated blood pressure and only about one-seventh had their condition adequately controlled. The VA study has led to major attempts to change this situation. NHLBI has data, collected in 1973 and 1974 from 14 communities, showing that 29 percent of hypertensives were unaware of their condition, 23 percent were aware but not undergoing therapy, 19 percent were aware but on inadequate therapy, and 29 percent were both aware and on adequate therapy. Although these data are not comparable to the national survey data, they are encouraging. In addition, patient visits for hypertension have increased dramatically in recent years (405).

The number of untreated individuals underscores the problem of "compliance," or convincing patients to take the medication. A person with hypertension must take the drugs throughout life, despite the absence of symptoms. Side effects, financial cost, and lack of explanation from physicians are some reasons that patients who feel well may not want to take prescribed drugs.

The cost of treating the entire population with diastolic blood pressures of 105 mm or greater (and a few below this level) is estimated by NHLBI at about $4.5 billion to $5 billion annually. The total cost that would be incurred if these hypertensives (those with the disease) were not treated cannot be estimated, but all cardiovascular disease, to which hypertension is a major contributor, costs society about $40 billion to $50 billion annually. Cost-benefit calculations carried out by NHLBI suggest that every dollar invested in controlling hypertension returns a benefit to society of $1.25 (405).

The Federal Government is significantly involved in the hypertension problem. FDA regulates the devices to diagnose hypertension and the drugs used to treat it. VA and NIH are sponsoring clinical trials aimed at improving knowledge. NHLBI coordinates a National High Blood Pressure Education Program, for both professionals and the public. NHLBI has also used hypertension as an example for building consensus (see chapter 5) and produced recommendations for the optimal diagnosis and treatment of hypertension for the practicing physician (285). VA has a nationwide program of screening patients for possible therapy, the Department of Defense (DOD) provides screening and therapy, and Medicare and Medicaid reimburse for treatment for hypertension, except that Medicare does not cover drugs for outpatients. Despite these efforts, a large number of patients with severe hypertension remains inadequately treated. Hypertensives are found especially in low-income groups, and blacks constitute a disproportionately large number of the individuals not being adequately treated (96).

In summary, drug treatment for hypertension has been subjected to a well-designed study for efficacy. On balance, such treatment is clearly indicated for approximately 6.1 million citizens with diastolic pressures above 104 mm mercury. It may be indicated, depending on the individual situation, for a significant portion of the estimated 20 million additional persons with blood pressures at or above 160/95, Calculations indicate that such treatment would probably be cost-beneficial. Nonetheless, despite considerable Federal activity and good efficacy and safety information, many affected individuals are not adequately treated.

Case 13: Drug Treatment for Otitis Media in Children*

Otitis media is the technical term for infection of the middle ear, a small cavity connecting the throat and the sinuses behind the ear that is necessary for effective hearing. Otitis media is believed to begin when bacteria enter the middle ear from the throat. Multiplication of these bacteria attracts white blood cells into the cavity, forming pus. The pus may burst through the eardrum and extend into the sinuses behind the ear or into the skull. Fluid can also collect in the middle ear and decrease hearing. If this fluid and the attendant loss of hearing persist, children can suffer delayed language development and impaired learning.

Ear infections are common in children. In a prospective study of 246 infants, approximately one-third were found to have ear infections at least once during the first year of life. Nineteen (8 percent) had two infections in the first year, and 4 percent had three or more infections in the first year (167). By the age of 6, 76 to 95 percent of children have had at least one ear infection. About 20 to 26 percent of children will have experienced six or more episodes by that age (172).

A variety of treatments is used for otitis media. Antibiotics are usually prescribed. Frequently, a medication for pain and a decongestant or an antihistamine are also suggested. Occasionally, a myringotomy, a simple surgical operation in which the eardrum is cut to release pus from the middle ear, is done. In about 40 percent of children, fluid persists after recovery from the acute infection (317). In these cases, antihistamines and decongestants are often prescribed and tubes are sometimes placed in the middle ear cavity through an eardrum form for draining.

Although antibiotics are accepted as efficacious therapy for ear infections, they have not been fully evaluated. They came into widespread use without careful testing about 20

*This case is adapted from material prepared for OTA by Philip Brunell, M. D., a member of the advisory panel for the study.

years ago. Controlled clinical trials to demonstrate the general efficacy of antibiotics for acute infection have been done only recently (127,317). Howie and coworkers carried out a controlled clinical trial in which the control group was given a placebo. Persistence of the middle ear infection occurred in all 45 cases of otitis caused by Pneumococcus and in 12 of 21 cases due to *Haemophilus* influenza when treated with a placebo; the most effective antibiotics cured more than 95 percent of similarly studied patients (172).

Antibiotics are also used prophylactically in children with recurrent otitis media. When Perrin, et al., tested sulfonamides in a group of children up to the age of 8 they found that prophylactic sulfonamides reduced the rate of otitis media by 7 times, with little morbidity. While sulfonamides are cheaper than most other antibiotics that might be used for prophylaxis, their nondiscriminate widespread use could be expensive for the medical care system.

The role of antibiotics in preventing the complications of otitis media is not known. Though it is difficult to find data showing a reduction in pyogenic (from pus) complications (317), most authorities agree that antibiotic therapy has decreased the incidence of acute mastoiditis, chronic eardrum perforation, and chronic mastoiditis.

The few trials of widely used decongestants and antihistamines have not shown these drugs to be effective in preventing serious otitis media (207).

FDA regulates all the drugs used for safety and efficacy. Government and private health insurance programs that include coverage for children routinely cover antibiotic treatment for otitis media as a benefit, and sometimes cover the other drugs as well. Special programs have been established for population groups with high rates of complications from otitis media, such as American Indians.

In summary, antibiotics are universally used in otitis media. After years of use, controlled clinical trials confirmed their efficacy. It appears, however, that clinical experience was adequate to demonstrate efficacy in this case, and one may question the ethics of using a placebo in studying treatments for this disease. A controlled clinical trial of prophylactic use of sulfonamides demonstrated efficacy, yet more expensive antibiotics are often prescribed. Other drugs, especially decongestants and antihistamines suggested by physicians and readily obtained over the counter in pharmacies, have no demonstrated efficacy.

Case 14: Cast Application for Forearm Fracture

Some bones, such as those in the forearm, are often fractured. Usually, the broken ends of the bone stay close to each other and, if immobilized, will heal in a period of weeks. If the ends are not close together, they are forcibly adjusted, often under anesthesia. Surgical “open” reduction with fixation by pins or other materials is also often used, despite the risk of infection or delayed healing. Experience indicates that without support during the healing process bones may not heal properly (32,156,427).

Through the centuries, various methods have been used to provide the necessary support for the bone. Ancient Egyptians, for example, used stiffened linen in a splint. The use of gypsum (plaster of paris) was first reported in 1798. Early attempts were plagued with complications such as pressure sores and gangrene caused by tight casting, stiff joints and wasting of the muscles. Techniques improved and by 1918 Bohler had developed methods still largely in use today (246).

Cast application for forearm fracture is a common procedure in medical practice. More than 1 million patient visits to office-based physicians in 1973 were for forearm

fracture, according to data from the National Ambulatory Medical Care Survey (373). Forearm fracture is the most common fracture in that study. Cast application has not been subjected to a controlled clinical trial. It is generally accepted as quite efficacious without such evaluation.

Alternatives to cast application exist, however. Traditional Chinese medicine uses different techniques. Instead of being forcibly reduced or alined, the bone ends are gradually brought into alinement, day by day. Bamboo splints are used and replaced every day. Movement of the limb begins as soon as satisfactory reduction is achieved. Horn has noted strengths and weaknesses of this method, especially its lack of complications, and described how modern and traditional methods are being merged in China (170).

Plaster of paris cast materials are regulated by FDA as medical devices. No federally supported research on cast application seems to be underway. All Government medical care programs and medical care reimbursement programs include cast application for forearm fracture as a benefit. Estimates for the annual cost of this procedure are not available.

In summary, cast application for forearm fracture is a technology whose efficacy has been established by experience in medical settings. It illustrates a technology whose efficacy could be called "manifest," that is, whose efficacy and safety are obvious to the observer. Although alternatives to cast application might be as efficacious, its widespread acceptance in this country makes development and testing of other methods unlikely and probably unnecessary.

Case 15: Treatment of Hodgkin's Disease

Hodgkin's disease, the most common neoplasm of young adults in the United States, is a form of cancer that primarily affects the lymphatic system. In 1977 there were an estimated 7,400 new cases of, and 2,900 deaths from, this disease (8).

Treatment of Hodgkin's disease primarily consists of two methods: supervoltage X-ray radiation and a four-drug combination treatment (vincristine, procarbazine, prednisone, and nitrogen mustard) known as MOPP (89). Supervoltage X-ray treatment is used for early and more localized stages of the disease and MOPP treatment for more advanced stages, although combinations of the two treatments are sometimes used.

The 3-year survival rate for patients with Hodgkin's disease increased from 35 percent in 1940-46 to 61 percent in 1965-69. From 1969 to 1973, the 5-year survival rate reached a level of 87 percent (8). The improvement resulted from new understanding of the pathology and natural history of the disease as well as development of the treatment.

In diagnosing Hodgkin's disease, pathologists classify the disease according to the predominating type of abnormal cell growth (histologic type). Laboratory tests and diagnostic X-rays are then used to determine whether the disease is confined to one lymph node region or has spread to other parts of the body. Such tests for extent of disease are called "staging. " The development of histologic: and staging criteria allowed patients to be grouped into relatively homogeneous populations according to the type and extent of disease. Knowledge of both the histologic class and the clinical stage of the disease are essential for planning the most appropriate treatment (106). Because such knowledge also permits the conduct of controlled clinical trials that are methodologically sound, the safety and efficacy of various treatments can be compared and evaluated.

Study of supervoltage X-ray treatment began in the 1930's. Controlled clinical trials of this technology have shown that 50 percent of patients with early stages of the disease

may now survive 15 years or more (107,188,273). When more extensive radiotherapy is used for limited disease, 90 percent are alive after 10 years, and most have no evidence of disease 4 or more years after treatment (205,329).

The four-drug combination treatment was developed at NCI, and its efficacy has been studied in controlled clinical trials. After completion of this treatment, 80 percent of patients with advanced Hodgkin's disease survive 5 years or more, and 47 percent remain completely free of disease (101).

Current trials are comparing new treatments and combinations with established treatments rather than with placebos or with no treatment. Controlled clinical trials are now being funded by NIH to demonstrate whether combined X-ray and drug therapy offer better results than either method alone. Other clinical trials are examining the long-term results of existing treatments (161,247,401).

In addition to evaluating the efficacy of these treatments, clinical trials provide a careful evaluation of risks. Each treatment has risks that can themselves be lethal, such as overwhelming infection (99), bone marrow suppression, pericarditis, and pneumonitis (273). A second malignancy may develop as a result of either radiotherapy or chemotherapy. In fact, recent evidence suggests that the incidence of second malignancies may be far higher in those patients receiving both radiotherapy and chemotherapy. This higher incidence may increase the risks of the therapy relative to the benefits (252). Compared to the possible benefits of a normal life span, however, these risks are considered acceptable (3).

FDA regulates the chemotherapeutic agents used in Hodgkin's disease, and FDA's Bureau of Radiological Health regulates the X-ray equipment used in treatment. In addition, the cost of supervoltage X-ray machines is high enough to require that the institution purchasing one secure a certificate-of-need (CON) from the State health planning agency. Treatments for Hodgkin's disease have been covered by third-party payers, including Medicare and Medicaid, since they first became available. Demonstration of efficacy has thus had little, if any, effect on reimbursement. In fact, ongoing trials of drugs, which could be considered experimental, are largely funded by payments of third-party payers for health services.

In summary, the efficacy and safety of treatments for Hodgkin's disease have been well demonstrated by a series of well-designed clinical trials. Insurance funds for medical services have helped to finance testing of treatments for Hodgkin's disease. The case demonstrates that testing of efficacy and safety can depend on other technologies, such as staging techniques. Additionally, the case shows that efficacy is not absolute, but relative, and requires judgments as to benefits and risks.

Case 16: **Chemotherapy for Lung Cancer**

Chemotherapy for cancer involves introducing **a** chemical or hormonal agent into the body in order to disrupt or destroy cells. It is used most frequently when surgical removal of the cancer is impossible or unsuccessful. Between 1940 and 1950, only one-third of patients diagnosed as having lung cancer were treated. From 1960 to 1970, 75 percent were treated (97). Four treatments for lung cancer have been developed: chemotherapy, irradiation (X-ray therapy), surgery, and immunotherapy. These therapies are used both individually and in combination.

Because at least 80 percent of lung cancer is caused by cigarette smoking, it is largely a preventable disease. It is nonetheless the most common form of fatal cancer in the United States, ranking first among males and fifth among females. ACS estimated that

89,000 deaths would occur from this disease in 1977 and that 98,000 new cases would be detected, a rate 14 times higher than that of 40 years ago (8). Despite the high percentage of patients who are treated, the overall 5-year survival rate for lung cancer (8 percent of males and 10 percent for females) did not change between 1950 and 1970 (97).

Multiple clinical trials of chemotherapy have led to three general conclusions about its efficacy in treating lung cancer:

1. The rate of survival of patients treated with chemotherapy for certain types of lung cancer limited to one side of the chest is similar to that of patients treated with radiotherapy and increasingly better than that of placebo-treated patients (213,432). The average increase in longevity from chemotherapy ranges from 2 to 15 months (30,58,213);
2. For extensive lung cancer, certain types of chemotherapy increase survival approximately 2 months over a placebo-treated group (30,74); and
3. The effects of chemotherapy used in combination with other therapy are unclear (58).

Durant and his coworkers compared irradiation, chemotherapy, and their combination in treating all types of inoperable lung cancer clinically confined to the chest. They found no significant difference in mean survival among the three groups. More important, they found no evidence that immediate treatment at the time of diagnosis improved either survival or quality of life when compared to the initiation of treatment when symptoms appeared, Although the study was not double-blind, it does raise important questions concerning the treatment of lung cancer patients without symptoms, especially in view of the complications of the treatment (106).

Recent evidence, however, indicates some improvements in results. According to information furnished by NCI, 20 percent of patients with oat cell carcinoma (a form of lung cancer) limited to the thorax now survive 2 years when treated with combination chemotherapy. NCI further reports that 30 to 40 percent of patients with limited non-oat cell carcinomas have increased survival periods of 14 to 15 months, up from the former median survival of 6 months.

The risks of chemotherapy are considerable and may increase in combination treatments. Many agents affect the bone marrow by lowering the number of white blood cells and thus leaving the subject liable to serious infection and even death. Another common complication is nausea or loss of appetite, with resultant weight loss and poor physical condition. Hospitalization, which affects quality of life and adds to financial costs, is often necessary during therapy.

Both methodological and ethical issues have confounded the execution of valid and reliable clinical trials. The definition of “inoperable lung cancer” has varied from study to study. Outcome measures are difficult to define. The most frequent measures have been patient survival rates and decreasing tumor size. Patients with lung cancer, however, die from other causes, and interpretation of tumor size is complicated by noncancerous disease conditions, such as infection and emphysema (74,416). These problems are further complicated by the fact that many trials compare one chemotherapeutic agent with another, rather than with **a** placebo.

Ethical problems arise in conducting such trials. If a study begins to demonstrate less improvement or greater deterioration in the treatment group than in the control or alternate treatment group, the researcher may feel ethically obligated to stop the trial.

The estimated cost for the drug for treating one patient is from $50 to $150. Approximately 60,000 new inoperable patients were treated for lung cancer with chemotherapy in 1977. Such chemotherapy is covered under most third-party reimbursement programs, including Medicare and Medicaid. Because third-party payers fund testing of chemotherapeutic agents as cancer therapy, such trials are among the least expensive at NIH.

NCI is supporting several trials of chemotherapy for lung cancer, as is VA. Chemotherapeutic agents used for lung cancer are regulated and approved for investigational use by FDA.

In summary, chemotherapy for lung cancer has been extensively studied for efficacy and safety. Efficacy is very limited. Drugs and hormones are inherently risky. Costs are high. Methodological and ethical problems plague studies in this area. Current chemotherapy for lung cancer may be a technology being diffused inappropriately.

Case 17: Hyperbaric Oxygen Treatment for Cognitive Deficits in the Elderly*

Surveys have shown that 10.0 percent of those over 65 years of age display mild to moderate cerebral dysfunction and that 4.4 percent in that age group are seriously demented, or approximately 2.2 million Americans in the first category and about 900,000 in the latter. Life expectancy is reduced to about a third of normal for the majority of seriously demented patients. The impact of mild to moderate cerebral dysfunction is more difficult to evaluate but must be highly significant in economic, social, and personal terms.

Consequently, considerable excitement was generated in both the scientific and general community when an article appeared in 1969 in the New *England Journal of Medicine* reporting enhanced cognitive functioning in elderly, male, organic brain syndrome patients following repeated exposure to pure oxygen, under pressure, in a hyperbaric chamber (1). Up to that time there was no known effective treatment for memory loss associated with brain changes due to arteriosclerotic disease or Alzheimer's disease. This finding by Jacobs and her associates (1) was even more compelling as five control subjects exposed to an air mixture failed to show improvement initially, but did improve later when they were crossed over to oxygen.

Five published reports confirmed Jacobs' observation (2,3,6,8,9). However, only one of these studies utilized a control group. Two studies failed to replicate the original Jacobs findings (10,11). One of these used 21 experimental subjects and four control subjects (11). These authors failed to note any significant differences between the experimental and control subjects.

Thus one of the major problems in evaluating the efficacy of hyperbaric oxygen as a treatment for cognitive impairment in the elderly was the paucity of studies that employed control subjects and the small number of control subjects in those that did. One reason for investigators' reluctance to include control subjects is that the control condition is more dangerous than the experimental condition. Experimental subjects breathe pure oxygen, but control subjects breathe an air mixture containing nitrogen, with some danger of the bends if care is not taken with decompression times.

Because of the importance of the Jacobs results and the obvious need for a replication study with enough control subjects to provide an adequate test of the efficacy of hyperbaric oxygen, a collaborative study was undertaken, in 1973, between the Psycho-

*This case is adapted from material prepared for OTA by the Alcohol, Drug Abuse, and Mental Health Administration.

pharmacology Research Branch of the National Institute of Mental Health (NIMH) and the New York University Medical Center.

Subjects in the study were 40 ambulatory individuals between 60 and 85 years of age residing in the community who had documented evidence of significant memory loss. There were approximately equal numbers of male and female subjects; circulatory disturbances were cited as the possible cause of organic brain syndrome in half the cases and senile brain disease was noted for the other patients.

Simply put, the results of this study failed to sustain the view that oxygen administered under pressure improves cognitive functioning in the elderly. Efforts were also made to identify subgroups of patients for whom oxygen may be especially efficacious. Again, there was no evidence of differential treatment effects as a function of initial severity of illness, sex, or presumed evidence of cerebrovascular disease. Subjects who entered this study had well-documented evidence of memory problems but were still sufficiently intact to reside in the community and to respond meaningfully to an intelligence test and to other psychological and psychometric tests. On the basis of the findings of Jacobs et al. (1) and others (2,3,6,8), one would have expected many of these patients to show a favorable response to hyperbaric oxygen treatment. The study findings clearly indicated this was not the case.

For a variety of reasons early dissemination of these negative findings was deemed in the public interest. The Jacobs findings had been picked up by the news media, especially the more sensational press, and hyperbaric oxygen was widely touted as a cure for a variety of the infirmities of old age, in addition to memory loss. A number of hyperbaric centers in this country were offering hyperbaric oxygen as a treatment for memory loss in the elderly at substantial fees. For example, at one center the fee was $5,000 for 15 days of treatment. This was not an easy issue to resolve, as scientific findings are generally not widely disseminated prior to publication in a respected scientific journal, where lag time between receipt of a manuscript and publication generally runs a year or more. To offset this delay, it was decided to present these findings at a meeting of the American Geriatric Society and to release a statement to the press once word was received that the paper had been accepted for publication (12).

Although publication of the study findings and dissemination of the results through the press and television have not completely eliminated the practice of offering this treatment to the public, it did appear to significantly dampen enthusiasm; a number of hyperbaric centers have since stopped offering this treatment. The study findings also appear to have had some impact on health insurance carriers and on the Social Security Medicare program, which at one time had considered paying for this treatment. The insurance carriers and Medicare have since ruled that hyperbaric oxygen is not a medically accepted or effective treatment for cognitive deficits in the elderly, and they will not pay for it.

The case points out the importance of appropriate dissemination of scientific findings. Information that promises relief to suffering individuals may be disseminated quickly and extensively—perhaps exceedingly so—if testing has been inadequate. It is critical that subsequent, contradictory (but more valid) findings be given the widest and most rapid dissemination.

4.

ESTIMATING EFFICACY AND SAFETY

4.

ESTIMATING EFFICACY AND SAFETY

Techniques used for estimating efficacy and safety range from the informal methods of individual physicians to randomized clinical trials with complex methodological designs. No technique is universally applicable for every medical technology. In many instances less complex methods may be more appropriate than the more sophisticated approaches. Frequently, combinations of techniques are used. This chapter describes five techniques used in evaluating safety and efficacy: preclinical, informal, epidemiological and statistical, controlled clinical trials, and formal consensus development.

Various laws have been enacted to regulate the efficacy or safety of drugs and medical devices since the passage of the Federal Pure Food and Drugs Act in 1906. Surgical and other procedures that depend primarily on providers' techniques have not been subject to similar controls. Rather, responsibility for assessing the efficacy and safety of these procedures is contained within the profession (125,332,334).

Assessments of efficacy and safety for "products" (drugs and devices) usually differ from assessments of medical and surgical procedures in terms of the source of evaluation and the kinds of techniques applied. The physical nature of products implies a highly consistent formulation that may be unattainable in surgical technique evaluation. Also, investigators can learn much about products before they are tested clinically (394). Many procedures, however, heavily rely on testing for their development.

PRECLINICAL

Many medical technologies are evaluated in biochemical and animal tests prior to human experimentation. These preclinical tests maybe part of the developmental effort, or a requirement for Federal or private approval, or both. The required tests may be of two types: 1) preliminary evidence to gain the right to test with humans (364), and 2) performance standard compliance to establish marketability.

Chemical analyses for purity, quantity, and quality of the active agents are typically undertaken. Other filler and stabilizing substances are evaluated for potential pharmacological activity.

Animal testing provides a guide to potential therapeutic activity as well as capacity to induce toxicity (85). Determining the degree of toxicity, or safety, is the major function of animal studies. A prime factor analyzed in safety tests is the level of median lethal dosage. Toxic effects are evaluated in terms of chemical and physiological analysis. Therapeutic effects may be measured in terms of bioavailability (transport across gastrointestinal membranes) and pharmacokinetics (distribution throughout the body).

The accuracy of animal models in determining the probable effects of drugs on people is a controversial issue. In particular, carcinogenic agent evaluation in animals is a

very complex, multifaceted problem. Questions that arise in these evaluations include short-term high dose versus long-term low dose, animal species selection, population size, and controls (191). Despite some of the inherent problems in utilizing animals, the report by the Office of Technology Assessment (OTA), *Cancer Testing Technology and Saccharin (353),* concludes that they are acceptable models for cancer studies and probably should be regarded as reasonable precursors to clinical studies.

Medical devices are evaluated by chemical and physical laboratory testing in addition to animal studies. Physical testing may seek to determine mechanical strength, material properties, and electrical performance. General manufacturing techniques, such as quality control, precision machining, and sterility, may also be evaluated. Chemical tests using culture or hematologic techniques may determine biocompatibility. Other chemical tests evaluate long-term dissolution in body fluids and the possible presence of toxic residues in the production of plastic materials. Implantable devices also are subjected to complete preclinical animal testing.

INFORMAL

Despite the increasing need to formally estimate the efficacy and safety of medical technologies, the majority of such evaluations are still based on informal approaches. White (426) estimated that 80 to 90 percent of all procedures have been evaluated by informal methods. These informal assessments of medical technologies may take place during medical school and specialty training and through personal peer experience.

Physicians and other health care personnel are constantly exposed to medical technologies throughout medical school, residency, and special courses. Students generally assume that these technologies are efficacious and safe, Technologies recommended to the student have undergone formal statistical studies or professional consensus exercises. However, it is more likely that the suggested uses of technology are based on previous experiences or training received by the instructor.

Personal experience is perhaps the oldest and most common informal method of judging the efficacy and safety of a medical technology. This technique is dominated by qualitative impressions. The control groups are primarily envisioned as experiencing the end result that would occur if there were no clinical intervention (85). Despite its limited statistical value, this technique does have some advantages compared to the more rigorous methods used in certain situations. For example, personal knowledge of the patient may promote beneficial adjustments to the type and level of treatment. Also, many rare side effects are reported in letters to the editor columns by individual physicians (85). Perhaps more importantly, personal experience is the primary method that determines whether or not a medical technology is adopted into widespread practice (79,187).

Peer experience is more explicit than personal experience; information may be exchanged by personal communication, journal articles, pamphlets, and the like. Again, there is little control over the scientific quality of these technical assessments. However, this peer interaction is the core concept of the more formal group consensus discussed later.

It is important to point out that many medical advancements have properly and successfully proceeded without rigorous statistical methodology of evaluation. For example, vitamin B12 treatment for pernicious anemia clearly is justified. Cast application for

forearm fracture (see chapter 3, case 14) is a technique whose efficacy has been established experimentally in medical settings. Alternatives such as bamboo splints exist (170); however, the widespread acceptance and success of casting makes evaluation of other methods unlikely and probably unnecessary. An earlier OTA report, *Development of Medical Technology: Opportunities for Assessment, * (354),* made two points that summarize the utility of informal methods: 1) "despite complexity, and cost, some procedures are so effective in restoring function that few would question their social utility, " and 2) ". . for a disease for which the natural history is fairly well known and the benefits of a new technology are dramatic, alternative methods of evaluation (as compared to controlled clinical trials) may be appropriate. "

Informal techniques are based on the clinical approach of qualitative, artful decisions as compared to the scientific approach of quantitative, mathematical decisions. Ingelfinger, et al. (178) point out the critical issue of statistically significant findings versus clinically significant results. Other sources (24) describe further causes both for separating the informal from the rigorous technique and developing new methodologies to improve medical decisions.

Three concepts summarize the necessity of both the informal and the rigorous techniques for assessing efficacy and safety. First, each extreme may be appropriate in certain situations. Second, many assessments require various combinations of techniques. And third, cooperation between clinicians and statisticians must exist to attain appropriate decisions when more rigorous techniques are used.

EPIDEMIOLOGICAL AND STATISTICAL

Epidemiology is the study of the determinants and the distribution of diseases and injuries in human populations. The term also incorporates the study of the impact of medical interventions on diseases and injuries. Three types of epidemiological methods that are particularly useful in evaluating the efficacy and safety of certain medical technologies are described in this chapter. These three methods are: retrospective, prospective, and controlled clinical trials. The last type of study warrants discussion in a separate section from the other two because of its importance and prevalent use.

Retrospective studies compare groups of people who have a disease with those that do not. These studies are designed to determine whether the two populations differ in terms of percentage exposed to certain critical factors. In addition, attempts may be made to compare standard factors, such as age, sex and race, between the two groups. Data obtained from retrospective studies are summarized as an "odds" ratio** which is defined as the ratio of incidence rate among the exposed group to the incidence rate among those not exposed. Both the relationship between oral contraceptives and thromboembolism* * * and the positive correlations demonstrated between smoking and lung cancer were established by retrospective studies.

Most information used in retrospective studies is derived directly from the patients, their relatives and friends, and individuals' medical and other records. Consequently, the

---

*This report, released in August 1976, described the development and assessment of cardiac pacemakers for heartblock.

* ● The "odds" ratio is a close approximation of the relative risk.

●● *Users of oral contraceptives are four or five times more likely to develop thromboembolic disease than nonusers (81).

uniformity, accuracy, and completeness of information (especially on death certificates) are often in doubt. In addition to incomplete or biased data, the selection of appropriate comparison groups represents another major problem in this type of research.

Despite some inherent problems, general utility of retrospective studies has been frequently substantiated by other experiments in which there is more control (81). Even marketing and manufacturing data may provide critical links to unsafe technologies. Atomizers containing isoproterenol were linked to cardiac arrythmia deaths. Improper usages and overdoses due to poor quality control in manufacture were shown to be probable causes of death. Utility, low cost, and quick results are the major advantages of these studies (237).

Prospective studies follow the histories of persons both exposed and unexposed to a particular factor under study. The incidence of deleterious effect resulting from such exposure is then determined for persons in the two groups. If records of individuals exposed to a particular factor exist, then the study also may utilize past data; however, prototypic prospective studies deal with ongoing events (43). Statistical results from such studies include incidence rates in addition to relative risk.

A major advantage of prospective studies is the relatively clear designation and selection of both the study and the comparison groups by means of matching characteristics with minimum bias before the disease develops. Some of the disadvantages of these studies include their high cost and the possible occurrence of changes in patients and methods over the duration of the test (237).

The Boston Collaborative Drug Surveillance Program (244) is an example of a large study that assesses drug efficacy and safety by utilizing epidemiologic methods. * To date, approximately 12 percent of the drug exposures studied by this program have yielded unsatisfactory results. In addition, statistical techniques were useful in discovering and estimating the frequence of unsuspected adverse drug reactions. The Framington Heart Study, which has been in progress since 1948, has shown a clear correlation between high blood pressure and the occurrence of cardiovascular disease in adults also using epidemiologic methods (81). Currently, some epidemiologic methods are aimed at assessing the efficacy and safety of various antihypertensive treatments.

CONTROLLED CLINICAL TRIALS

All subjects who agree to participate in controlled clinical trials (or simply, randomized clinical trials) are assigned to experimental and control groups. Subjects in these trials are assigned randomly to either the experimental or control group. These trials, and their impartial test and control group establishment, are direct experimental extensions of prospective studies that have no control over the physician's choice of treatment. In a controlled clinical trial intended to assess efficacy and safety, the experimental group would be treated or diagnosed by the technology under examination; usually the control groups would be either treated by an established standard technology or given a placebo. However, in some cases, a standard technology is administered to one of the study groups while a second (control) group receives no treatment. Clinical tests and examina-

---

*The program was initially funded by the Pharmaceutical Manufacturer Association Foundation. Since 1967, it has been supported by a number of other organizations, including FDA and the National Institute of General Medical Sciences of NIH.

tions of the members of each group are used for evaluations of the relative benefits and risks of the technology.

Many controlled clinical trials require a long period of time and large commitments of money, resources, and subjects, The National Institutes of Health (NIH) estimated that the total amount of money* expended for trials underway in FY 1975 (new starts and continuing studies) was $641.8 million for 755 trials. * * Efficacy and safety research often requires money contributions from several sources. For example, it may be appropriate sometimes for the third-party payers to finance part of the evaluation of an established, presently reimbursable technology. In addition, the Food and Drug Administration (FDA) estimates that private drug firms spend $1 million to $4 million to bring a drug to market after it has been developed in the laboratory (406).

Many professionals who conduct research into the efficacy of medical technologies have focused attention on the randomized controlled clinical trial because critical assessments of the efficacy and safety of medical technologies require high-quality research (65). For example, Cochrane (72), Hill (163), and others strongly support the use of the randomized clinical trial in evaluating efficacy or safety. Conversely, others (133) suggest that nonrandom, less well-controlled trials and statistical manipulation of available data can provide results that are as useful as randomized clinical trials.

Randomized controlled trials are the most useful when: 1) the benefit of a new technology is uncertain (e. g., amniocentesis, see chapter 3, case 2), and 2) the relative benefits of existing therapies are disputed (55) (e.g., tonsillectomy, see chapter 3, case 9). There is much statistical theory that supports the scientific utility of such randomization procedures in clinical trials. Byar, et al. (55) discussed three major advantages to randomization. First, and most familiar, bias may be eliminated from the assignment of treatment. Often double-blind techniques are utilized in which neither the patient nor the physician knows the technology used on any specific individual. (However, in comparing drug to surgical treatments, bias may well occur because both the surgeon and the patient know which method is being utilized; and only lower risk patients may be candidates for the surgical operation.) Secondly, randomization prevents bias with respect to variables that exist in the experiment but are not directly considered in the design. This allows comparisons between treatment groups. The third advantage of randomization is the validity of the statistical tests of significance that are used to compare treatments. It should be noted that complete randomization may be inappropriate under certain circumstances; in such cases modifications in the randomizing process may be used (151).

There are many areas of controversy surrounding the use of randomized clinical trials, perhaps the greatest of which is ethical (21). Arguments against randomization and other aspects of these trials are based on a concern for both patient and physician rights and responsibilities. Critiques of randomization include the following statements: physicians must make clinical judgments and act according to their consciences (431); personal physicians must influence whether their patients enter a trial and what treatment is administered; patients must be given the best possible information in consent forms (335); and, patients should be able to choose which treatment is delivered.

Critics of controlled trials or of some of the processes used in trials also point out that certain groups of patients have rights that are easily violated. Appropriate questions regarding the rights of children in particular are raised, For example, when can informed

---

● The total amount here refers to the entire cost of completing trials that were underway in FY 1975.

.● Trials supported by the NIH vary widely in costs. One of the most expensive, the Multiple Risk Factor Intervention Trial (MRFIT), is budgeted at $115.7 million.

consent be given by a child?; at what age?; with what medical conditions or illnesses?; and, who, if not the child, will guard those rights? In addition, the long-term effects of treatments or other medical technology interventions can be especially serious and very long in evidencing themselves in children. Clinical trial protocols must be established with all these and more questions in mind. Similar questions may occur regarding the rights of other groups composed of convicts, the aged, and the mentally retarded, for example.

Many articles defend the ethics of using controlled clinical trials. Byar, et al. (55) state that physicians cannot do just what they "believe" best, their practice must be based upon sound scientific evidence. Similarly, an honest acceptance of the fact that the relative benefits and risks of the best current therapy are not known is the first step in recognizing the need for clinical trials. If each patient is so unique as to be ineligible for statistical randomization, how can the individual physicians use clinical judgments based on past experience as the optimal guideline for determining the treatment of the next patient (55)? Mosteller (249) contends that the rights of patients are protected in their ability to refuse participation in the trial. In addition, proper diagnosis of a patient must precede a decision regarding trial participation. In some cases, patients (or physicians) may also choose to select a treatment but randomize on dosage level. This choice also provides the patient with more control. A final point in favor of randomization is the apparent improvement (although not perfection) of the statistics and planning of recent randomized clinical trials.

There are no unequivocable answers to these concerns. Certain technical improvements in statistical methods allow faster identification of intermediate results, thereby leading to sounder decisions regarding the termination date of certain types of trials. Improved consent mechanisms are being developed and could be applied more widely, Interestingly, many articles note serious complaints about randomization but still recommend cautious use of the technique (335,423).

FORMAL CONSENSUS DEVELOPMENT

The assessment of a specific medical technology may include one or more studies which use any or all of the techniques previously described. If the evidence clearly supports or rejects the relative utility of a treatment, then the analysis of efficacy and safety may be complete (though it may need periodic re-examination). In many cases, however, the evidence does not lead to such an unequivocable decision. Consequently, a consensus group may be formed both to evaluate all pertinent information, which may range from informal to detailed statistical studies, and to recommend its findings to the medical community.

There are two types of consensus groups relevant to this report which are discussed further in the next chapter. Briefly, one type of consensus group evaluates the current state of efficacy and safety knowledge regarding either a particular medical technology or technologies that relate to a specific medical condition. An example of this type of consensus development is the "technical consensus-building" effort of NIH. A second type of group both analyzes a medical technology, particularly devices, and recommends possible standards to be used in the conduct of future efficacy and safety assessments. This type of consensus process is used in the programs of the Association for the Advancement of Medical Instrumentation and the American Society for Testing and Materials.

5.

CURRENT ASSESSMENT ACTIVITIES

5.

CURRENT ASSESSMENT ACTIVITIES

This chapter describes Federal Government and private sector activities for assessing the efficacy and safety of medical technologies. It is not an evaluation of the performance of the agencies, except where such performance is affected by the presence or absence of policies relating to efficacy and safety.

Federal Government Activities

FOOD AND DRUG ADMINISTRATION

The Food and Drug Administration (FDA) of the Department of Health, Education, and Welfare (HEW) is one of the principal Federal regulatory agencies designed to protect the health of the American public. Over the past two decades, FDA's responsibilities in the protection of health have increased significantly. In 1970, there were only three product-oriented bureaus: foods, drugs, and veterinary medicine. Subsequently, the agency has taken on responsibilities encompassing a broad range of medical technologies, such as X-ray equipment and other radiation-emitting medical and consumer devices, blood banks, vaccines and allergenic, organ transplants, and other biological products.

The growth in the agency's jurisdiction has been accompanied by a concomitant increase in its budget and staff. Between 1954 and 1977, FDA's budget grew from $5.5 million to $250 million; the staff increased from less than 1,000 to 7,300 (123,240). FDA's FY 1977 budget represented approximately 4 percent of the Public Health Service budget and 40 percent of total Federal outlays for consumer protection.

The specific role FDA envisions for itself is regulating the transfer of medical technologies from the level of medical researcher to the level of health practitioner and consumer. The agency particularly emphasizes regulation in those areas where consumers cannot make reasonably informed judgments. These regulatory responsibilities give FDA one of the most direct Federal roles in assuring the efficacy and safety of two major classes of medical technologies: drugs and medical devices.

Prescription Drugs: Statutory Authority

FDA is responsible for implementing the Food, Drug, and Cosmetic Act of 1938. This Act mandates Federal regulation of all drugs. As the Act's principal enforcer, FDA is required to approve all new drugs before they are marketed. Such approval is contingent upon the demonstrated efficacy and safety of a new drug.

The requirement that the efficacy of a new drug be demonstrated before approval was added to the Act by amendment in 1962. Previously, the 1938 Act restricted FDA's

33-405 O - 78 - 6

review to the safety of drugs. Therefore, the fact that a particular drug was not shown to be efficacious could not, in most cases, serve as the basis for disapproval of its marketing application.

Two statements from the 1962 amendments form the basis for FDA's definition of efficacy. According to the legislation, the FDA Commissioner must refuse approval of a drug marketing application if, after notice and opportunity for hearing, he or she determines that "there is lack of substantial evidence that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the proposed labeling thereof. . ."* Substantial evidence is defined in the Act as: "evidence consisting of adequate and well-controlled investigations, including clinical investigations, . . . on the basis of which it could fairly and responsibly be concluded. . . that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof. " No distinction is made either in the Act or its implementing regulations between the terms efficacy and effectiveness.

Safety is assessed as a separate factor from efficacy. FDA must weigh the relative benefits and risks associated with the use of the drug, and the drug may enter the market only when the benefits derived from its use clearly outweigh the risks.

Prescription Drugs: Regulation

All drugs which are not already on the market, or generally recognized by experts as safe and effective under prescribed conditions of use, must undergo premarket review. This process begins with the submission of a "new drug application" (NDA) by the manufacturer to FDA. Minimally, NDAs must contain a full report both of the investigations conducted to determine a drug's efficacy and safety, and the methods, facilities, and controls used in its manufacture, processing, and packaging. In addition, labeling samples to be used for the drug must be included. FDA must either approve the application or notify the applicant of an opportunity for a hearing within 180 days after an application is filed. During the 180-day period, FDA attempts to determine if the therapeutic benefits of the new drug justify its potential risks.

FDA may provide exemptions from the NDA process to those intending to use the drug solely for investigational purposes. However, a "Notice of Claimed Investigational Exemption for a New Drug" (IND) must be filed by anyone planning to conduct research involving the use of new drugs by human beings. An IND must contain chemical, manufacturing, and control information, results of animal studies, and a description of the protocol for the clinical study, including information about the investigator and facilities for the study. If FDA does not prohibit the research within 30 days following the IND filing, the research may commence.

Medical Devices: Statutory Authority

FDA was first provided authority to regulate medical devices in the Food, Drug, and Cosmetic Act of 1938. This Act extended FDA control over foods and drugs and gave FDA new powers with regard to cosmetics and medical devices. Under the Act the FDA had to prove that a product was in fact dangerous or fraudulent before any action could be taken to remove the product from the market.

---

*These amendments also required drug firms to demonstrate the efficaciousness of all drugs marketed between 1938 and 1962.

The development and use of medical devices has expanded greatly since the passage of the 1938 Act. As a result of the dynamic growth of the industry, more complex, sophisticated, and technologically challenging products were being developed that had the potential to cause serious patient injury or even death. In response to some of the possible dangers inherent in such growth, Congress enacted the Medical Device Amendments of 1976, which bestowed FDA with significant new authority to ensure the safety and efficacy of medical devices. These amendments were enacted primarily to provide regulatory safeguards commensurate with the potential consumer risks associated with the use of increasingly sophisticated medical devices. Accordingly, they require evaluations of efficacy and safety to be made "weighing any probable benefit to health from use of the device against any risk of injury or illness from such use. " To achieve such evaluations, Congress both expanded FDA's operative definition of medical devices and required classification of all devices into one of three regulatory categories, differentiated according to the extent of control necessary to ensure their efficacy and safety.

As defined in the amendments, a medical device is any health care product that does not achieve any of its principal intended purposes either by chemical action within or on the body, or by being metabolized. Examples of devices included in this definition are optical prescription lenses and frames, hearing aids, intrauterine devices, surgical instruments, cardiac pacemakers, and CT scanners. The definition also applies to *in vitro* diagnostic products, including those that were previously defined and regulated as drugs. It is estimated that there are more than 8,000 products currently on the market that conform to the expanded definition.

FDA designed its system of classification and regulatory controls to prevent unnecessary regulation of device manufacturers while simultaneously providing maximum protection to consumers. Devices placed in the Class I category are subject only to general controls which include premarket notification, adherence to good manufacturing practices, and recordkeeping requirements. * *Class II* medical devices must meet FDA's performance standards which may relate to their construction, components, ingredients, and properties. A manufacturer must seek premarket approval of a medical device both when general controls would not ensure its safety and efficacy and when there is insufficient information available to develop performance standards. Premarket approval is the overriding regulatory requirement for Class III devices. Devices that are life sustaining, life supporting, or implanted into the body usually must be placed in the Class III category.

A manufacturer who develops a new device must notify FDA at least 90 days in advance of its placement on the market. During this 90-day period FDA determines the class in which the device belongs through regulations which require the manufacturer to supply: 1) proposed labels, advertising, and directions for use; 2) statements regarding the similarity to or difference from products already on the market; and 3) descriptions of how a device complies with existing standards regulations.

Medical Devices: Regulation

Implementation of the 1976 Amendments began with the assignment of devices into the three regulatory categories. Nineteen panels were created to recommend classifications for a total of 3,500 generic categories of devices. Each panel was composed of seven voting members who were professionals with training and experience in the clinical, scientific, and engineering aspects of devices. One industry and one consumer represen-

'General controls apply to devices in all three categories.

tative served as nonvoting members on each panel. The panel reports, which were recently released and contained over 7,000 pages of documentation, provide both the recommended classification for each device reviewed, and evaluations of each generic category of devices. Approximately 37, 59, and 4 percent of the devices were recommended for placement in Classes I, II, and III, respectively (360).

Regulations pertaining to Class I devices (section 520(f) of the amendments) direct FDA to "prescribe regulations requiring that the methods used in, and the facilities and controls used for, the manufacturing, packing, storage, and installation of a device conform to current good manufacturing practices (GMP). " FDA has published a proposed GMP regulation which applies to all devices, and is therefore known as the "umbrella" GMP. Failure to comply with the GMP renders a device "adulterated," and regulatory action can be initiated against the manufacturer.

Section 514 of the amendments authorized FDA to develop and promulgate performance standards for devices in *Class II*. Standards maybe developed by FDA or outside organizations. For example, contracts have been awarded to extramural organizations to develop standards for electrocardiographs (EKG), electromagnetic compatibility, electrosurgical devices, and infant incubators, among other devices. FDA may also adopt an existing standard as the mandatory one.

To review the adequacy of existing standards and to guide the development of new ones, FDA annually prepares a comprehensive list of current national and international standards activities for medical devices and diagnostic products. Criteria for such attributes as performance, sensitivity, accuracy, materials, safety, and durability are included in the standards listed. FDA works closely with voluntary standards organizations to review and possibly adopt some consensus standards already developed by these agencies.

As stated previously, devices classified into *Class III* are required to undergo a process of premarket approval. This process entails the submission of a premarket approval application. These applications are then referred to the appropriate classification panel for review and subsequent recommendation to FDA. The application must include, among other things, a summary presenting a sound case for approval and a review of all known data published that demonstrates the product's safety and efficacy. Following the panel's recommendation, FDA will approve or disapprove the application within 180 days of its receipt, unless a longer period of review time is agreed to by FDA and the applicant. Manufacturers of pre-enactment *Class III* devices (any device in commercial distribution before May 1979) have 30 months after final classification to develop data demonstrating the efficacy and safety of such devices before FDA can require the submission of a premarket approval application.

All testing of devices that involves the use of human subjects will be required to follow FDA's regulations governing the investigational use of devices, after these regulations are published in final form and become effective.

NATIONAL INSTITUTES OF HEALTH

The principal biomedical research agency within the Federal Government operating under HEW, is the National Institutes of Health (NIH). It was established in the immediate post-Second World War years both to consolidate the Government's medical research activities and to conduct, encourage, and support medical research and develop-

ment. NIH currently receives approximately two-thirds of all Federal dollars allocated to biomedical research, although more than a dozen other Federal agencies also conduct such research. NIH provided an estimated $2.24 billion in biomedical research support in 1977; this amount represents approximately 40 percent of all moneys expended for medical research in the United States during that year (260).

Biomedical research conducted by NIH includes studies of drugs, devices, and medical and surgical procedures. These studies usually are accomplished through grant and contract awards to academic and other research institutions. However, NIH generally does not synthesize the evidence regarding efficacy and safety gained from these studies.

Statutory Authority

Section 301 of the Public Health Service Act provides NIH with its basic research authority. This section of the Act authorizes the Surgeon General of the Public Health Service (the parent agency of NIH) to encourage and assist "research, investigations, experiments, demonstrations, and studies relating to the causes, diagnosis, treatment, control, and prevention of physical and mental diseases and impairments of man. " In addition to the general statutory authority provided by the Public Health Service Act, 8 of the 11 institutes comprising NIH have specific legislative mandates to fulfill particular research functions for certain categories of disease. For example, the National Cancer Institute (NCI) and the National Heart, Lung, and Blood Institute (NHLBI), the two largest components of NIH, are governed by statutes which include requirements to engage in demonstration and control programs relevant to those disease categories.

Specific references to efficacy or effectiveness do not appear in any of the NIH legislative authorities. However, NIH concern regarding the efficacy and safety of medical technologies can be assumed from the general language it uses to describe its mission: 1) advancing knowledge and understanding of the normal and pathological processes of the human body, and 2) developing ways in which the providers of medical care can safely and effectively intervene to prevent, treat, or cure diseases and disabilities.

Clinical Trial Support

Clinical trials provide the basis for the testing and orderly application of fundamental research knowledge prior to its general introduction into the health care system. These trials assist in preventing the premature introduction of new diagnostic and treatment hypotheses into general practice. Often, such trials are the only methods used for testing and evaluating the safety and efficacy of new diagnostic and treatment developments.

NIH investment in both the support and conduct of clinical trials has increased substantially in recent years. Four out of the eleven institutes* nearly tripled their total obligations for major clinical trials between 1971 and 1974. In FY 1975 alone, NIH provided approximately $110 million to support clinical trials; this figure represents 5 percent of the total NIH budget for FY 1975. Completion of these trials was estimated to cost another $345 million.

Tables 2, 3, and 4, on the following pages illustrate NIH support for clinical trials during FY 1975. Table 2 delineates clinical trial investment both by institute and by type

● The four institutes were the National Cancer Institute; the National Heart, Lung, and Blood Institute; the National Institute of Neurological and Communicative Disorders and Stroke; and the National Eye Institute.

Table 2.—National Institutes of Health 1975 Inventory of Clinical Trials

Amount of NIH Support for Clinical Trials Active in Fiscal Year 1975 by Institute and Type of Support

(in millions of dollars)

NIH Institute*	Extramural support: Grant	Contract **	Grant & contract	Total	Intramural support ***	Amount of support
NEI	$0.8	$2.1	—	$2.9	$0.2	$3.1
NHLBI	4.3	37.8	0.1	42.2	0.5	42,7
NIAID	1.3	1.5	—	2.8	0.3	3.1
NIAMDD	1.8	0.9	—	2.7	0.7	3.4
NCI	14.0	7.5	2.6	24.1	2.5	26.5
NICHHD	1.3	1.8	0.2	3.4	0.5	3.9
NIDR	0.8	0.2	—	1.1	0.7	1.7
NINCDS	0.8	0.6	—	1.4	2.0	3.4
NIGMS	0.1	—	—	0.1	—	0.1
Total	$25.1	$52.6	$2.9	$80.6	$7.3	$87.8

*Names of Institues: NEI = National Eye Institute; NHLBI = National Heart, Lung, and Blood Institute; NIAID = National Institute of Allergy and Infectious Disease; NIAMDD = National Institute of Arthritis, Metabolism, and Digestives Diseases; NCI = National Cancer Institute; NICHHD = National Institute of Child Health and Human Development; NIDR = National Institute of Dental Research; NINCDS = National Institute of Neurological and Communicative Disorders and Stroke; NIGMS = National Institute of General Medical Sciences.

**Contract includes interagency agreements without intramural support.

***Intramural support includes intramural support in combination with interagency agreements.

Table 3.—National Institutes of Health 1975 Inventory of Clinical Trials

Number of Clinical Trials Supported by NIH in Fiscal Year 1975 by Institute and Type of Support

NIH Institute	Number of trials supported extramurally: Grant	Contract *	Grant & contract	Total	Number of trials conducted intramurally**	Total number of trials
NEI	13	2	—	15	5	20
NHLBI	6	16	1	23	3	26
NIAID	78	21	—	99	10	109
NIAMDD	29	14	—	43	6	49
NCI	253	66	48	367	38	405
NICHHD	24	12	1	37	4	41
NIDR	25	6	—	31	13	44
NINCDS	13	8	—	21	38	59
NIGMS	2	·	—	2	—	2
Total	443	145	50	638	117	755

*Contract includes interagency agreements without intramural support.

**Intramural support includes intramural support in combination with interagency agreements.

Table 4.—National Institutes of Health 1975 Inventory of Clinical Trials

Number of and Amount of Support for NIH Supported Clinical Trials Active in Fiscal Year 1975 **by Institute and Type of Intervention**

(in millions of dollars)

NIH Institute	Total trials supported in FY 1975		Type of intervention					
			Therapeutic		Prophylactic		Diagnostic	
	Number	Amount	Number	Amount	Number	Amount	Number	Amount
NEI	20	$3.1	11	$2.4	—	—	9	$0.7
NHLBI . . .	26	42.7	16	27.0	9	$14.6	1	1.2
NIAID	109	3.1	53	1.2	29	1.5	27	0.5
NIAMDD .	49	3.4	45	3.3	2	0.0	2	0.0
NCI	405	26.5	372	23.6	12	1.3	21	1.6
NICHHD. .	41	3.9	11	0.6	21	2.7	9	0.6
NIDR	44	1.7	16	0.7	17	0.7	11	0.3
NINCDS. .	59	3.4	49	3.0	5	0.3	5	0.2
NIGMS. . .	2	0.1	2	0.1	—	—	—	—
Total. . .	755	$87.8	575	$61.8	95	$20.9	85	$5.1

of expenditure, Table 3 indicates the number of clinical trials conducted by each institute, As evidenced in this table, the average expenditure per trial ranged widely from $1.6 million for NHLBI to $28,000 for NIAID and NIGMS.

Table 4 outlines expenditures by three functions of technology: therapeutic, prophylactic, or diagnostic. Clinical trials investigating therapeutic technologies were predominant in 1975. Supplemental information provided by NIH indicates that a total of 535 trials were conducted to test drugs either in isolation or in combination with another type of technology. Four hundred of these trials tested drugs in isolation. More than **300** trials tested cancer chemotherapies; only **25** evaluated surgical procedures. Eighty-five trials examined such diagnostic technologies as CT scanning for brain tumors and fluorescent scanning in thyroid disease. However, few clinical trials examined the efficacy of screening or early diagnosis. Trials of primary prevention were quite rare.

NIH interest in conducting and disseminating the results of clinical trials continues to grow. For example, a summary of clinical trials under NIH support is assembled annually, which is divided according to type of trial and level of expenditure. In addition, the agency has established an NIH Clinical Trials Committee to coordinate work in the areas of design, taxonomy, and trial monitoring strategies. NIH held a major conference in the fall of 1977 for persons engaged in this type of research to impart information recently generated about clinical trial methodology.

Consensus Development

According to NIH, the present process for diffusion of medical technologies "leads to a situation in which the practicing community at large is not prepared to react promptly and in the best informed state to rapid advances in technology. . . . While the Food and Drug Administration has stringent requirements for the safety and efficacy of drugs, biologics, and devices, many procedures existing in current medical practice and new interventions entering the medical arena and adopted by practitioners are not amenable to such regulatory action and require more critical appraisal of effectiveness" **(382).** Although there have been many situations where a clinical trial has firmly established the efficacy and safety of a particular medical technology, there are other situations in which

the results of a clinical trial have been equivocal. Also, in some cases controlled trials may indicate that a technology is of limited benefit. Thus, the technology is efficacious but the value of this limited efficacy must be evaluated by other techniques besides the controlled trials. In some cases clinical trials may be prohibitively expensive. In other cases trials may pose difficult ethical and moral considerations. In such cases clinical experience can bean important factor in determining what use should be made of the technology.

Due to both the inherent limitations in clinical trials and the need for improved methods of disseminating research information, NIH initiated a process for developing a consensus among representative experts regarding the proper role of a given medical technology. NIH entitled that process "technical consensus development. " Representatives of various segments of the medical community are asked to agree on five issues: the clinical significance of the new findings; the adequacy of efforts to validate efficacy and safety; the need to identify cost, ethical, or other social impacts as points for caution; the need for feasibility demonstrations in community settings; and whether research results are phrased for easy understanding; and acceptance by health practitioners (381).

Hypertension was one of the first areas in which technical consensus development was applied (see chapter 3, case 12). Initially, NIH appointed a Committee on Detection, Evaluation, and Treatment of High Blood Pressure, which included individuals representing a wide range of professional groups, including the American Medical Association, the American College of Cardiology, the American College of Physicians, and the American Heart Association. This committee developed detailed recommendations on the management of hypertension, which included diagnostic procedures and a listing of effective therapies. Because the committee reflected such a broad base of interested parties, and therefore had great credibility, the recommendations were widely adopted.

A second major example of consensus development application at NIH was the 1977 Meeting on Breast Cancer Screening (see chapter 3, case 4). The meeting was held to coincide with the completion of a review of the Breast Cancer Detection Demonstration Project (BCDDP), which involved periodic screening of large numbers of women for breast cancer using clinical history, physical examination, mammography and thermography. Critics had questioned whether the use of radiation (by mammography) to detect cancers might not subsequently trigger development of malignancies.

NIH convened a 16-member panel composed of scientists, epidemiologists, and physicians from various disciplines, including radiology, medical oncology, surgery, and general medicine. Representatives of the clergy, legal profession, and lay public were also asked to participate on the panel.

Subsequent to the gathering of evidence, the panel developed 12 recommendations regarding the risks, benefits, and ethical considerations involved in the BCDDP, in particular, and screening, in general. The recommendations ranged from specific suggestions for determining which risk groups should continue to undergo periodic screening and the appropriate radiation dose, to general recommendations regarding the need for additional research in particular subject areas,

In January 1978, NIH established the position of Associate Director for Medical Applications of Research as a response to the success of the consensus development process. The Associate Director and staff work with individual institutes to increase awareness of each institute's activities in consensus development. Additionally, they coordinate consensus development efforts which involve a number of institutes simultaneously. This office has recently developed guidelines for methods to be utilized in: 1) the identification

of new knowledge pertinent to health care, 2) consensus development conferences, and 3) the dissemination of research information. Other technical consensus conferences are being planned by the Associate Director in conjunction with the other institutes. Table 5 lists conferences being coordinated during 1978 and 1979.

Table 5.—Consensus Development Conferences
National Institutes of Health

Time	Institute	Title	Format
May 1978	NCI	Medical Aspects of Asbestos	Conference
June 1978	NIEHS	International Cadmium Conference	Conference
June 1978	NIDR	Dental Implants	International conference
June 1978	NIH Nutrition Coordinating Committee	Nutrition in the Eighties	Panels and formal presentations; 2 days
June 1978	NCI	Mass Screening for Colo-Rectal Cancer	Cconference of European & American scientists
July 1978	NIA	Treatable Brain Diseases in the Elderly	2-Day meeting
July 1978	NINCDS	Indications for Tonsillectomy and Adenoidectomy	Advisory group
August or October 1978	NHLBI	Early Hospital Discharge of Patients with Uncomplicated Myocardial Infarction	Panel
September 1978	NIAID	Availability of Insect Sting Kits to Non physicians	Panel
September 1978	NICHHD	Antenatal Diagnosis	Panel
November 1978	NIGMS	Supportive Therapy in Burn Care	2-Day workshop
Summer 1978	DRS (BEIB)	The Use of Microprocessor-Based, 'Intelligent, ' Machines in Patient Care	Workshops (series of 4)
December 1978	NIAMDD	Intestinal Bypass Surgery in Treatment of Massive Obesity	Panel
1978	NIEHS	Standards for Laboratory Use of Toxic Substances Posing a Potential Risk	Panel
1978	NIA	Postmenopausal Estrogen Treatment	Preliminary planning meeting
1978	NCI	Mass Screening for Lung Cancer	Conference
1978	NCI	Rehabilitation for Cancer Patients	Conference
1978	NIEHS	Toxicological Evaluation of Hair Dyes	Workshop
1978	NCI	Health Education Workshop	Workshop
1978	NCI (ION, NIA)	Palliative Care of the Terminally Ill	Conference
1979	NHLBI	Prophylactic Use of Low Dose Heparin in the Prevention of Venous Thrombosis and Pulmonary Embolism	Panel
1979	NHLBI	Use of Extracorporeal Membrane Oxygenator (ECMO) in the Treatment of Adult Respiratory Failure	Public meeting
1979	NEI	Photocoagulation Therapy for Diabetic Retinopathy	Panel
1979	NIEHS	Validation of Short-Term Tests as Predictors of Carcinogenic and Mutagenic Activity	Conference
1979	Interagency Committee or New Therapies for Pain and Discomfort	Pain and Its Relief	Panels and formal presentations; 2 days

Source: Information furnished by staff of the Associate Director for Medical Applications of Research, NIH.

ALCOHOL, DRUG ABUSE, AND MENTAL HEALTH ADMINISTRATION

The Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA), another agency within HEW, incorporates programs of basic and applied research, service, and training, which are relevant to the understanding and treatment of mental illness, drug abuse, and alcoholism, in its three component institutes: the National Institute on Alcohol Abuse and Alcoholism (NIAAA), the National Institute on Drug Abuse (NIDA), and the National Institute of Mental Health (NIMH). ADAMHA has conducted research to establish the safety and efficacy of medical technologies since the 1950's. In 1975, however, ADAMHA established Treatment Assessment Research (TAR) as a separate research category, specifically designed to study the relative safety and efficacy of various substances and procedures applied to human subjects. This research includes prospective clinical trials, case reports, retrospective surveys, and reanalysis of early data. The three ADAMHA institutes provided $19 million to support TAR in FY 1975.

TAR was identified as a major agency priority in 1978 (359). To assist TAR, a work group was established with the following overall aims:

1. To develop a plan that will assess the current state of TAR and develop research programs in selected high-priority areas,
2. To advise the Administrator and the institute Directors on priorities for the areas of treatment assessment studies that are important to public health and are feasible within the next 2 years,
3. To develop a long-term plan that will keep the agency abreast of both methodological and substantive developments in order that the institutes' programs can rapidly reflect these developments and changing needs.

Tables 6 and 7 outline the FY 1975 ADAMHA investment in TAR. Table 6 delineates TAR investment by institute and by type of support. Table 7 presents both the number of studies and amount of support by type of intervention. As evidenced in table 6, NIMH programs are the most developed, particularly in the evaluation of the safety and efficacy of drugs used to treat the mentally ill. The $1.5 million figure listed for NIMH under the "grant and intramural" support largely represents that agency's program in collaborative clinical trials. Examples of such trials include the study of hyperbaric oxygen treatment for cognitive defects in the elderly (see chapter 3, case 17) and a study of intensive social casework and neuroleptic drugs in treatment of outpatient schizophrenia (166, 142).

Table 7 indicates that, similar to NIH clinical trials, most research moneys are allocated to therapeutic interventions (82 percent). A smaller portion of money is devoted to diagnostic interventions (12 percent). Only 6 percent of the funds went to study prophylaxis in FY 1975.

HEALTH SERVICES ADMINISTRATION

Several components of the Health Services Administration (HSA), also an HEW agency, conduct assessments of efficacy and safety and support other activities closely related to such assessment. The Indian Health Service and Public Health Service (PHS) hospitals and clinics are extensively involved in testing computer applications to improve the handling of medical information. In addition, PHS hospitals and clinics are involved in clinical drug trials and other research studies supported by both intramural funds and

Table 6.—Alcohol, Drug Abuse, and Mental Health Administration 1975 Inventory of Treatment Assessment Research

Number of and Amount of Support for ADAM HA Supported Treatment Assessment Research Projects Active in Fiscal Year 1975 by Institute and Type of Intervention

(in millions of dollars)

ADAM HA Institute"	Total TAR'S supported in FY 1975		Type of intervention					
			Therapeutic		Prophylactic		Diagnostic	
	Number	Amount	Number	Amount	Number	Amount	Number	Amount
NIAAA . . .	8	$0.4	8	$0.4	–	–	–	
NIDA	52	5.3	49	5.1	–	—	3	$0-1
NIMH	237	13.6	187	10.2	18	$1.2	32	2.1
Total. . .	297	$19.3	244	$15.8	18	$1.2	35	$2.2

•Institute names: NIAAA = National Institute on Alcohol Abuse and Alcoholism; NIDA = National Institute on Drug Abuse; NIMH = National Institute of Mental Health.

Table 7.—Alcohol, Drug Abuse, and Mental Health Administration 1975 Inventory of Treatment Assessment Research

Amount of ADAM HA Support for Treatment Assessment Research Projects Active in Fiscal Year 1975 by Institute and Type of Support

(in millions of dollars)

ADAM HA Institute	Extramural support				Intramural support	Amount of support
	Grant	Contract	Grant & intramural	Total		
NIAAA.	$ 0.4	—	—	$ 0.4	—	$0.4
NIDA	4.0	$1.3	—	5.3	—	5.3
NIMH	10.3	0.0	$1.6	11.9	$1.7	13.6
Total	$14.7	$1.3	$1.6	$17.6	$1.7	$19.3

competitively acquired extramural funds. The nation's largest effort in studying Hansen's disease is conducted by PHS Hospital in Carville, La., with considerable technology development and new technology transfer in the area of treatment of insensitive limbs. The Indian Health Service also has been involved in the evaluation of space technology developed by the National Aeronautics and Space Administration (NASA) for application to remote rural health facilities.

NATIONAL CENTER FOR HEALTH SERVICES RESEARCH

The National Center for Health Services Research (NCHSR) is a component agency of the Office of the Assistant Secretary for Health of HEW. It was established by Public Law **93-353,** the Health Services Research, Health Statistics, and Medical Libraries Act of 1974. NCHSR is authorized to undertake a broad range of research documentation and to evaluate activities pertaining to nearly all aspects of health care delivery.

According to the agency, the "assessments supported by NCHSR are best characterized in a general sense as cost-benefit/cost-effectiveness studies' '(369). These assessments often take place during demonstrations, whereby a technological innovation is studied in the context of the actual health care delivery setting. A mixture of efficacy, safety, effectiveness, and cost information is often developed during these types of assessments. NCHSR has demonstrated, and sometimes developed, a number of technologies, many of which are computer based. The agency also has supported an investigation, conducted by the American College of Radiology, of the efficacy of various X-ray procedures. "Technical consensus" techniques similar to those of NIH are sometimes used by the Center. For example, it sponsored an American College of Cardiology conference and report, *Optimal Electrocardiography* (371).

OFFICE OF HEALTH PRACTICE ASSESSMENT

The Office of Health Practice Assessment (OHPA) is located within the Office of the Assistant Secretary for Health, HEW. OHPA is responsible for providing coverage recommendations to the Social Security Administration (SSA) when questions arise regarding reimbursement coverage under Medicare for new services, devices, or procedures.* OHPA only synthesizes existing information on the efficacy and safety of a given technology; it does not conduct new studies. The agency collects available data and translates that information into recommendations to Medicare regarding coverage. Final authority for deciding issues of Medicare coverage resides in the Health Care Financing Administration (HCFA).

OHPA recommendations are based on evidence in four areas: efficacy, safety, stage of development (i. e., the progression of a technology from the experimental stage to full clinical application), and acceptance by the medical community. Upon receipt of a coverage question, the OHPA staff members conduct literature reviews and contact relevant experts both inside and outside the Government. The opinions of these consultants are contained in a memorandum of recommendation to Medicare that cites relevant evidence regarding the four criteria listed above. The opinion provided by OHPA affects only Medicare; Medicaid coverage is decided by the States. To date, recommendations have been developed for only a minority of the technologies Medicare reimburses.

PHS and HCFA have had several discussions about the four areas of evidence (criteria for coverage). Both agencies agree that cost-effectiveness and cost-benefit considerations should be included in coverage determinations. Some modification of the criteria is actively being considered (319).

A second technology-related activity that OHPA conducts is the Medical Practice Information Demonstration Project. This Project addresses a two-fold problem: that medical practice, including the use of medical technology, is based on information that ranges from hard scientific knowledge to judgment, speculation, and assumption; and that the differences in validity of these sources of medical practice information are not explicit. The Medical Practice Information Demonstration Project is an attempt both to develop and test the feasibility of a technique designed to elicit consensus from recog-

*Before the reorganization of HEW in 1977, reimbursement questions were referred to the Public Health Service's Bureau of Quality Assurance. That Bureau is now the Health Standards and Quality Bureau of the Health Care Financing Administration.

nized experts in a particular field of medical practice regarding the epidemiology, diagnosis, therapy, and economics of a disease entity, and to identify and validate the most authoritative scientific data supporting those opinions. The information derived from such a technique is expected to have these results:

1. Those conclusions about a disease entity or medical technology that rest on a valid information base can be put to immediate use in making regulatory and reimbursement decisions and in quality assurance programs;
2. Those conclusions and assumptions that are unsupportable or rest on an invalid base will help set priorities for biomedical and health services research;
3. The entire profile of validity from well-documented, scientifically supportable knowledge to mere assumptions will find immediate application in medical education.

HEALTH STANDARDS AND QUALITY BUREAU

The Health Standards and Quality Bureau (HSQB) is part of HCFA, HEW. The Bureau is composed of three distinct programs: Professional Standards Review Organization (PSRO), End Stage Renal Disease (ESRD), and Standards/Certification. The most significant of these programs is the PSRO, which is designed to assure that rendered services are medically necessary, consistent with professionally recognized standards, and delivered at an appropriate level of care. It is responsible for reviewing the provision of health services under the federally financed programs of the Social Security Act, i.e., Medicare, Medicaid, and Title V. PSRO medical necessity review determinations are used as conditions for the payment or denial of claims under Medicare and Medicaid.

The PSRO program provides a mechanism for developing consensus regarding the appropriate use of particular medical technologies through the criteria and standards development process. Each PSRO is responsible for developing its own criteria and standards. These standards are based on local patterns of medical practice. The National Professional Standards Review Council may adopt exemplary norms, standards, and criteria, and distribute them to PSROs for their adoption and use. Any locally developed norms, criteria, or standards of care that differ significantly from those developed by the National Council may be disapproved by it.

While the National Council has provided general guidance to the PSROs and sent actual criteria sets to PSROs for use, these have not been officially adopted, and therefore, are used only as technical assistance. HCFA has stated that the National Council will begin to adopt exemplary sets and review local PSRO sets to determine if they differ significantly. If the differences cannot be justified, the National Council is expected to disapprove the use of those norms, criteria, and standards by that PSRO.

Information for the criteria sets issued to date by the National Council was developed under contracts with such groups as the American Medical Association, the American College of Physicians, and various university hospitals. The primary purpose of the studies was to develop criteria on medical necessity for hospitalization for different disease categories. However, some of the criteria sets include indications for the effective use of drugs, devices, and procedures.

OTHER FEDERAL PROGRAMS

Only selected Federal programs involved in evaluations of efficacy and safety of medical technologies have been described. However, more than a dozen Federal agencies conduct or support biomedical research, some of which involve testing for efficacy or safety. Two agencies that conduct such testing are covered below: the Veterans Administration (VA) and the Department of Defense (DOD). Other agencies that conduct this type of research but are not covered in this report include NASA and the National Science Foundation (NSF).

Veterans Administration

The health programs administered by VA are designed to provide quality medical care to veterans. In order to furnish such care to veterans, VA spends approximately 80 percent of its health-related budget on direct provision of services. VA's Department of Medicine and Surgery runs the largest centrally directed patient care system in the United States and serves an eligible population of about 30 million. The authorizing statutes for VA do not include specific references to either efficacy or effectiveness. Nevertheless, the Department of Medicine and Surgery has promulgated regulations, manuals, and circulars related to the efficacy and safety of medical technologies and services.

VA (and DOD, see below) conducts a full range of technology activities, including research and development, validation assessment, transfer to practice, and dissemination of information. The Department of Medicine and Surgery has a Research and Development Division involved in basic medical research, clinical trials, health services research, and rehabilitative engineering. The FY **1976** budget for research and development was $96 million. All research is conducted within the VA system.

Clinical trials aimed at testing efficacy and safety can be funded either through the existing budgets of each facility or the Research and Development Division. In 1976, 24 multicenter cooperative studies were in progress. One of these studies tested various drug treatments for hypertension (see chapter 3, case 12). Other examples of such VA-sponsored studies include a trial of the efficacy of immune serum globulin for the prevention of post-transfusion hepatitis and a trial of methadyl acetate and methadone as maintenance treatment for heroin addiction (303,216).

Department of Defense

DOD operates a network of hospitals and clinics that are intended to provide health care for active-duty personnel and retired members of the uniformed services. The authorizing statutes covering the health programs in DOD do not mention efficacy or safety explicitly. The Assistant Secretary of Defense of Health Affairs, however, has responsibility for establishing uniform policies, standards, and procedures for medical care. Many directives and instructions establish criteria and standards for aspects of medical technology related to efficacy and safety, such as standards for the purchase of hardware. In addition, FDA performs the functions related to quality assurance of drugs, devices, and biologics procured by the Department.

DOD supports a considerable amount of health-related research. In 1976 alone, expenditures for health research totaled more than $114 million. DOD research and development activities are directed toward providing medical knowledge and expertise in those areas that primarily affect the military. These activities include clinical trials that

test the efficacy and safety of medical technologies. For example, DOD spends about $15 billion annually on the development and assessment of field medical care and evaluation systems (see chapter 3, case 5). In addition, approximately $40 million is spent each year to develop and assess new technologies to ensure troop readiness through disease prevention. A third area of DOD research and development activities involves the collection of scientific data for use in establishing safety criteria for exposure to hazards arising from military environments.

Private Sector Activities

The private sector supports many activities intended to evaluate the efficacy and safety of medical technologies. In addition, many Federal programs depend upon private sector facilities and personnel to produce much of the data used for evaluations of safety and efficacy. In fact, most federally financed clinical trials take place in private sector hospitals and clinics, many of which are university-affiliated.

The work of individual physicians or medical center research teams has resulted in a number of innovative medical and surgical procedures. Although there are few formal requirements that mandate new procedures be shown to be efficacious before their use, a substantial amount of testing is still conducted with or without Federal funds.

In addition, there are a number of indirect controls on the use of new technologies. Swazey **(332)** has identified four such controls: 1) professional training and socialization, 2) peer group controls, 3) design and conduct of clinical research, and 4) physician/patient or investigator/subject relationships.

Some professional associations have developed formal mechanisms for reviewing accumulated evidence regarding the proper use of a technology. * In late 1976, the Medical Practice Committee of the American College of Physicians recommended that the College "explore the feasibility of forming an organization to develop a mechanism for the systematic review of the efficacy of diagnostic and therapeutic procedures. " The American Academy of Pediatrics has developed recommendations on immunization practices. The American Public Health Association periodically compiles a list of effective preventive and therapeutic procedures for infectious diseases. The Council of Medical Specialty Societies, the American College of Surgeons, and the American College of Physicians have provided advice to the National Blue Shield on the efficacy of lumbodorsal sympathectomy, uterine suspension, and basal metabolic rate determinations—all questionable procedures for which Blue Shield was continuing to reimburse. The American Hospital Association and the American College of Radiology also have been involved in similar activities.

Professional associations are becoming increasingly involved in standards setting. Standards are seen as the means by which professionals, consumers, industry, and the Government can accept and communicate technical recommendations for certain characteristics of technologies. The usual mechanism for the development and approval of such standards is **a** multiorganization group that is composed of all relevant and affected disciplines and interests. The largest source of voluntary consensus standards is the American Society for Testing and Materials (ASTM). ASTM has promulgated voluntary standards

*In these reviews, distinctions between efficacy and effectiveness are hazy; however, they seem to emphasize effectiveness.

in several medical areas, including implantable devices and prosthetics. Standards approved by the members of ASTM are submitted to the American National Standards Institute (ANSI) for acceptance as the American National Standard. ANSI is a voluntary federation of more than 400 standards-writing bodies in the United States.

Another medical standards-setting group is the Association for the Advancement of Medical Instrumentation (AAMI). The Association, which represents 5,000 professional, corporate, and institutional members, provides a forum in which health care professionals, manufacturers of medical devices, and Government representatives can interact to develop standards that promote patient safety. This is accomplished by establishing basic performance and user information requirements. AAMI has committees operating in Ambulatory Monitoring, Autotransfusion, Human Engineering, and Otolaryngology, among other areas. These committees take into consideration all factors including technological and economic impacts relevant to the establishment of a reasonable level of safety and efficacy.

The Alliance for Engineering in Medicine and Biology also has assessed a number of medical technologies over the past few years, particularly in the area of ultrasonic diagnosis (41). The alliance usually does not conduct formal evaluations of the efficacy and safety of specific medical technologies. However, the alliance's assessments may increase practitioner awareness regarding the importance of considering both the type and validity of efficacy and safety information as it relates to technologies they use or plan to use. The alliance report on technology procurement in health care institutions (5) is an example of their efforts to provide practitioners and administrators with a process by which they can evaluate technologies.

6.

STATUS AND IMPLICATIONS OF EFFICACY AND SAFETY ASSESSMENT

6.

STATUS AND IMPLICATIONS OF EFFICACY AND SAFETY ASSESSMENT

Efficacy and safety are extremely important starting points in determining if technologies will be safe and effective in use. If a technology does not provide benefit with acceptable risk under optimal, controlled, research conditions, then it will not do so under average conditions of use. Simply stated, efficacy is essential to effectiveness. *

Chapter 1 briefly mentioned the general importance of efficacy and safety data. That theme is further developed in this chapter, which presents information on the uses and users of such data. This chapter also describes a normative model of the generation, processing, and dissemination of efficacy and safety information, and contrast current programs and systems for assessment to the normative system. Finally, it examines the status of information on efficacy and safety.

USES AND USERS OF EFFICACY AND SAFETY DATA

Any person or institution using or directly affecting the use of medical technologies is a user of efficacy and safety information. There are two basic types of users: "passive" and "active. " Patients or consumers of medical care often can be viewed as "passive" users of efficacy and safety knowledge. Many Government and private sector programs, for example, several of the grant programs of the Department of Health, Education, and Welfare's (HEW) Health Services Administration (HSA), also are "passive" users. HSA, for example, may award a grant to a community for the establishment of certain specific health services. The agency does not require that technological services provided with these funds be of demonstrated efficacy and safety. This situation represents a passive use of efficacy and safety information, because the usefulness of the grant program depends in part on the effectiveness, and thus the efficacy, of the services purchased. "Active" users of efficacy and safety information include physicians, biomedical and health services researchers, nurses, and other health professionals, many public and private third-party payers, and personnel in Government regulatory programs and medical schools, and so on. Table 8 lists many of these users of information, the uses, and the sources of information.

Information from well-designed and valid studies of effectiveness can be of higher utility than studies of efficacy to most of the users listed, because many of them are concerned primarily with the benefit of a technology under actual or average conditions of use. Because of the difficulty of conducting evaluations of effectiveness, information on effectiveness is often lacking. Efficacy information, the next best source of guidance on

● However, even if a technolog, were safe, efficacious, and effective, it might lack social benefit if overriding ethical or other societal concerns were not addressed satisfactorily y.

Table 8.—Users of Efficacy and Safety Information

User	Actions taken on the basis of efficacy and safety information	Major sources of information
Non" Federal public or private programs:		
Physicians (and nurses, other health professionals)	• Clinical decision making relative to diagnosis, treatment, and prevent ion of health problems • Decisions to adopt new technologies • Publishing, communicating to professional associations, colleagues, etc.	• Own experience • Colleagues • Professional meetings • Professional literature • Detail men, other manufacturers' representatives
Professional associations	• Set standards for use of technologies • Assess competence for certification, etc. • Communicant ion to membership, etc.	• Professional literature • Experience of members and other health professionals
Schools of medicine or public health	• Instruction • Set agendas for future research	• Knowledge and experience of faculties • Professional literature
Private sector third-part y payers	• Decisions to place a technology on the coverage schedule • Decisions to reimburse for specific uses of a technology	• Professional opinion • Professional literature • Associations
Federal Government programs:		
Food and Drug Administration, PHS	• Decisions to allow investigational use of drugs or devices • Decisions to allow marketing of drugs or devices • Decisions to allow products to stay on market	• Manufacturer or sponsor • Professional literature • Staff knowledge • Outside professional advisors
Medicare program, HCFA	• See private third-part y payers	• Office of Health Practice Assessment, PHS • NIH, and other Federal programs • See private third-party payers
Medicaid program, HCFA	• See private third-party payers (HCFA recommends such decisions but the States have the decision authority)	• Medicare decisions • See private third-party payers
National Institutes of Health, PHS	• Decisions on research agendas • Decisions on demonstrate ion and control programs • Disseminate ion of information	• Research conducted at or supported by NIH • Professional literature • Outside advisors • Staff knowledge
Health Resources Administration, • PHS	• Set national guidelines for health planning • Develop planning guidance for certificate-of-need determinations	• Other Federal agencies • Contracts with private organizations • Professional literature
Office of Professional Standards Review Organizations, ••HCFA	• Set guidelines for medical care reviews • Set guidelines for reviews of institutional and length-of-stay admissions	• See Health Resources Administration

•And State and private sector programs linked to HRA, such as health systems agencies,
••And private sector, local PSROs

appropriate use of technology, is therefore utilized more frequently. For example, regarding new technologies, there is usually little or no experience with them under average conditions of use for the development of even informal professional consensus of effectiveness. For these reasons, it is important to develop and disseminate the most valid and comprehensive efficacy and safety information possible, within resource and methodological constraints.

A SYSTEM FOR ASSESSING EFFICACY AND SAFETY

The adoption and use of medical technologies by health care professionals should be based on well-validated information regarding their benefits and risks. This statement does not imply that every aspect of every technology must or can be subjected to randomized, controlled clinical trials. That would be an impossible task for several reasons, including financial and human resource limitations, the excessive time requirements, philosophical and political considerations, the complexity of medical technologies and their uses, etc. However, it does imply both the existence of accurate and relevant information, which is developed to the extent desired and practical, regarding the effects of technologies and the dissemination of such information to the individuals and groups in need of it. Also, this information should pertain to the benefits and risks of a technology under the conditions in which it will actually be used. Because of the difficulty of obtaining effectiveness and safety data, decisionmakers substitute efficacy and safety data as a somewhat equivalent measure of the technical effects of technology.

This section presents a model of the process of generating, processing, and disseminating information on efficacy and safety. This model is then compared to the current systems and programs in order to examine whether shortcomings exist in the current systems.

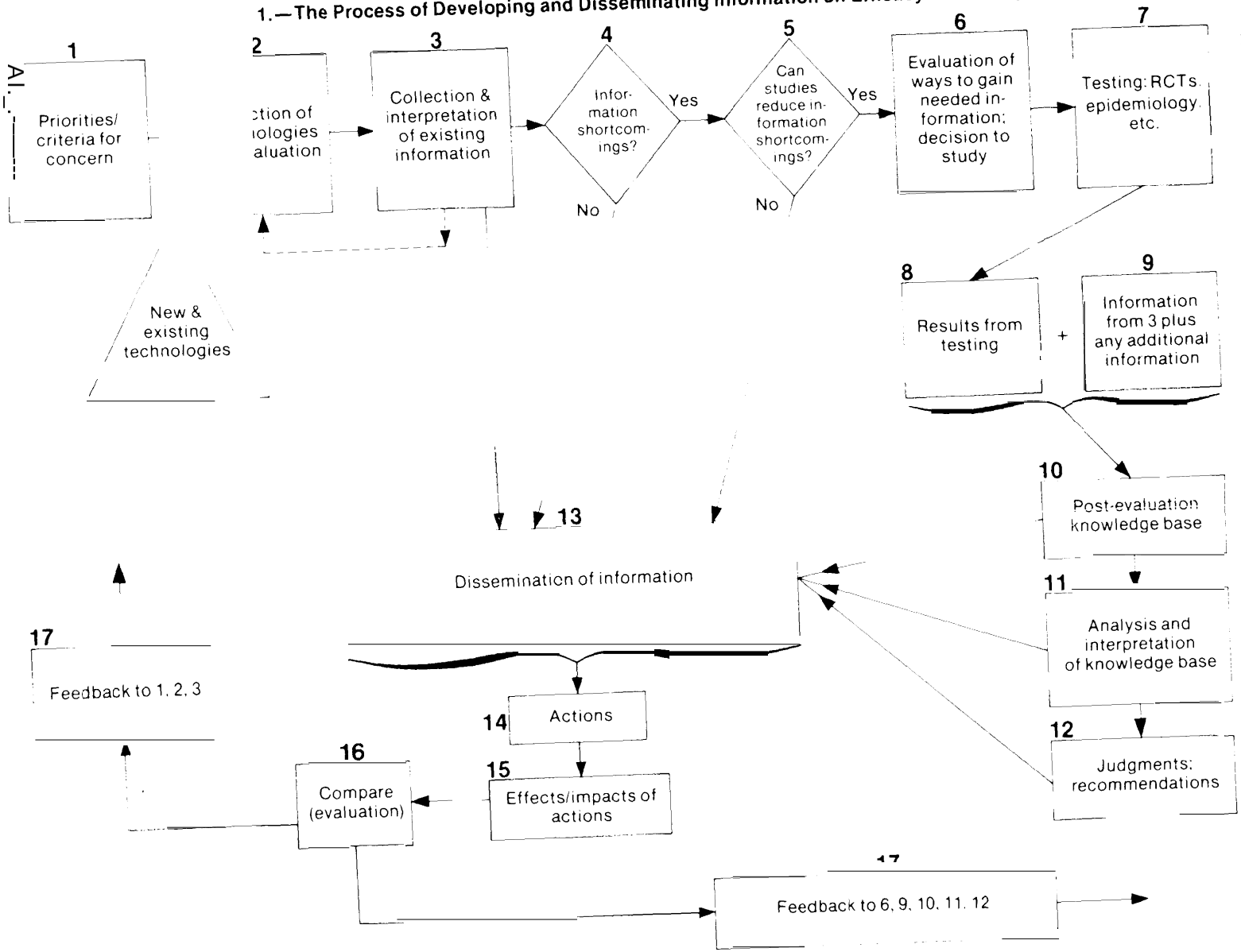
Developing and disseminating information on efficacy and safety is a tremendously complex process. Although many of the intricate details of the process are not germane for the purposes of this report, the complexity of this process should not be forgotten. To illustrate some of this complexity, figure 1 depicts many of the elements involved in assessing efficacy and safety. Even that relatively complicated process described in figure 1 represents a simplified abstraction of the reality.

In this report, the process is viewed as an interdependent and nondiscrete flow of four types of actions:

- Identification: Monitoring technologies, selecting those in need of study, and deciding which to study. (Steps 1-6 of figure 1)
- Testing: Conducting the appropriate analyses or trials. (Step 7)
- Synthesis: Collecting and interpreting existing information and the results of the testing step, and, usually, making recommendations or judgments of efficacy and safety. (Steps 8–12, and often 3)
- Dissemination: Providing the synthesized information, or any other relevant information, to the appropriate parties who use or make decisions concerning the use of medical technologies. (Step 13)

The action steps represented in figure 1 are not within the scope of this report. For a description of some of the possible actions, see table 8, Also, an HEW report on medical

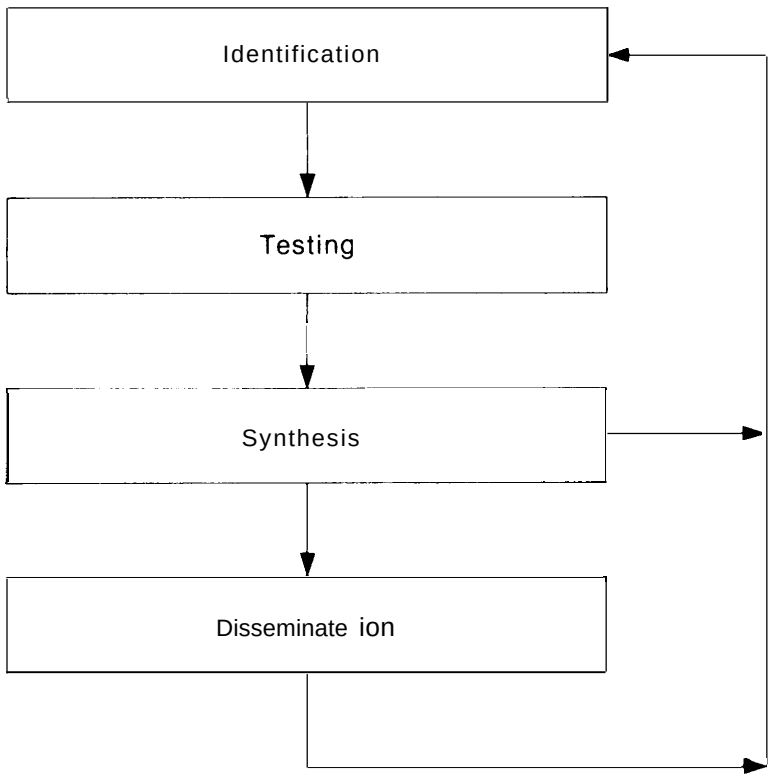
1.—The Process of Developing and Disseminating Information on Efficacy and Safety



technology management *(369)* describes in greater detail the potential actions and relates them to a similar model.

The four elements of a normative system for developing and disseminating efficacy and safety information are depicted in figure 2.

Figure 2.—Simplified Process for Developing and Disseminating Efficacy and Safety Information



This model represents only one possible method of viewing the process of assessing medical technologies. It is designed to serve as a logical standard against which existing assessment programs may be evaluated.

SHORTCOMINGS OF CURRENT SYSTEMS AND PROGRAMS

The primary shortcoming in current assessment methods is the lack of a formal *or well-coordinated "system"* for developing and disseminating safety and efficacy data (53,250,357,369). Some elements of the process are operating and performing well. However, the elements are not linked together and do not follow each other logically. The Assistant Secretary for Health of HEW has stated (357):

> There are, of course, informal mechanisms for the assessment of the health-care technology. It is probably true that such informal approaches served us reasonably well in the past. But for a variety of reasons, we can no longer rely on such informality.

HEW recognized the lack of a "strategy for managing medical technology. . . and . . .an analytical paradigm upon which to develop such a strategy" **(369).** A report to the Secretary in December 1977, outlined the components of "such a strategy. " Responding to that study, the Secretary of HEW established an Office of Health Technology in January 1978. The Office was designed to include these functions: testing and demonstrating the strategy developed in the 1977 study, serving as a focal point for health technology policy development in the Department, and providing recommendations to the Health Care Financing Administration (HCFA) on the advisability of reimbursement for specific medical technologies **(287).** As of September 1978, however, insufficient implementation of the proposed HEW system had taken place. Consequently, the Office of Technology Assessment (OTA) was unable to analyze the actual functions being fulfilled.

Development and dissemination of information on the efficacy and safety of drugs and devices more closely approximates a coherent system than does the assessment of medical and surgical procedures. Beginning in 1906 with the passage of the Federal Pure Food and Drugs Act, various laws have been enacted to regulate the safety and/or efficacy of both drugs and medical devices. Surgical and other procedures that depend primarily on providers' techniques have not been subject to similar Federal controls. Assessment of safety and efficacy for these procedures has remained primarily in the hands of the profession.

There are a number of factors which help explain the differences in the safety and efficacy evaluations for products and procedures. One of these is the physical nature of products. Investigators can learn much about products before they are tested clinically (394). For procedures, however, clinical testing is the essence of their development. In addition, procedures are complex, and therefore, their evaluations are correspondingly complex.

Source of sponsorship also distinguishes products and procedures. Drugs and devices usually are developed for marketing by profitmaking firms. Mechanisms have been created to regulate industries. Procedures, however, are usually developed by an individual physician or medical team. Given the history of relative autonomy the medical profession has enjoyed in our society, it is not surprising that the profession has been given the responsibility for regulating its own members and their use of technology (125,332,334). It appears, therefore, that one major problem in assessing efficacy and safety centers on procedures which develop without control or planning in the private sector of medical practice.

Identification

Presently, there is no complete list or catalog of either existing medical technologies or those that particularly require assessment for efficacy and safety. Partial lists do exist. The Food and Drug Administration (FDA), for example, has lists of approved drugs and devices. The fact remains, however, that many medical procedures, which are not on reimbursement schedules, but are important to assess (bed rest for certain diseases, for example) are not cataloged in one source.

No existing system completely identifies developing technologies that will need evaluation for safety and efficacy. The National Institutes of Health (NIH) does a yearly study of its clinical trials and publishes a catalog of those trials it supports. Other agencies, such as the Veterans Administration (VA), have similar catalogs or lists. Through

its premarket approval process, FDA gathers information on drugs and devices that are being developed. If medical and surgical procedures were to be evaluated before they came into widespread use, however, some comprehensive system for recognizing them in a timely fashion would be necessary. A variety of sources could produce such a catalog. Professional literature is one source. Another is institutional committees that review research for adherence to ethical standards. Complete lists of clinical trials would provide the beginning of an "early warning system. "

Even if funds for, and numbers of, clinical trials were greatly expanded, setting priorities for study would still be necessary, because it is neither possible, nor desirable, to study every efficacy- or safety-related aspect of medical technology. Such priorities might help to ensure that all areas of medicine, such as prevention, are considered. Priorities for assessment might include beneficial technologies that are neglected or technologies that are suspected to be useless or dangerous. Technologies that are, or are expected to be, either expensive or widely used also could be given priority. For new technologies, potentially important advances could be assessed rapidly.

In sum, there is no formal process for selecting which technologies are to be studied; indeed, there is not even a set of priorities for such selection. New drugs and new devices are, however, subject to the FDA market approval process and thus are automatically identified for study, at least in regard to the efficacy and safety claims of the manufacturers.

Testing

The testing phase includes stimulating, requiring, funding, or conducting studies. Shortcomings related to the testing phase center around four issues: 1) the state of the methodologies for conducting controlled trials, consensus activities, and other tests; 2) the level of financial support, particularly for controlled clinical trials; 3) the relative appropriateness of the questions and technologies being studied; and 4) the number of personnel qualified to conduct such research.

Although the state of clinical trial methodologies has improved dramatically in the past 30 years, there are still uncertainties involved in the design of each trial. This report is not directly concerned with the technical methodologies for testing, but it should be noted that "there is no standard textbook on clinical-trial methodology" (147), and that the further development and dissemination of methodological information would complement efforts to assess efficacy and safety.

There is no "correct" level of financial support for clinical trials; no one can set an exact figure for the amount that should be invested in trials and other forms of testing. Does the current level of funding, then, represent a shortcoming? This question must be answered positively because important areas of health care are not receiving adequate investigation, according to the evidence gathered by OTA. New or developing immunization and screening technologies and new procedures are studied relatively infrequently, as are existing technologies of all types. This discussion applies to both the second and the third shortcomings listed above.

Often, the decision to investigate a certain question (for example, what specific effects of a technology are being examined?) has been influenced by such factors as investigator curiosity, research needs, and so on. * The concerns and information needs, for

*Many of the shortcomings of the testing phase are intimately related to the inadequate identification phase.

example, of health planning agencies or Professional Standards Review Organizations (PSROs) are much less frequently considered in these decisions (369). Changes in the level or direction of the Nation's activities in assessment of efficacy and safety would highlight the limited number of personnel presently qualified to conduct such research. Biostatistics and epidemiology have been less affluent areas of health research (57). Consequently, the number of epidemiologists, statisticians, and others essential to efficacy and safety assessments may be inadequate for future needs.

In short, the country has the potential to develop a good capability for testing efficacy and safety, but the actual effort could perhaps be expanded or at least organized according to somewhat different priorities. Such an effort may require an expanded base of qualified research personnel.

Synthesis

Synthesis involves a critical analysis of the results of testing (available data from preclinical to clinical experience, epidemiological studies, and controlled trials) and all other available and relevant information. This analysis involves a "putting together" of the data into a summary of the efficacy and safety of the technology in question. It usually takes the form of judgments or recommendations regarding the appropriate indications for use of the technology. Consensus development, which is described in chapter 4, also can be considered a synthesis activity. Syntheses are most commonly found as review articles in the medical literature. However, this literature varies in quality and is usually not directed toward the needs of practitioners. Williamson notes that "many, if not most, health sciences publications are detailed, highly technical research reports directed by the investigator to his fellow researchers, " and that "interpretation of many. . requires an understanding of technical terminology, research design, and analytical statistics that is beyond the scope of the average professional. . . ." (428).

The validity of published information also has been questioned. Two studies of research reports in leading medical journals found nearly 75 percent of the publications analyzed to have invalid or unsupportable conclusions as a result of statistical problems alone **(115,300).** Other studies that focused on research design, data collection, and analysis in specific areas of medicine found that none of the articles studied yielded valid or supportable results (137, 189), When Juhl and his coworkers examined the literature in gastrointestinal diseases, they found that few well-designed trials were conducted. Additionally, they observed a preponderance of positive trials, indicating a bias toward positive results (186). Furthermore, 80 percent of the trials dealt with new treatments; few were concerned with evaluating "established treatments. "

Federal Government synthesis activities are expanding. The consensus development activities of NIH are too new for evaluation of their effects. The hypertension synthesis (see chapter 5) seems to have had positive impact. Many of the consensus exercises planned for 1978 by the Institutes of NIH, however, appear to be modifications of seminars and conferences planned previously. How well these activities fulfill the synthesis function remains to be seen, but there is great potential. The Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) has used **a** technique related to consensus development in the area of psychosurgery. However, that agency contends that a more formal and quantitative technique should be developed. The process of recommending coverage decisions to Medicare (chapter 5) by the Office of Health Practice Assessment represents another synthesis activity. That Office has stated, however, that because of the ad hoc nature of the process there is "no assurance that the best and most reliable data are utilized in a given case" (369).

Despite the recent expansion in synthesis activities, they still represent a modest level of activity that have suffered, at least in part, from lack of quality in both content and process. Furthermore, synthesis activities are hampered by the lack of well-validated information on efficacy and safety.

Dissemination

Many of the comments relating to synthesis also apply here. Federal agencies have not assigned a high priority to disseminating information. FDA sometimes sends letters to all physicians as one mechanism for distributing important information. The National Center for Health Services Research (NCHSR) frequently disseminates information to a wide audience by issuing a series of NCHSR Research Reports that describe the results of projects funded or conducted by that agency. Also, NIH has provided information primarily to the professional community through its demonstration and control projects, through the National Library of Medicine, and through other activities, including a regular feature in the *Journal of the American Medical Association.*

As described in chapter 5, the private sector also has multiple channels which encourage the flow of information. Professional societies are expanding their activities in this area.

The Federal Government provides little information for such public agency activities as health planning programs. In the case of the computed tomography (CT) scanner, for example, the Bureau of Health Planning and Resources Development, the Federal agency which administers health planning activities, contracted with a private firm to produce planning guidelines for such devices. Likewise, third-party reimbursers, such as the Medicare program, seldom receive assistance from such agencies as NIH in deciding benefits.

STATUS OF EFFICACY AND SAFETY INFORMATION

The shortcomings described above would be much less deleterious if the state of knowledge about the efficacy and safety of medical technologies were adequate. Conversely, if the state of information were inadequate and there were no shortcomings in the processes and systems of assessment, perhaps little could be done to improve the information base. The data inadequacies, and the corresponding difficulties in using technologies, might then be the inevitable result of the inherent complexities in the field of medicine. However, there *are* shortcomings in the current ways in which efficacy and safety information are developed and disseminated. Therefore, data inadequacies and their effects— inappropriate diffusion and use of technologies—need examination.

Many technologies have been shown to lack efficacy or be unsafe only after enjoying widespread use. A psychosurgical procedure called leucotomy or lobotomy, for example, was widely adopted in the early 1950's and was subsequently abandoned when its efficacy and safety were seriously challenged. The Wassermann test for diagnosing syphilis was used for over 40 years until it was discovered that only half of the patients with positive test results actually had the disease (223). More recent examples include internal mammary artery ligation (see chapter 3, case 8), colectomy (surgical removal of the large intestine) for epilepsy (162), carotid-jugular shunts for mental retardation, lumbo-dorsal sympathectomy, uterine suspension, and gastric freezing.

Questions of efficacy have been raised recently regarding a number of medical technologies currently in use (72,124,162,179,223). As mentioned earlier, White has stated that only 10 to 20 percent of all procedures used in present medical practice have been shown to be of benefit by controlled clinical trials; many of the other procedures may not be efficacious (426). In fact, many technologies in use have had their efficacy and safety questioned, including oral drug treatment for diabetes (64,236), respiratory therapy (19,24), oral decongestants (207), thermography for diagnosing breast cancer (248), ergotamine for migraine headache (410), immune serum globulin for preventing hepatitis (303), intensive care for pulmonary edema (152), coronary care units (233), and radical mastectomy (228).

Such widely used technologies as tonsillectomy, appendectomy, and the Pap smear have not been completely assessed for efficacy (see chapter 3, cases 1, 9, and 10). Others, such as electronic fetal monitoring (EFM) and coronary bypass surgery, have been diffused rapidly before careful evaluation (see chapter 3, cases 7 and 8). Concern about risks has led to questions regarding the use of mammography and skull X-ray (see chapter 3, cases 4 and 6).

The above are only examples. Others could be listed. The systems for assessing efficacy and safety have made the compilation of such a list possible. However, the same systems were not able to provide early and adequate information in order to prevent or delay the spread of technologies until their effects had been predicted more clearly. Further, since these examples can be cited, there are probably many others. Although perfect information on efficacy and safety can never be attained, shortcomings in assessment systems may be impeding a closer approximation of that goal. The status of efficacy and safety information cannot be exactly determined, but the combination of long lists of examples of technologies inadequately assessed and shortcomings in assessment procedure processes may indicate that improvement is possible.

7.

POLICY ALTERNATIVES

7.

POLICY ALTERNATIVES

This chapter outlines a number of policy alternatives intended to correct some of the shortcomings in the assessment process presented in earlier chapters. Many of these options do not require new legislation because sufficient authority already has been written into law. In certain cases, desired actions could be stimulated by congressional oversight. Alternatives are presented for each of the four phases of the assessment process: identification, testing, synthesis, and dissemination. Although not previously discussed, several policy alternatives which attempt to translate efficacy and safety information into improved management of the utilization of technologies also are presented. Many of the alternatives and all of the steps of the process are both relevant and applicable to other types of assessments of medical technologies. Cost-effectiveness assessments, for example, could follow the four-step process. In that context, the advantages and disadvantages presented in this chapter would have to be modified to reflect the expanded functions.

The first question to be addressed is to what extent, if any, should the Federal Government either change or expand its activities in the process of assessing efficacy and safety. As described in chapter 5, existing Federal and private mechanisms execute important parts of the task of assessing the efficacy and safety of medical technologies. The Food and Drug Administration (FDA) has the statutory responsibility for assuring safety and efficacy of drugs and devices. Other Federal agencies, such as the National Institutes of Health (NIH), fund clinical trials that produce information on safety and efficacy. The private sector supports a large number of clinical trials, some mandated by FDA legislation. If Federal action were desirable, the four functions described above could be assigned to one agency or divided among several agencies in the Federal Government, They could be developed in one or more existing agencies, or an entirely new agency could be developed. Alternatively, the private sector could be encouraged or provided incentives to expand its activities in these areas. Or, some combination of Federal and private strategies could be pursued. Again, the first question is whether the Federal Government should or should not act; that question must be decided by Congress.

SECTION ONE: CONGRESSIONAL ALTERNATIVES

Alternative A-1: Any change or expansion in the development of information on the safety and efficacy of medical technologies could be left to the private sector. This alternative does not imply that there are no problems in existing private sector activities. This alternative would give Government a twofold role: to stimulate the private sector and to monitor its activities.

Alternative A-2: The Federal Government could expand activities relating to the development of information on efficacy and safety of medical technologies. A series of possibilities is presented later in this chapter which could be followed if this alternative

were desirable. This alternative could include legislative mandates for the performance of certain activities.

Alternative A-3: Some combination of alternatives A-1 and A-2 could be pursued.

Any agency or agencies involved in assessing efficacy and safety could complete this task better if certain criteria were met. As examples, such an agency (or agencies) might need:

- An explicit mission concerning efficacy and safety assessment, The agency must accept this role and be held accountable for its performance.
- Statutory or regulatory authority to accomplish its mission. For example, it should be able to gain access to information it needs, including access to FDA materials considered to be proprietary.
- Adequate funding for the assigned mission. This might require an existing agency to reorder its spending priorities. In addition, new funding would probably be necessary.
- A competent, multidisciplinary staff with expertise in technology development and technology evaluation. Statisticians, physicians, epidemiologists, sociologists, economists, and others would be essential.
- Credibility with the health professions, scientists, industry, and third-party payers. It would be desirable if the agency already had relationships with these groups. Relationships with practicing physicians are important, particularly because information dissemination to that group would be an important task. Working relationships with other Government agencies involved in technology development and use would also be necessary.

The following sections discuss a series of alternatives to current policy in each of the four areas mentioned earlier. The functions could be addressed in many ways. The alternatives given are not exhaustive, but rather illustrative. Nor are they mutually exclusive. Furthermore, any agency could use a variety of programmatic mechanisms for meeting its objective: grants, contracts, intramural research, and mandating or requesting assessment from those who are able to provide a service. Any or all of these mechanisms could be used by any one agency. The alternatives that follow do not discuss or compare these approaches. (Table 9 summarizes the possible responsible organizations for conducting the four basic functions in efficacy and safety assessment.)

SECTION TWO: IDENTIFYING TECHNOLOGIES THAT NEED ASSESSMENT

A system for identifying technologies that need assessment could be developed in a number of agencies at various levels.

Alternative B-1: A special commission could be established to identify technologies needing assessment. This task will be a lengthy one requiring a special commitment. Establishing a special commission for that purpose would have some advantages. It could include prestigious physicians as well as experts from other disciplines and lay “representatives. Its deliberations could be open to public scrutiny. The major disadvantage in choosing this alternative is that such a commission would be far removed from sources of

Table 9.—Possible Sites for Carrying Out Four Key Tasks in Efficacy and Safety Assessment

Identifying technologies that need assessment

B-1. A new commission
B-2. Institute of Medicine
B-3. National Institutes of Health
B-4. Agencies involved in technology development
B-5. Food and Drug Administration
B-6. A new Federal office or agency, or the Office of Health Technology

Requiring, stimulating, conducting, or funding studies

c-1. National Institutes of Health
C-2. Other Federal agencies
C-3. Food and Drug Administration
C-4. A new Federal office or agency, or the Office of Health Technology

Synthesizing information

D-1. A new commission
D-2. Institute of Medicine
D-3. National Institutes of Health
D-4. Agencies involved in technology development
D-5. Food and Drug Administration
D-6. Office of Health Practice Assessment
D-7. A new Federal office or agency, or the Office of Health Technology

Disseminating information

E-1. National Institutes of Health
E-2. Other Federal agencies
E-3. A new Federal agency, or the Office of Health Technology
E-4. A new office in HEW

new technologies, including those that might prove to be problematic. Furthermore, new staff and multiple subcommittees would be necessary.

Alternative B-2: The task could be assigned to the Institute of Medicine of the National Academy of Sciences, This is a desirable option because it chooses an extant, prestigious organization for the task. (The National Academy of Sciences previously carried out the task of evaluating evidence of the safety and efficacy of drugs on the market at the time of the passage of the 1962 Food and Drug Amendments.) The institute would probably have good sources of information about development of procedures in academic medical centers. As **a** quasi-governmental body, the institute could bridge the gap between Government and private sector medicine. The disadvantages of using the institute are the relatively small number of practitioners in its membership and the uncertainty as to whether it would perform such a task.

Alternative B-3: The task could be assigned to NIH. This arrangement is advantageous because NIH administers most of the Federal biomedical research support and a large percentage of the national expenditure. Staff at NIH could be expected to be cognizant of developments even in areas in which NIH has not committed funds. However, NIH has exhibited a stronger interest in developing medical technologies than in assessing them. To some extent, this potential problem could be ameliorated by placing

the function high in the administration of NIH, possibly in a new division or bureau. Such placement might avoid the parochial concerns of the various disease-oriented institutes. Nonetheless, if NIH were assigned this function, careful oversight by the higher echelons of the Department of Health, Education, and Welfare (HEW) and Congress would be essential to assure the effective completion of the task. Another potential problem in choosing this alternative is that d-e accomplishment of the basic mission of NIH could be hampered by such a new function.

Alternative B-4: Each agency (for example, the National Center for Health Services Research (NCHSR)) developing medical technologies could be asked to develop a list of its technologies that would need evaluation. This option would avoid the creation of another bureaucracy, It would also make an important function even more diffuse than it already is, and would lead to a great deal of overlap. In addition, it might **leave** many extant and new procedures unassessed. There are also potential, informal conflicts of interest associated with this alternative.

Alternative B-5: FDA could be assigned the task. FDA has experience in evaluating new technologies, and many of the same principles used in evaluations of drugs and devices could be applied to the area of procedures, with or without a regulatory program specifically concerned with procedures. The major disadvantage of using FDA is that it has had much more experience in working with private firms than in completing the type of function described here. Furthermore, FDA lacks technical resources and has image problems in the practicing community.

Alternative B-6: A new agency or office could be developed, possibly within HEW, that would be assigned the responsibility for efficacy and safety assessments. Its mission could include any combination of identifying technologies to be assessed, conducting and funding the studies, evaluating and synthesizing the information, and disseminating that information. The advantage in choosing this option is that no existing agency is deeply committed to assessing the efficacy and safety of medical and surgical procedures. Alternatively, it is difficult to establish a new agency, assign it a mission, document its need for a new budget, and recruit expert staff. Furthermore, it may not be desirable to develop a new bureaucracy that would handle all four functions when existent agencies and programs could do some, or most, of the job.

HEW has established an Office of Health Technology that would probably have the identification function within its mandate. The future structure and functions of that Office are unclear, however. If the Office of Health Technology begins functioning, it could engage in any or all of the activities specified in this report. Similarly, there are bills in Congress which would establish Federal agencies or offices that could be assigned many of the assessment functions, including identification.

SECTION THREE: REQUIRING, STIMULATING, CONDUCTING, OR FUNDING STUDIES

Expanded support for efficacy and safety testing could be developed in a variety of ways:

Alternative C-1: NIH could assume a larger role in testing both new and existing technologies for efficacy and safety. This option has the advantage of assigning the function to an agency that is already familiar with the field and, therefore, best equipped to

identify developing technologies. This alternative is disadvantageous because not only has NIH been reluctant to assume such an expanded role without new funding, but also NIH has resisted becoming deeply involved in existing medical practice. One method of realizing this option might be to develop a new program or bureau at NIH. The option would be most effective if new money were appropriated to NIH.

Alternative C-2: Other Federal agencies could be asked to expand their roles. The Veterans Administration (VA) is an obvious choice because it offers an excellent field for testing efficacy and safety due to its activities within a medical system that is quite practice oriented. However, VA's funds for medical research are limited, and most of its population is comprised of adult males. Furthermore, VA lacks connections both to HEW and the general community of practitioners. Nonetheless, VA and other agencies could make important contributions.

Alternative **C-3:** FDA could be given a larger role. However, FDA's experience is in administering a regulatory program, and it is not clear that procedures could be studied in a way analogous to regulation of drugs and devices. In addition, FDA has limited contacts with clinical researchers who could conduct the requisite studies.

Alternative C-4: A new agency could be developed in HEW to fund and conduct efficacy and safety testing, This option incorporates recognition of the fact that the function requires new staff and funds and an organizational focus, and that it would be difficult to change dramatically the mission of an extant agency. The major problem associated with this alternative is that of developing an entirely new agency, This problem could be partially overcome by assigning experts from existing agencies to the new agency. If a new agency were developed, it also might be an appropriate site for identifying technologies that need assessment. An agency with a vested interest in evaluating efficacy and safety could be expected to be active in identifying candidates for evaluation.

Studies would not have to be federally funded. Under FDA statutes, for example, the greatest expense of testing is borne by the manufacturers. If proof of the efficacy and safety of procedures were required by private and public third-party payers, private funding could support more of this testing. Third-party payers also could fund studies directly; National Blue Cross, for example, has funded a study by the Institute of Medicine on the efficacy of the computed tomography (CT) scanner. If successful, this model probably could be used more often. Much of the current testing of medical and surgical procedures is already supported by private funds, including service funds.

SECTION FOUR: SYNTHESIZING INFORMATION

Merely executing numerous research studies will not solve the problems of assessing the efficacy and safety of medical technologies. More data certainly will be helpful, but gaps in knowledge still will remain. Furthermore, value judgments are an integral part of making decisions of efficacy and safety. For example, the net benefit of a technology includes both efficacy and safety; yet, these two pints of the concept cannot be measured in fully comparable terms (see chapter 4). Value-based decisions must still be made regarding whether the positive benefit (efficacy) justifies the risk. Furthermore, study design and the general validity of research findings will need evaluation.

Many agencies and programs could synthesize information. Examining the literature available on a particular technology could highlight the need for further studies in certain areas. Thus, additional studies could appropriately be conducted by the same program

that identifies technologies needing assessment. Wherever this function is performed, it should be open to the public and other parties of interest; it also should have public and professional visibility.

Alternative D-1: The task of synthesizing information could be undertaken by the same commission that identifies technologies needing evaluation (Alternative B-l). The advantage associated with this option is that such a commission would be involved in developing information as a result of trials it stimulated. One disadvantage of this alternative is that such additional responsibility would necessitate the increased capability of staff and advisory committees. Also, such a commission might have little credibility with the practicing community.

Alternative D-2: The Institute of Medicine could be asked to undertake this task, in addition to identifying candidates for assessment. The same advantages found in Alternative B-2 would also apply here.

Alternative D-3: NIH could undertake the task of synthesizing safety and efficacy information. NIH already has the largest extant activity in this area and has begun to use the mechanism for developing consensus effectively in at least one area. However, NIH has shown little inclination to make judgments that could be used by regulatory agencies. * Perhaps NIH could continue to develop consensus in areas in which little controversy exists and in which consensus could have immediate benefits, such as that of diagnosis and treatment of hypertension.

Alternative D-4: Agencies involved in technology development could also synthesize the information derived from trials. One concomitant disadvantage with this option is the diffusion of the function among numerous agencies. The disadvantages mentioned directly above in Alternative D-3 also would apply.

Alternative D-5: FDA could undertake the performance of this task. It already has extensive experience synthesizing and evaluating information submitted both by drug and device manufacturers and physicians. It also has a mechanism for forming expert committees and using outside consultants which would be desirable and applicable to this alternative. However, FDA is basically a regulatory agency and may not be able to attract the scientists necessary for regulating procedures. Again, FDA's negative image with the practicing community would hamper its work.

Alternative D-6: The Office of Health Practice Assessment (OHPA) could undertake the task. OHPA already makes synthesis decisions for the Medicare program. Given adequate resources and access to appropriate experts, it could accomplish the task of synthesizing safety and efficacy information. However, OHPA currently lacks credibility with the practicing community and lacks expertise and access to the information required to complete the task.

Alternative D-7: A new Federal agency could undertake the entire task, including synthesis (see Alternative B-6).

*NIH does provide some information, in the form of judgments or recommendations, to agencies such as the Food and Drug Administration. However, the 1977 Department of Health, Education, and Welfare technology management study concludes that the needs of regulatory agencies remain generally unfulfilled.

SECTION FIVE: DISSEMINATING INFORMATION

Synthesized information—regardless of how valid, understandable, or relevant—is of little value if it is not disseminated to those individuals and organizations which need it. This task is more complex than it seems. The agency responsible for such dissemination must not only have access to the synthesized efficacy and safety information, and any other relevant information, but also must develop, improve, or expand methods of communication to appropriate parties, identify those parties, evaluate the effects of its actions in terms of information conveyed, and perform other related tasks.

Alternative E-1: NIH could refine and expand its dissemination efforts. That agency is one of the most active in disseminating information; and in addition, it contains the National Library of Medicine. However, NIH is reluctant to expand its role in this area, particularly in regard to practicing physicians and health care delivery-related information, partly because of budgetary constraints.

Alternative E-2: This function could be assigned to the Federal agencies involved in testing or synthesis that already perform the dissemination task to a limited degree. The utility of increasing activities by all those agencies, however, would be qualified by at least three factors: parties in need would receive information from a multitude of sources; the function might require a degree of talent, skill, and technique development that many of the agencies could not attain; and, many of the agencies do not have the necessary contact or credibility with the parties who need the data.

Alternative E-3: A new Federal agency, as described in Alternative B-6, could be given the funds and personnel for this task. A close working relationship with NIH would have to be established,

Alternative E-4: Instead of assigning the task to a new agency, either one created to perform the dissemination task or one created to perform alternative tasks, a new office perhaps could be developed either at the level of the Assistant Secretary for Health, HEW, or within an existing Public Health Service (PHS) agency. Presently, there is no focus within HEW for health professional information dissemination as there is now for consumer information. Placing a new office at the Assistant Secretary level would have the advantage of proximity to the National Center for Health Statistics (NCHS). In addition, it would be at a level high enough for access to information and resources of PHS agencies, particularly NIH. It may also facilitate communication with the Health Care Financing Administration (HCFA). A disadvantage of a new office would be its having to start with little credibility or few contacts with many of the parties who need the information. Also, functional conflicts with NIH would have to be anticipated as in Alternative E-3.

USING INFORMATION

This report has primarily addressed a specific problem: the lack of accessible, reliable information on the safety and efficacy of medical technologies. The mere availability of such information, however, does not assure the efficacy and safety of medical technologies currently in use. The development and dissemination of efficacy and safety information leads to a fifth step, namely, the application of such knowledge.

As illustrated in chapters **3** and *6,* many Federal programs use, or could use, information regarding efficacy and safety, According to health planning legislation, ap-

proval of capital investments depends on establishing "need," and such establishment requires scientific information regarding the health benefit expected from application of a particular technology. Professional Standards Review Organizations (PSROs) that examine services for appropriateness, depend on such information. Federal programs that finance and provide medical care also must make some evaluation of efficacy and safety in determining reimbursement of a particular procedure. All these programs must make decisions based partially on efficacy and safety. These decisions often have been made passively or by default.

The following are intended to serve only as examples of possibilities for using information on efficacy and safety to assist providers and consumers in making informed decisions.

Example 1: Medical and surgical procedures could be subject to regulation. In this option, all procedures would be evaluated for safety and efficacy, and only those approved by an agency such as FDA could be used. Such an approach, while theoretically possible, would be difficult to enforce. Because procedures are developed in many sites and are not products, they cannot be regulated through such measures as controlling advertising and interstate transport. In addition, physicians would undoubtedly resist such regulation. The process would be expensive and could retard innovation.

Example 2: When a new technology shows promise, and when a group responsible for the identification task has judged it worthy of full-scale evaluation, medical centers that have the resources to conduct evaluation studies could be allowed to use the technology. Third-party payers would fund this evaluation on **a** prospective budget basis; they would not pay fee-for-service charges for use of the technology until its efficacy, safety, and indications for use were evaluated. No additional public funds would be required if this option were utilized; yet, private insurance companies would spend less on the testing than they would otherwise spend on reimbursement for unproven procedures. No legislation or regulations would be required, and any provider could offer the technology to anyone willing to pay for it out-of-pocket. To be successful, such a mechanism would need a panel of well-recognized professional experts whose plan for testing the technology would have credibility. The plan would include specified testing sites and conditions of use. A similar mechanism could be used for technologies already in use, but payment would not be withdrawn while they were being tested. Once testing was completed and the technology proved to be relatively unsafe or lacking efficacy, reimbursement for its use could be terminated, or specific conditions for reimbursement could be outlined by third-party payers.

APPENDIXES

Appendix A

DEVELOPMENT AND DIFFUSION OF MEDICAL TECHNOLOGIES*

This appendix describes the nature of medical technologies, offers a model of their diffusion, and considers the place of efficacy and safety assessment in the diffusion process. The analysis also reveals the importance of information on efficacy and safety and demonstrates the possibility of making the assessment of safety and efficacy an integral part of the development of medical technologies.

THE NATURE OF MEDICAL TECHNOLOGIES

Medical technologies are of many different types and serve a variety of functions. Nonetheless, they can be classified into sets. Schemes of classification can help in evaluating the efficacy and safety of a particular technology and in judging new technologies on the basis of previous experience or evaluation *(223,277).*

A useful system for classifying medical technologies distinguishes these technologies according to two dimensions—medical purpose and physical nature (3.54). Each of these two dimensions can be broken down further as follows.

Medical Purpose: 1) A *diagnostic* technology helps in determining what disease processes occur in a patient; 2) A *preventive* technology protects an individual from disease; 3) A *therapeutic* or *rehabilitative* technology relieves an individual from disease and its effects (therapeutic technologies can be further divided into those few technologies that cure disease and the many technologies that give symptomatic relief, but do not alter the underlying disease process); 4) An *organizational* or *administrative* technology is used in management and administration to ensure that health care is delivered as effectively as possible; and 5) A *supportive* technology is used to provide patients, especially those in hospitals, with needed services (e.g., hospital beds and food services).

Physical Nature: 1) A *technique* is a purposive application of skills or knowledge, or both, by a health care provider to a patient; 2) A *drug* is any chemical or biological substance that may be applied to, ingested by, or injected into humans in order to prevent, treat, or diagnose disease or other medical conditions; 3) A *device* is any physical item, excluding drugs, used in medical care, and may range from a machine requiring large capital investment to a small instrument or implement; and 4) A *procedure* is a combination, often quite complex, of provider skills or abilities with drugs, devices, or both *(354).*

Drugs and devices are products; procedures, on the other hand, are utilization of a product or products according to the knowledge or skills of a medical care provider. In some cases, the drugs or devices involved are not predominant factors in a procedure. In-

*A more detailed discussion of these issues maybe found in reference (354).

stead, the technique or the provider performing the procedure are most important. A surgical procedure, for example, involves the use of scalpels, clamps, and anti-infection drugs; the key to the procedure, however, is the surgeon's actions. The case of coronary artery surgery in chapter 3 illustrates this point: mortality for such surgery ranges from 0.8 to 12 percent, a very large range in which the skill of the surgeon performing the surgery is clearly a key factor.

HOW MEDICAL TECHNOLOGIES ARE DEVELOPED AND DIFFUSED

The development, diffusion, and use of medical technologies is a process that has been described as including at least seven steps. *

1. Discovery, through research, of new knowledge, and relation of this knowledge to the existing knowledge base;
2. Translation of new knowledge, through applied research, into new technology, and development of a strategy for moving the technology into the health care system;
3. Evaluation of the safety and efficacy of new technology through such means as controlled clinical trials;
4. Development and operation of demonstration and control programs to demonstrate feasibility for widespread use;
5. Diffusion of the new technology, beginning with the trials and demonstrations and continuing through a process of increasing acceptance into medical practice;
6. Education of the professional and lay communities in use of the new technology; and
7. Skillful and balanced application of the new developments to the population.

This sequence of technology development and use is attractive because it offers a logical, linear model for understanding the development process. The model highlights the fact that it it usually possible to identify a medical innovation prior to widespread diffusion, and thus test it in advance for safety and efficacy.

But medical technologies, like others, in fact emerge from a process that is far less systematic and certainly less linear than that which this model implies (345). An additional weakness of this model is that it does not acknowledge the importance of epidemiological research (39). Epidemiological methods have been used in testing efficacy and safety of medical technologies and have led to advances in the prevention and control of disease. The causes of such diseases as cholera, scurvy, and lung cancer have been identified through epidemiological research; epidemiological data have made control programs possible. For example, epidemiological data have shown that cigarette smoking is the major cause of lung cancer, and thus, as noted in the case study in chapter 3, lung cancer is almost totally preventable. Yet basic research has not discovered the mechanism by which cigarette smoking causes cancer **(39).**

Once a technology has been developed through the complex of activities referred to as "basic or fundamental research" and "applied research, " it usually must be tested on

● Modified from reference (392).

human subjects. This area of clinical investigation and testing encompasses a range of activities from first human use to large-scale clinical trials in patients. Occasionally, the first human use of a new technology is spectacularly successful, as it was in the case of the cardiac pacemaker (354). More often, however, it is not, and modifications in the technology are required. After a new technology is shown to be useful in scattered clinical experiments, organized trials may be carried out; increasingly these are controlled clinical trials. The issue of testing for efficacy and safety by the use of clinical trials is discussed in chapter 4.

After human trials have been conducted, and in some cases, before adequate trials are completed, diffusion and adoption of the technology takes place. If clinical trials of a new technology are promising, Government-supported demonstration projects may be organized to show that a technology which is efficacious under controlled clinical conditions is also useful in the community, where social, economic, and other factors may modify its impact. Usually, however, practitioners are persuaded to adopt new developments through less formal channels (79).

Extensive work in primarily nonmedical areas has shown that the diffusion of technology usually follows a sigmoid ("S" shaped) curve in which the rate of adoption accelerates as time goes on **(289).** Diffusion of some medical technologies also follows this curve. A slow initial diffusion rate often is interpreted as an indication of caution on the part of potential users, but in fact may also reflect poor communication between sellers and buyers and among buyers. Those who accept the new technology soonest are referred to as innovators. Early adopters and late adopters account for subsequent diffusions (187,289).

Not all medical technologies follow the diffusion pattern of the sigmoid curve. One major type of departure from the standard model occurs when diffusion reaches a high rate soon after the technology becomes available. This pattern has been referred to as the "desperation reaction model" *(407).* Initial rapid diffusion seems to occur in the absence of evidence of efficacy or safety because of a lack of a suitable alternative technology combined with desperation on the part of patients and of providers responsible for treatment. Later, however, the results of clinical tests and experience begin to influence physician's behavior. If the results are positive, diffusion of the new technology may continue rapidly. More ambiguous results may give rise to physician caution, possibly slowing diffusion. When later evidence is negative, use of the new technology may decline.

Whatever its initial pattern of diffusion, a technology may be partially or completely abandoned if it proves to be of little use clinically. Medicine is replete with examples of procedures fallen out of use.

Both private industry and the Federal Government invest large sums of money in the development of medical technology. The applied research leading to new pharmaceuticals occurs primarily in the drug industry itself. Likewise, most of the research and development leading to medical devices or equipment takes place among manufacturers of medical devices.

Drugs, devices, and procedures in general, especially technique-dependent procedures, pose quite different problems for the evaluation of safety and efficacy. Drugs can be assessed for chemical purity and often have effects that can be tested in the laboratory and can also be tested appropriately in animals. Medical devices, resting on a solid theoretical basis of science in electronics and physics, also may be evaluated by methods not involving human subjects. Procedures, however, involve the use of human skills. The efficacy of a procedure depends on the skill of the provider carrying out the pro-

cedure. Furthermore, drugs and devices are developed in more or less well-defined sites, while medical procedures are developed in many settings. Also, established drugs and devices are often used in an entirely new procedure. For example, the Food and Drug Administration (FDA) may certify anticoagulants as efficacious ("doing what they purport to do") in preventing the coagulation of blood. Such certification, however, does not establish whether use of anticoagulants is an efficacious procedure for the treatment of myocardial infarction or stroke.

Medical procedures, such as surgery, that depend primarily on provider skills are complex and have a correspondingly complex development. New procedures of this type are often developed and tested in hospitals, many of which are university affiliated. Support for their early use and testing often comes from Federal research funds, but considerable funding also comes from service funds, that is, payment for medical services. Chemotherapy for lung cancer (see chapter 3) is an example of an experimental procedure that is often covered by insurance programs.

Appendix B

METHOD OF THE STUDY

Studies in the Office of Technology Assessment (OTA) are frequently done with the assistance of an advisory panel of experts. Panel members suggest source materials and subject areas, assist in data collection and interpretation, review staff drafts for accuracy and validity, suggest conclusions based on the facts, discuss alternatives for the consideration of Congress, and give arguments for and against specific alternatives. The panel, however, does not determine the content of the report and is not responsible for the conclusions and options,

An advisory panel of experts was formed for the study of efficacy and safety of medical technology. Dr. Lester Breslow was named panel chairman, With the help of Dr. Breslow, other panel members then were selected to represent a wide range of disciplines, viewpoints, and expertise. Two members of the OTA Health Advisory Committee, who had expressed particular interest in this study, were named to the panel.

The first meeting of the panel was held in Washington, D. C., on October **26, 1976.** At this meeting, the panel considered the work plan prepared by the staff. The panel endorsed the use of specific case studies of medical technologies to illustrate the benefits and problems involved in assessing the efficacy and safety of medical technologies. The panel also discussed the concepts of efficacy and safety.

After the October meeting, all panel members submitted lists of technologies for the proposed cases. Staff developed criteria for selection of the final list of cases. These criteria were designed to include:

1. Examples of types of technology by function (preventive, diagnostic, and therapeutic and rehabilitative);
2. Examples of different stages of development and diffusion (not yet diffused, experimental or pilot, established in medical care, abandoned);
3. Examples from different areas of medicine (such as general medical practice, pediatrics, obstetrics, and surgery);
4. Examples addressing medical problems that are important because of their high frequency or significant impacts;
5. Examples with associated high costs;
6. Examples of technologies in widespread use; and
7. Examples with sufficient evaluable literature.

Based on the chosen criteria and panel suggestions, *16* cases were selected, and the literature on each was reviewed. (Case 17 was added during 1978.)

The second meeting of the panel was held in Washington, D, C., on December **10, 1976.** At this meeting, the panel reviewed a brief precis on each of the suggested cases. The panel made several suggestions concerning selection of cases, corrected mistakes of fact and interpretation in the case descriptions and suggested additional references. Fur-

ther, the panel reviewed a staff paper on methods for evaluating the efficacy and safety of medical technologies, which was the basis for chapter 4 of this report. Two panel members agreed to develop cases for the final report, and one panel member agreed to develop a brief paper on private sector activities,

After the second meeting of the panel, data collection activities were intensified. In addition to review of the scientific literature, the staff read many Government and private sector reports. All Government agencies and departments listed by the Office of Management and Budget (OMB) as having health activities were sent a survey asking them to summarize their involvement in efficacy and safety issues. Almost 100 private sector organizations also were sent a survey requesting information about their activities in the areas of efficacy and safety of medical technologies. Finally, officials of a large number of public and private agencies and organizations were interviewed, either in person or by telephone.

The third meeting of the panel was held in Washington, D. C., on February 11, 1977. At this meeting, four guests made comments and answered questions from staff and panel: Dr. Seymour Perry, Special Assistant to the Director, National Institutes of Health (NIH); Dr. Mark Novitch, Deputy Associate Commissioner for Medical Affairs, Food and Drug Administration (FDA); Dr. Michael Goran, Director, Bureau of Quality Assurance; and Dr. Clifton Gaus, Director of Health Insurance Studies, Social Security Administration (SSA). These witnesses also commented on a staff draft concerning the involvement of their agencies in efficacy and safety assessment. During the remainder of the meeting, the panel discussed a staff draft that was the basis for chapter 5 of this report and suggested conclusions and policy alternatives that might result from the study.

From February 11 to March 11, 1977, OTA staff wrote a first complete draft of the report. NIH was particularly helpful in this effort, submitting material on almost all of the selected case studies. The responses to the survey of Government agencies and departments were incorporated in chapter 5.

The final meeting of the panel was held, again in Washington, D. C., on March 11, 1977. The panel reviewed the first draft and offered comments and criticisms.

After the meeting, revised cases were sent to NIH for substantive review and to all agencies of the Department of Health, Education, and Welfare (HEW) for confirmation of their roles as described in the cases. Each case was also reviewed by experts in the private sector.

A second draft of the report was then prepared, and in May *1977* was sent *to* the study advisory panel, to the Health Advisory Committee, to the OTA Technology Assessment Advisory Council, and to approximately 100 individuals both within and outside the Federal Government, including officials of Government agencies described in the report.

Changes in staff and time devoted to preparation of other OTA reports, particularly, Policy *Implications of CT Scanners,* delayed work on the third draft of this report, except for the incorporation of comments on the second draft.

The third, and final, draft was prepared during spring of 1978. This draft was reviewed by the study advisory panel, the Health Advisory Committee, and by approximately 80 additional individuals and organizations from both within and outside of Government. Also, several of the cases were revised by contract. All cases were again reviewed by specialists in the particular subject areas. The Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA); the National Center for Health Services Re-

search (NCHSR); and, again, the NIH were particularly helpful in their reviews of substantive material. The final report was written in accordance with the comments and suggestions provided.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Acute Abdominal Conditions Research Group, "Some Problems in Nonoperative Treatment of Acute Appendicitis," *Chin. Med. J.* 2:21, 1974.
2. Advisory Committee on the Biological Effects of Ionizing Radiations, *The Effects on Populations of Exposure to Low Levels of Ionizing Radiation*, Washington, D. C.: National Academy of Sciences, 1972.
3. Aisenberg, A., and Quasi, R,, "Improved Survival in Hodgkin's Disease," *Cancer* 37:2423, 1976.
4. Alliance for Engineering in Medicine and Biology, *System Design of a Clinical Facility for Diagnostic Ultrasound*, Technical Report N-1977-2, September 1977.
5. Alliance for Engineering in Medicine and Biology, *Summary* Guidelines *for* Technology *Procurement in Health Care Institutions*, Technical Report N-1977-I, September 1977.
6. American Biology Council, *Contributions of the Biological Sciences to Human Welfare: Fed. Proc. (Vol. 31), 1972.*
7. American Cancer Society, *'76 Cancer Facts and Figures, 1975.*
8. American Cancer Society, *'78 Cancer Facts and Figures, 1977.*
9. American Medical Association, *The Best of Law and Medicine 70/73*, Chicago: AMA, 1974 (Reprinted from *J A)) / Med. A).*
10. Anderson, B.; Korner, B.; and Ostergaard, A. H., "Topical Ampicillin Against Wound Infection After Colorectal Surgery," *Anti. Surg.* 176:129, 1972.
11. Appelgate, J.; Haverkamp, A. D.; Orleans, M.; et al., "Electronic Fetal Monitoring: Implications for Obstetrical Nursing," presented to Nursing Research, 1978.
12. Asano, Y,; Nakayama, H.; Takehiko, Y.; et al., "Protective Efficacy of Vaccination in Children in Four Episodes of Natural Varicella and Zoster in the Ward," *Pediatrics* 59:3, *1977.*
13. Asano, Y.; Nakayama, H.; Yazaki, T.; et al., "Protective Efficacy of Vaccination in Children in Four Episodes of Natural Varicella and Zoster in the Ward," *Pediatrics* 59:8, *1977.*
14. Ashford, N. A.; Butler, S.E.; and Zolt, E. M., "Comment on Drug Regulation and Innovation in the Pharmaceutical Industry" (draft), presented to the Department of Health, Education, and Welfare Review Panel on New Drug Regulation, Jan. 10, 1977.
15. Ayers, W'. R., "Cost and Technology: End-Point\ in Medical Technology Assessment," *Clin. Eng.* 5(6): 1, 1977.
16. Bailar, 1. C., "Mammography: A Time for Caution," *J. Am. Med. A.* 237:997, 1977.
17. Bannerman, R, M.; Gillick, D.; Van Coevering, R.; et al., "Amniocentesis and Educationl Attainmerit," *N. Eng. J. Lied*, 297:449, 1977.
18. Banta, H. D., and Sanes, J. R., "Assessing the Social Impacts of Medical Technologies," *J. Comm. Health* 3:245, 1978.
19. Barach, A. L., "The Indiscriminate Use of IPPB," *J. Am. Med. A.* 231:1141, 1967.
20. Baram, M., "Medical Device Legislation and the Development and Diffusion of Health Technology," paper presented at the Boston University Conference on Health Care Technology and Quality of Care, November 1976.
21. Barber, B., "The Ethics of Experimentation With Human Subjects," *Sci. Am.* 234:25, 1976.
22. Barker, K.; Graham, N. G.; Mason, M. C.; et al., "The Relative Significance of Preoperative Oral Antibiotics, Mechanical Bowel Preparation, and Preoperative Peritoneal Contamination in the Avoidance of Sepsis After Radical Surgery for Ulcerative Colitis and Crohn's Disease of the Large Bowel," *Brit. J. Surg.* 58:270, 1971.
23. Barnes, B. A., "Discarded Operations: Surgical Innovation by Trial and Error," in *Costs, Risks, and Benefits of Surgery*, edited by J. P. Bunker, et al., New York: Oxford University Press, 1977.
24. Baroon, S., and Wolfe, H., *Measuring the Effectiveness of Medical Decisions*, Springfield, Ill. Charles C Thomas, 1972.

25. Barton, G. G., et al., "Respiratory Care Warrants Studies for Cost Effectiveness," *J. Am. Hosp. A.* 49:61, *1975.*

26. Beard, R. W., and Edington, P. T., "Should Fetal pH Be Used in the Diagnosis of Fetal Asphyxia?," in *Preventability of Perinatal Injury, New* York: Alan R. Liss, Inc., 1975.

27. Beard, R. W.; Edington, P. T.; and Sibanda, J., "The Effects of Routine Intrapartum Monitoring on Clinical Practice," *Contr. Gynec. Obstet.* 3:14, 1977.

28. Bell, B. S., and Loop, J. W., "The Utility and Futility of Radiographic Skull Examination for Trauma," N. *Eng. J. Med.* 284:236, *1971.*

29. Benin, L., Commissioner of Health of the City of New York, memorandum addressed to all hospitals in New York, Jan. 3, 1975.

30. Benefield, J., et al., "Current and Future Concepts of Lung Cancer," Ann. *of Intern.* Meal, 83:93, *1975.*

31. Ben-Yishay, Y.; Diner, L.; Warga, C.; *et* al., "The Alleviation of Cognitive and Functional Impairments in Senility by Hyperbaric Oxygenation Combined With Systematic Cuing," in *Fifth* International *Hyperbaric Conference* (Vol. 1), edited by W.G. Trapp, et al., Burnaby, Canada: Simon Fraser University Press, 1974.

32. Bleck, E. E.; Duckworth, N.; and Hunter, N., *Atlas of Plastic Cast Techniques,* Chicago: Year Book Medical Publishers, Inc., 1974.

33. Blue Cross/Blue Shield, "Will There Come a Time When It's Just Too Costly to Save a Life?," *Time Magazine* (advertisement), 1976.

34. Blumberg, B. D.; Golbus, M. S.; and Hanson, K, H., "'The Psychological Sequelae of Abortion Performed for a Genetic Indication, 'r *Am. J. Obstet. Gynecol.* 122:799, 1975.

35. Bombardier, C., et al., "Socioeconomic Factors Affecting the Utilization of Surgical Operations," N, *Eng. J. Med.* 297:699, 1977.

36. Bonadonna, G.; Brusamolino, E.; Balagussa, P.: et al., "Combination Chemotherapy as an Adjuvant Treatment in Operable Breast Cancer," N. *Eng. J. Med.* 294:405, 1976.

37. Boyes, D., cited by M. Sloan, National Cancer Institute, in a personal communication, 1976.

38. Boyle, E.; Aparicio, A.; Canosa, F.; et al., "Hyperbaric Oxygen and Acetazolamide in the Treatment of Senile Cognitive Functions," in *Fifth International Hyperbaric Conference* (Vol. 1), edited by W.G. Trapp, et al., Burnaby, Canada: Simon Fraser University Press, 1974.

39. Breslow, L., testimony, "Basic Issues in Biomedical Research," before the U.S. Senate Committee on Labor and Public Welfare, Subcommittee on Health, June 17, 1976.

40. Brook, R. H., "Quality of Care Assessment: Policy Relevant Issues." *Pol. Sci.* 5:317, 1974.

41. Brook, R.; Avery, A.; Greenfield, S.; et al., *Quality of Medical Care Assessment Using Health Outcome Measures; An Overview of the Method,* prepared for the National Center for Health Services Research, Santa Monica, Calif.: RAND Corp., 1976.

42. Brooks, S. M., *McBurney's Point: The Story of Appendicitis, New* York: A.S. Barnes and Co., 1969.

43. Brown, B. W., and Hollander, M., *Statistics: .4 Biomedical introduction, New* York: John Wiley and Sons, 1977.

44. Brown, C. J., "Deaths From Appendicitis," Med. *J. Aust.* 2:407, 1934.

45. Brown, S. S., "A Case Study of Medical Efficacy: Tonsillectomy" (manuscript), Washington, D. C.: National Academy of Sciences, Institute of Medicine, 1975.

46. Brownfield, R. L., and Ives, E. R., "Creating a Data Base for the Laboratory Universe," *Lab. Mgt.* 14:22, 1976. (1975 figures furnished by Mr. Brownfield in a personal communication.)

47. Brudnell, M.; Cox, B. S.; and Taylor, C. W., "The Management of Dysplasia Carcinoma in Situ and Microcarcinoma of the Cervix,"]. *Obstet. Gynecol. Br. Comw.* 80:673, 1973.

48. Brunell, P. A., "Protection Against Varicella," *Pediatrics* 59:1, 1977.

49. Bunker, J.P. "Elective Hysterectomy: Pro and Con," *N. Eng. J. Med.* 295:264, 1976.

50. Bunker, J. P., "Surgical Manpower: A Comparison of Operations and Surgeons in the United States and in England and Wales, "N. *Eng. J. Med.* 282:135, 1970.

51. Bunker, J. P.; Barnes, B. A.; Mosteller, F., eds., *Costs, Risks, and Benefits of Surgery, New* York: Oxford University Press, 1977.

52. Bunker, J. P., and Brown, B. W., "The Physician-Patient as an Informed Consumer of Surgical Services," N. *Eng. J.* Med. 290:1051, 1974.

53. Bunker, J. P.; Hinkley, D.; and McDermott, W. V., "Surgical Innovation and Its Evaluation," *Science* 200:937, 1978.

54. Bull, J. P., "The Historical Development of Clinical Therapeutic Trials,"]. *Chron. Dis.* 218, September 1959.

55. Byar, D.; Simon, R.; Friedewald, W.; et al., "Randomized Clinical Trials," N. *Eng. J.* Med. 295:74, 1976.

56. Caldeyro-Barcia, R., et al., "Fetal Monitoring in Labor," in *Maternal and Child Health Practices: Problems, Resources and Methods of Delivery*, Montevideo, Uruguay: Latin American Center for Perinatology and Human Development, 1974 (est.).

57. Califano, J. A., "Remarks Before the Annual Meeting of the American Federation for Clinical Research," San Francisco, Calif., Apr. 29, 1978.

58. Carbone, P., "Lung Cancer: Perspectives and Prospects," *Ann. Intern. Med.* 83:93, *1975.*

59. Carnegie-Mellon University, Department of Engineering and Public Policy, *Regulation of Orthopedic Surgical Implants: An Investigation of the Effects of the 1976 Medical Device Amendments*, Pittsburgh: Carnegie-Mellon University, Mar. 4, 1977.

60. Cassell, E. J., "The Case of Hypertension," manuscript prepared for the Commission on Protection of Human Subjects of Biomedical and Behavioral Research, June 1976.

61. Center for Development Technology, "A State of the Art Review of the Effects of Regulation on Technological Innovation in the Chemical and Allied Products Industries," presented to the National Science Foundation, February 1975.

62. "Cervical Cancer Screening Programs: The Walton Report," *Can. Med. Assn. J.* 114:2, *1974.*

63. Cetrulo, C. L., and Freeman, R., "Problems and Risks of Fetal Monitoring," in *Risks in the Practice of Modern Obstetrics*, edited by S. Aladjem, St. Louis: C.V. Mosby Co., 1975.

64. Chalmers, T. C., "Settling the UGDP Controversy,"]. *Am. Med. A.* 231:624, *1975.*

65. Chalmers, T.; Block, J.; and Lee, S., "Controlled Studies in Clinical Cancer Research," N. *Eng. J. Med.* 287:75, *1972.*

66. Chalmers, T. C.; Matta, R. J.; Smith, H.; et al., "Evidence Favoring the Use of Anticoagulants in the Hospital Phase of Acute Myocardial Infarction," N. *Eng. J. Med.* 297:1091, *1977.*

67. Charlson, M. E., and Feinstein, A. R., "The Auxometric Dimension: A New Method for Using Rate of Growth in Prognostic Staging of Breast Cancer,"). *Am. Med. A.* 228:180, *1974.*

68. Christopherson, W. M., et al., "Cervical Cancer Control," *Cancer* 38:1357, *1976.*

69. Cibils, L. A., "Clinical Significance of Fetal Heart Rate Patterns During Labor: I. Baseline Patterns," *Am. J. Obstet. Gynecol.* 125:290, *1976.*

70. Clark, R. L.; Copeland, M. M.; Egan, R. L.; et al., "Reproducibility of the Technic of Mammography (Egan) for Cancer of the Breast," *Am, J. Surg.* 109;127, 1965.

71. Clymer, H. A., "The Economic and Regulatory Climate: U.S. and Overseas Trends," in *Drug Development and Marketing, edited* by R. Helms, Washington, D. C.: American Enterprise Institute, 1975.

72. Cochrane, A., *Effectiveness and Efficiency* (Nuffield Provincial Hospitals Trust), London: Burgess and Son, Ltd., 1972.

73. Cohen, L. S., "Coronary Artery Surgery: Past, Present, and Future," Corm. *Med.* 40:509, *1976.*

74. Cohen, M., "Lung Cancer: A Status Report,"]. *Natl. Cane, Inst.* 55:505, 1975.

75. Coldrey, E., "Five Years of Conservative Treatment of Acute Appendicitis,"]. *Int.* Coil. *Surg.* 32:255, 1959.

76. Cole, P., "Epidemiology of Breast Cancer: An Overview," in *Report to the Profession: Breast Cancer*, Washington, D. C.: National Institutes of Health, National Cancer Institute, 1974.

77. Cole, P., remarks, a part of "Elective Hysterectomy: Pro and Con," *N. Eng. J. Med.* 295:264, *1976.*

78. *Cole*, P., and Berlin, J., "Elective Hysterectomy" (manuscript), 1976.

79. Coleman, J. S.; Katz, E.; and Menzel, H., *Medical Innovation: A Diffusion Study*, Indianapolis, Ind.: Bobbs-Merrill, 1966.

33-496 O - 78 - 10

80. Collins, S. D.; Lehmann, J. L.; and Trantham, K., "Surgical Experience in Selected Areas of the United States," *Publ. Health Monogr.* (Vol. 38), 1956,

81. Colombo, F.; Shapiro, S.; Slone, D.; et al., *Epidemiological Evaluation of Drugs*, Amsterdam: Elsevier/North Holland Biomedical Press, 1977.

82. Committee on Biomedical Research in the Veterans Administration, *Biomedical Research in the Veterans Administration*, Washington, D. C.: National Research Council, 1977.

83. Comroe, J. H., "The Road From Research to New Diagnosis and Therapy," Science 200:931, 1978.

84. Condon, R. E., "Rational Use of Prophylactic Antibiotics in Gastrointestinal Surgery," *Surg. Cl. N. A.* 55:1309, *1975.*

85. Cooper, J. D., *Decision-Making on the Efficacy and Safety* of *Drugs: Philosophy and Technology of Drug Assessment* (Vol. I), Washington, D. C.: Interdisciplinary Communication Associates, Inc., 1971.

86. Cooper, J. D., *The Philosophy of Evidence*, Washington, D. C.: Interdisciplinary Communication Associates, Inc., 1972.

87. Coppelson, L. W., and Brown, B., "Estimates of the Screening Error Rate From the Observed Detection Rates in Repeated Cervical Cytology," *Am. J. Obstet Gynecol.* 119:953, 1974.

88. Cornfield, J., "Recent Methodological Contributions to Clinical Trials," *J. Am, Med. A.* 104(4):408, *1976.*

89. Costan, B., et al., "Current Concepts of Leukemia and Lymphoma: Etiology, Pathogenesis, and Therapy," *Ann.* Intern. *Med.* 85:351, *1976.*

90. Cousins, N., "Anatomy of an Illness (as Perceived by the Patient)," N. *Eng. J. Med.* 295:1458, *1976.*

91. Cramer, D. W., "The Role of Cervical Cytology in the Declining Morbidity and Mortality of Cervical Cancer," *Cancer* 34:2018, *1974.*

92. Crile, G., "Peritonitis of Appendiceal Origin Treated With Massive Doses of Penicillin," *Surg. Gynecol. Obstet.* 83:150, *1946.*

93. Crile, G., and Fulton, J. R., "Appendicitis-With Emphasis on the Use of Penicillin," *U.S. Naval Med. Bull.* 45:464, *1945.*

94. Cromwell, J.; Ginsburg, P.; Hamilton, D.; et al., *Incentives and Decisions Underlying Hospitals' Adoption and Utilization of Major Capital Equipment*, prepared for the National Center for Health Services Research, Boston: Abt Associates, 1975.

95. Culliton, B. J., "Breast Cancer: Second Thoughts About Routine Mammography," *Science* 193:55, *1976.*

96. Curry, C. L., "Hypertension in the Black Population," paper presented at the meeting of the Institute of Medicine, Oct. 27, 1976.

97. Cutler, S. J., cd., *End Results in Cancer*, Washington, D. C.: U.S. Govt. Printing Office, 1968.

98. Davidson, S., cd., *PSRO Util[illegible]ation and Audit in Patient Care*, St. Louis: The C.V. Mosby Co., 1976.

99. Desser, R., and Ultmann, J., "Risk of Severe Infection in Patients With Hodgkin's Disease of Lymphoma After Diagnostic Laparatomy and Splenectomy," *Ann, Intern. Med.* 77:143, *1972.*

100. Detmer, D. E., and Tyson, T. J., "Delivery of Surgical Care: Inferences Based on Hospital Discharge Abstract Data," *Surg. Forum* 27:460,1976.

101. DeVita, V., et al., "Combination Chemotherapy in the Treatment of Advanced Hodgkin's Disease," *Ann. intern. Med.* 73:881, *1970.*

102. Dickinson, L., "Evaluation of the Effectiveness of Cytologic Screening for Cervical Cancer: 111. Cost-Benefit Analysis," *Mayo Clin. P.* 47:550, 1972.

103. Dodge, C. H., "Medical Devices: A Brief Review of Legislation and Issues Related to Regulation{" Washington, D. C.: Congressional Research Service, Library of Congress, 1974.

104. Donahue, V. C., remarks, a part of "Elective Hysterectomy: Pro and Con." N. *Eng. J.* Med. 295:264, 1976.

10s. Doresen, N., *Review Panel on New Drug Regulation*, Washington, D. C.: Department of Health, Education, and Welfare, May 1977.

106. Durant, K. R., et al., "Comparison of Treatment Policies in Inoperable Bronchial Carcinoma," *Lancet* 1:715, 1971.

107. Durant, R. E., "Treatment of Hodgkin's Disease: With What and By Whom, " *Ann. Intern. Med.* 73:1033, *1970.*

108. Ebert, R. V., "Anticoagulants in Acute Myocardial Infarction: Results of a Cooperative Clinical Trial, " *J. Am. Med. A.* 225:724, *1973.*

109. Ebert, R. V., "Use of Anticoagulants in Acute Myocardial Infarction, " *Circulation* 45:903, *1972.*

110. Edington, P. T.; Sibanda, J.; and Beard, R. W., "Influence on Clinical Practice of Routine Intrapartum Fetal Monitoring, " *Brit. J. Prev. Soc. Med.* 3:341, 1975.

111. Edwards, A. E., and Hart, G. M., "Hyperbaric Oxygenation and the Cognitive Functioning of the Aged, " *J. Am. Geriatr. Soc.* 22:376, *1974.*

112. Effler, D. B., "Myocardial Revascularization: Current State of the Art, " *Am. J. Cardiol.* 36:849, *1975.*

113. Elstein, A. S., "Clinical Judgment: Psychological Research and Medical Practice, " *Science* 194:696, *1976.*

114. Everett, M. T.; Brogan, T. C.; and Nettleton, J., "The Place of Antibiotics in Colonic Surgery: A Clinical Study, " *Brit. J. Surg.* 56:679, 1969.

115. Feinstein, A. R., "Clinical Biostatistics: XXV. A Survey of Statistical Procedures in General Medical Journals, " *Clin. Pharmacol. Ther.* 15:97, *1974.*

116. Fineberg, H. V.; Bauman, R.; and Sosman, M., "Computerized Cranial Tomography: Effect on Diagnostic and Therapeutic Plans, "]. *Am. Med. A.* 238:224, *1977.*

117. Fink, D. J., and Holleb, A. I., letter to project directors and coordinators of the Breast Cancer Detection Demonstration Projects, Aug. 23, 1976.

118. Fitz, R. H., "Perforating Inflammation of the Vermiform Appendix, " *Trans. Assoc.* Am. *Phys.* p. 107, 1886.

119. Fletcher, J. C., "Prenatal Diagnosis: Ethical Issues, " in *Encyclopedia of Bioethics* (in press).

120. Fowler, R., "Childhood Mortality From Acute Appendicitis: The Impact of Antibiotics, " *Med. J. Aust.* 2:1001, *1971.*

121. Fox, R. C., and Swazey, J. P., *The Courage to Fail,* Chicago: University of Chicago Press, 1974.

122. Franklin, B., "Some Account of the Success of Inoculation for the Smallpox in England and America, " London: Printed by W. Strahan, 1759, from the Library of the U.S. Surgeon General's Office.

123. Franklin, W., and Lowell, F. C., "Unapproved Drugs in the Practice of Mediciner " *N. Eng. J.* Med. 292:1075, 1975.

124. Frazier, H. S., and Hiatt, H. H., "Evaluation of Medical Practices, " Science 200:875, 1978.

125. Friedson, E., *Doctoring Together: A Study* of *Professional Social Control, New* York: Elsevier, 1975.

126. Frohlich, E. D., "Hypertension 1973: Treatment—Why and How, " *Ann. Intern. Meal,* 78:717, *1973.*

127. Fry, J., "Antibiotics in Acute Tonsillitis and Acute Otitis Media, " *Br. Med. J.* 2:883, *1958.*

128. Fuchs, F., and Riis, P., "Antenatal Sex Determination, " *Nature* 177:330, *1956.*

129. Fuchs, V. R., *Who Shall Live?, New* York: Basic Books, 1974.

130. Gabert, H. A., and Stenchever, M. A., "Continuous Electronic Monitoring of Fetal Heart Rate During Labor, " *Am, J. Obstet. Gynecol. 115:919,* 1973.

131. Gassner, C. B., and Ledger, W. J., "The Relationship of Hospital-Acquired Maternal Infection to Invasic Intrapartum Monitoring Techniques, " *Am. J. Obstet. Gynecol.* 126:33, 1976.

132. Gaus, C. R., testimony of the Social Security Administration before the President's Biomedical Research Panel, Sept. 29, 1975.

133. Gehan, E., and Freireich, E., "Nonrandomized Control in Cancer Clinical Trials, " *N. Eng. J. Med.* 290:198, 1974.

134. Gershon, A., "Varicella-Zoster Virus Infection, " prepared for the Institute of Allergy and Infectious Disease, Acute Viral Infections Panel, Virology Task Force, November 1976.

135. Gibbs, R. S.; Istea, H. M.; and Read, J. A., "The Effect of Internal Fetal Monitoring on Maternal Infection Following Cesarean Section, " *Obstet. Gyn.* 48:653, 1976.

136. Gibson, R. M., and Mueller, M. S., "National Health Expenditure Highlights, Fiscal Year 1976," *Research and Statistics Note No.* 27, Social Security Administration, Dec. 22, 1976.

137. Gifford, R. H., and Feinstein, A. R., "A Critique of Methodology in Studies of Anticoagulant Therapy for Acute Myocardial Infarction, " *N. Eng. J. Med.* 280:351, 1969.

138. Gilbert, J. P.; McPeek, B.; and Mosteller, F., "Progress in Surgery and Anesthesia: An Evaluation of Innovative Therapy" (manuscript), 1976.

139. Gilmore, O. J. A.; Brodribb, A. J.; Browett, J. P.; et al., "Appendicitis and Mimicking Conditions, " *Lancet* 2:421, *1975.*

140. Godber, G., "Constraints Upon the Application of Medical Advances, " *Proc. Roy. Soc. Med.* 67:1274, 1974.

141. Goldberg, C., "The 'Conservative' Approach in Surgical Decisions and Intervention: Evidence of Its Successfulness and the Implications for Health Care Delivery Systems, " Portland, Oreg.: Kaiser Hospital (unpublished).

142. Goldberg, S. C., et al., "Prediction of Relapse in Schizophrenic Outpatients Treated by Drug and Sociotherapy, " *Arch. G. Psyc.* 34:171, 1977.

143. Goldfarb, A. I.; Hochstadt, N. J.; Jacobson, J. H.; et al., "Hyperbaric Oxygen Treatment of Organic Mental Syndrome in Aged Persons,"]. *Gerontol.* 27:212, 1972.

144. Goldhaber, S. Z.; Bloom, B. S.; Sugerbaker, P. H.; et al., "Effects of the Fiberoptic Laparoscope and Colonoscope on Morbidity and Cost, " *Ann. Surg.* 179:160, *1974.*

145. Goodlin, R. C., and Haesslein, H. C., "When Is It Fetal Distress?, " *Am. J. Obstet. Gynecol.* 128:440, 1977.

146. Gordon, G., and Fisher, G. L., eds., *The Diffusion of Medical Technology*, Cambridge, Mass.: Ballinger Publishing Co., 1975.

147. Gordon, R. S., "Clinical Trials," *N. Eng. J. Med.* 298(7):400, *1978.*

148. Grace, N. E.; Muench, H.; and Chalmers, T. C., "The Present State of Shunts for Portal Hypertension in Cirrhosis, " *Gastroenty.* 50:684, *1966.*

149. Green, L. W.; Levine, D. M.; and Deeds, S., "Clinical Trials of Health Education for Hypertensive Outpatients: Design and Baseline Data, " *Prev. Med.* 4:417, 1975.

150. Greene, A. G., "Technology and Social Values, " in *The Management of Health Care*, edited by W.J. Abernathy, et al., Cambridge, Mass.: Ballinger Publishing Co., 1974.

151. Greenberg, B. G., "Conduct of Cooperative Field and Clinical Trials, " *Am. Statistn.* 13:13, June 1959.

152. Griner, P, F., "Treatment of Acute Pulmonary Edema: Conventional or Intensive Care?" *Ann. Intern. Med.* 77:501, *1972.*

153. Grosse, R. N., "Cost-Benefit Analysis of Health Service, " *Ann. Am. Acad. Polit. Soc. Sci.* 399:89, *1972.*

154. Gunby, P., "Meeting Reveals Growing Caution About 'Routine' Thyroidectomies, " *J. Am. Med. A.* 236:2477, *1976.*

155. Guzick, D. S., "Efficacy of Screening for Cervical Cancer: A Review, " *Am. J. Publ. Health* 68:125, *1978.*

156. Gypsona Technique, Hull, England: T.J. Smith and Neshew, Ltd., 19s3.

157. Hammand, K. R., and Adelman, L., "Science, Values and Human Judgment, " *Science* 194:389, *1976.*

158. Haverkamp, A. D., address to the 25th Annual National Health Forum, New York, N. Y., Mar. *23, 1977.*

1s9. Haverkamp, A. D.; Orleans, M.; and Murphy, J. R., "The Differential Effects of Fetal Monitoring on Mothers and Infants, " paper presented before the American Public Health Association, Oct. 31, 1977.

160. Haverkamp, A. D.; Thompson, H. E.; McFee, J. G.; et al., "The Evaluation of Continuous Fetal Heart Rate Monitoring in High-Risk Pregnancy, " *Am. J. Obstet. Gynecol.* 125:310, *1976.*

161. Hellman, S., "Current Studies in Hodgkin's Disease, " *N. Eng. J. Med.* 290:894, *1974.*

162. Hiatt, H.H. "Protecting the Medical Commons: Who Is Responsible?" *N. Eng. J. Med.* 293:235, 1975.

163. Hill, A. B., "The Clinical Trial," *N. Eng. J. Med.* 247:113, *1952.*

164. Hill, A. B., *Principles of Medical Statistics, New* York: Oxford University Press, 1971.

165. Hill, A. B., *Statistical Methods in Clinical and Preventive Medicine*, London: E. & S. Livingston, Ltd., 1962.

166. Hogarty, G. E., et al., "Drug and Sociotherapy in the Aftercare of Schizophrenic Patients, " *Arch. G. Psyc.31* :603, *1974.*

167. Hoekelman, R. A., "Infectious Illness During the First Year of Life, " *Pediatrics* 59:119, *1977.*

168. Hochuli, E.; Eberhard, J.; and Dubler, O., "The Effect of Modern Intensive Monitoring in Obstetrics on Infant Mortality and the Incidence of Hypoxia and Acidosis, " *J.Perinat. Med. 4:78, 1976.*

169. Hen, E. H., and Petrie, R. H., "Clinical Value of Fetal Heart Rate Monitoring, " *Clin.Obstet.Gynecol.* 18:1, *1975.*

170. Horn, J. S., *Away With All Pests, New* York: Modern Reader, 1969.

171. Howie, J. G. R., "Death From Appendicitis and Appendectomy, " *Lancet* 1:1334, *1966.*

172. Howie, V. M., and Ploussard, J. H., "Efficacy of Fixed Combination Separate Components in Otitis Media, " *Clin.Peal.* 11(4):214, *1972.*

173. Hughey, M. J.; Lapata, R. E.; McElin, T, W.; et al., "The Effect of Fetal Monitoring on the Incidence of Cesarean Section, " *Obstet.Gyn.*49:513, *1977.*

174. Humber, J. M., and Almeder, R. F., eds., *Biomedical Ethics* and *the Law, New* York: Plenum Press, 1976.

175. Husain, O. A. N., "Quality Control in Cytological Screening for Cervical Cancer, " *Tumori* 62:303, *1976.*

176. Iglehart, J. K., "Probing Not Only the Cause of Disease But the Effects of Research, " *Natl.Jour.* p. 142, Jan. *22, 1977.*

177. Imai, Y., "Hyperbaric Oxygen (OHP) Therapy in Memory Disturbances, " in *Fifth International Hyperbaric Conference* (Vol. 1), edited by W.G. Trapp, et al., Burnaby, Canada: Simon Fraser University Press, 1974.

178. Ingelfinger, F. J.; Ebert, R. V.; Finland, M.; et al., *Controversy in Internal Medicine,* Philadelphia: W.B. Saunders Co., 1974.

179. Institute of Medicine, "The Efficacy of Medical Care: A Review of Current Evidence on Major Procedures" (unfunded grant proposal), Washington, D. C.: National Academy of Sciences, 1975.

180. Jacobs, E. A.; Alvis, H. J.; and Small, S. M., "Hyperoxygenation: A Central Nervous System Activator?" *J.Geriatr.Psyc.*5:107, 1972.

181. Jacobs, E. A.; Winter, P. M.; Alvis, H. J.; et al., "Hyperoxygenation Effects on Cognitive Functioning in the Aged, " N. *Eng.J.Med.* 281:753, *1969.*

182. James, L. W.; Bowe, E. T.; and Rey, H. R., "The Status of Fetal Monitoring in Decision Making in Patient Management, " in *Controversy in Obstetrics and Gynecology* (Vol. II), edited by D.E. Reid and C.D. Christian, Philadelphia: W.B. Saunders Co., 1974.

183. Jennings, J. E.; Burger, H. E.; and Jacobi, M., "Acute Appendicitis, " *Arch. Surg.*44:896, *1942.*

184. Jergens, M. E.; Morgan, M. R.; and McElroy, C. E., "The Use of Skull and Cervical Spine Radiography in the Evaluation of the Injured Patient" (manuscript).

185. Jones, O. H., "Cesarean Section in Presentday Obstetrics, " *Am. J.Obstet.Gynecol.*126:521, *1976.*

186. Juhl, E.; Christensen, E.; and Tygstrup, N., "The Epidemiology of Gastrointestinal Randomized Clinical Trials, " N. *Eng.J.Med.* 296:20, 1977.

187. Kaluzny, A. D., "Innovation in Health Services, Theoretical Framework and Review of Research, " *Health Serv.Res. 9:101,* 1974.

188. Kaplan, H. S., "The Management of Hodgkin's Disease, " *Cancer 36(* Suppl.):796, *1975.*

189. Karch, F. E., and Lasagna, L., "Adverse Drug Reactions: A Critical Review,"]. *Am, Med. A.* 234:1236, 1975.

190. Kelly, K. M.; Madden, D. A.; Arcarese, J. S.; et al., "The Utilization and Efficacy of Pelvimetry, " *Am. J.Roentgenol. Radium Ther.Nucl.Med.* 125:66, 1975.

191. Kennedy, D., "Animal Testing and Human Risk, " paper presented before the Writers' Seminar, American Cancer Society, Houston, Tex., Apr. 1, 1978.

192. Kennedy, D., "Creative Tension: FDA and Medicine, " *N.Eng.J.Med.* 298(15):846, 1978.

193. Kern, W. H., and Zivolich, M. R., "The Accuracy and Consistency of the Cytologic Classification of Squamous Lesions of the Uterine Cervix, " *ActaCytol.* 21 :519, 1977.

194. Kimball, A. W., "Discussion of 'Recent Methodological Contributions to Clinical Trials', " *Am. J. Epidem.* 104(4):422, *1976.*

19s. Kisch, A, I., et al., "An Epidemiologic Approach to the Study of the Incidence of Surgical Procedures, " *Med. Care* 7:471, *1969.*

196. Klarman, H. E.; Francis, J.; and Rosenthal, G. D., "Cost Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease, " *Med. Care* 6:48, *1968.*

197. Kloster, F.; Kremkau, L.; Rahimtoola, S.; et al., "Prospective Randomized Study of Coronary Bypass Surgery for Chronic Stable Angina, " *Circulation* 52(Suppl. 11):90, 1975.

198. Knatterud, G. L.; Meinert, C. L.; Klimt, C. R.; et al., "Effects of Hypoglycemic Agents on Vascular Complications in Patients With Adult-Onset Diabetes,"]. *Am. Med. A.* 217:777, *1971.*

199. Knox, E. G., "Cervical Cancer, " in Screening *in Medical Care* (Nuffield Provincial Hospitals Trust), London: Oxford University Press, 1968.

200. Kolata, B. B., "Coronary Bypass Surgery: Debate Over Its Benefits, " *Science* 194:1263, 1976.

201. Koran, L., "The Reliability of Clinical Methods, Data, and Judgments" (Part 1 of 2), *N.Eng.J.* Med. 293(13):642, 1975.

202. Koran, L., "The Reliability of Clinical Methods, Data, and Judgments" (Part 2 of 2), *N.Eng.J. Med.* 293:695, *1975.*

203. Krause, E. A., Power *and Illness, New* York: Elsevier, 1977.

204. Kristein, M. M., "Economic Issues in Health Maintenance," *Prev. Med.* (in press).

205. Lacher, M., "Long Survival in Hodgkin's Disease, " *Ann. Intern. Med.* 70:7, *1969.*

206. Lambourne, A., and Lederer, H., "Effects of Observer Variation in Population Screening for Cervical Carcinoma, " *J.Clin. Path.* 26:564, *1973.*

207. Lampert, R. P.; Robinson, D. S.; and Soyka, L. F., "A Critical Look at Oral Decongestants, " *Pediatrics* 55:550, *1975.*

208. Ledger, W. J.; Norman, H.; Gee, C; et al., "Bacteremia on an Obstetric-Gynecologic Service, " *Am. J. Obstet.Gynecol.* 121: 205, 1975.

209. Lee, W. K., and Baggish, M. S., "The Effect of Unselected Intrapartum Fetal Monitoring, " *Obstet.Gyn.* 47:516, 1976.

210. Lelah, T., and Brook, R. H., "Ischemic Heart Disease: Assessing the Quality of Medical Care Using Short-Term Outcome Measures, " in *Quality of Medical Care Assessment* Using Outcome Measures: Eight Disease-Specific Applications, edited by A.D. Avery, et al., Santa Monica, Calif.: RAND Corp., 1976,

211. Lembcke, P. A., "Measuring the Quality of Medical Care Through Vital Statistics Based on Hospital Service Areas: 1. Comparative Study of Appendectomy Rates, " Am. *J.Publ.* Health 42:276, 1952.

212. Lesnick, G. J., "Detection of Breast Cancer in Young Women,"]. *Am. Med. A.* 237:967, *1977.*

213. Lewis, B., National Cancer Institute, National Institutes of Health, personal communication, June 1978.

214. Lewis, C. E., "Variations in the Incidence of Surgery, " *N.Eng.J.* Med. 281:880, 1969.

215. Li, M. C., and Ross, S. T., "Chemoprophylaxis for Patients With Colorectal Cancer: Prospective Study With Five-Year Follow-Up, " *J. Am. Med. A.* 235:2825, 1976.

216. Ling, W.; Charuvastra, V. C.; Kaim, S. C.; et al., "Methadyl Acetate and Methadone as Maintenance Treatment for Heroin Addicts," Arch. G. *Psyc.* 33:709, 1976.

217. Loew, C. R., "Breast Cancer, " in *Screening in Medical Care* (Nuffield Provincial Hospitals Trust), London: Oxford University Press, 1968.

218. Lowrance, W., *Of Acceptable Risk,* Los Altos, Calif.: William Kaufmann, Inc., 1976.

219. Lubic, R. W., "Fetal Electronic Monitoring vs. Home Delivery" (manuscript), Mar. 23, 1977.

220. Lusted, L. B., et al., "Evaluating the Efficacy of Radiologic Procedures by Bayesian Methods: A Progress Report, " in *Models in Metrics for Decision Makers,* edited by K. Snapper, Washington, D. C.: Information Resources Press (in press).

221. Lyon, J. L., and Gardner, J. W., "The Rising Frequency of Hysterectomy: Its Effects on Uterine Cancer Rates, " *Am. J.Epidem.* 105:439, *1977.*

222. MacGregor, J. E., "Evaluation of Mass Screening Programs for Cervical Cancer in N.E. Scotland," *Tumori* 62:287, 1976.

223. McDermott, W., "Evaluating the Physician and His Technology, " *Daedalus*, p. 135, Winter 1977.

224. McKeown, T., "Validation of Screening Procedures, " in *Screening in Medical Care*, (Nuffield Provincial Hospitals Trust) London: Oxford University Press, 1968.

225. McKeown, T., and Knox, E. G., "The Framework Required for Validation of Prescriptive Screening, " in *Screening in Medical Care*, (Nuffield Provincial Hospitals Trust) London: Oxford University Press, 1968.

226. McNeil, B. J., et al., "Cost Effectiveness Calculations in the Diagnosis and Treatment of Hypertensive Renovascular Disease, " *N. Eng. J. Med.* 293:221, *1975.*

227. McPeek, B.; Gilbert, J. P.; and Mosteller, F., "The End Result: Quality of Life, " in Costs, *Risks, and Benefits of Surgery*, edited by J.P. Bunker, et al., New York: Oxford University Press, 1977.

228. McPherson, K., and Fox, M. S., "Treatment of Breast Cancer, " in Costs, *Risks, and Benefits of Surgery*, edited by J.P. Bunker, et al., New York: Oxford University Press, 1977.

229. Macri, J. N.; Weiss, R. R.; Tillit, R.; et al., "Prenatal Diagnosis of Neural Tube Defects, " *J. Am. Med. A.* 236:1251, *1976.*

230. Magee, R. B.; Stowell, J. M.; and Macduffee, R. C., "Appendicitis: An Analysis of 2,117 Cases Treated Surgically in a Community Hospital During a Period of *20* Years, " *Penn. Med.* 70:47, *1967.*

231. Majerus, K. M., "Amniocentesis: Update 1976" (manuscript), prepared for the President's Committee on Mental Retardation.

232. Marx, J. L., "Hypertension: A Complex Disease With Complex Causes, " Science 194:821, 1976.

233. Mather, H. G.; Pearson, N. G.; Read, K. L. Q.; et al., "Acute Myocardial Infarction: Home and Hospital Treatment, " *Brit. Med. J.* 3:334, *1971.*

234. Mathier, V. W., "Surgical Treatment for Stable Angina Pectoris: Prospective Randomized Study, " N. *Eng. J.* Med. 292:709, 1975.

235. Mathur, V. S., and Guinn, G. A., "Prospective Randomized Study of Coronary Bypass Surgery in Stable Angina, " *Circulation* 52(Suppl. I):133, 1975.

236. Maugh, T. H., "Diabetes Therapy: Can New Techniques Halt Complications?" *Science* 190:1281, 1975.

237. Mausner, J. S., and Bahn, A. K., *Epidemiology*, Philadelphia; W.B. Saunders Co., 1974.

238. Mazzi, E.; Herrera, A.; and Bergman, M., "Prevention of Scalp Abscesses Secondary to Fetal Scalp Monitoring, " *J. Peal*, 90:664, *1977.*

239. "Meniere's Disease: Placebo Surgery Seems Effective, " Medical *World News* 19(17):33, *1978.*

240. Miles, R., *The Department of HEW, New* York: Praeger Publishers, Inc., 1974.

241. Miller, A. B.; Lindsay, J.; and Hill, G. S., "Mortality From Cancer of the Uterus in Canada and Its Relationship to Screening for Cancer of the Cervix, " *Int. J. Cancer* 17:600, 1976.

242. Miller, H., *Medicine and Society*, London: Oxford University Press, 1973.

243. Miller, R. R., "Drug Surveillance Utilizing Epidemiologic Methods. " *Amer. J. Hosp. Pharm.* 30:584, 1973.

244. Miller, R. R., and Greenblatt, D. J., eds., *Drug Effects* in Hospitalized Patients, New York: John Wiley and Sons, 1976.

245. Milunsky, A., "Current Concepts in Genetics: Prenatal Diagnosis of Genetic Disorders, " *N. Eng. J. Med.* 295:377, 1976.

246. Monro, J. K., "The History of Plaster of Paris in the Treatment of Fractures, " *Br. J. Surg.* 23:258, 1932.

247. Moore, M. R.; Bull, J. M.; Jones, S. E.; et al., "Sequential Radiotherapy and Chemotherapy in the Treatment of Hodgkin's Disease, " *Ann. intern. Med.* 77:1, 1972.

248. Moskowitz, M.; Milbrath, J.; Gartside, P.; et al., "Lack of Efficacy of Thermography as a Screening Tool for Minimal and Stage I Breast Cancer, " N. *Eng. J.* Med. 295:249, 1976.

249. Mosteller, F., Harvard University, personal communication, May 17, 1977.

250. National Academy of Sciences, "Medical Technology and the Health Care System: A Study of Equipment-Embodied Technology—Draft" (manuscript), February 1978.

251. National Academy of Sciences—National Research Council, "Postoperative Wound Infection: The Influence of Ultraviolet Irradiation of the Operating Room and of Various Other Factors, " *Ann. Surg.* 160(Suppl.):1, 1964.

252. Neufeld, H.; Weinerman, B. H.; and Kernel, S., "Secondary Malignant Neoplasms in Patients With Hodgkin's Disease, "]. *Am. Med. A.* 239(23):2470,1978.

253. Neuhauser, D., and Jonsson, E., "Managerial Response to New Health Care Technology: Coronary Artery Bypass Surgery, " in *The Management of Health Care*, edited by W.A. Abernathy, et al., Cambridge, Mass.: Ballinger Publishing Co., 1974.

254. Neuhauser, D., and Lewicki, A. M., "What Do We Gain From the Sixth Stool Guaiac?" N. *Eng. J. Med.* 293:226, 1975.

255. Neutra, R., "Indications for the Surgical Treatment of Suspected Acute Appendicitis: A Cost-Effectiveness Approach, " in *Costs, Risks, and Benefits of Surgery*, edited by J. P. Bunker, et al., London: Oxford University Press, 1977.

256. Neutra, R.; Fienberg, S. E.; and Friedman, E., "The Impact of Fetal Monitoring on Neonatal Death, " paper presented before the American Public Health Association, Oct. 31, 1977.

257. The NICHD National Registry for Amniocentesis Study Group, "Mid-Trimester Amniocentesis for Prenatal Diagnosis: Safety and Accuracy, " *J. Am. Med. A.* 236:1471, *1976.*

258. Nichols, R. L.; Broido, P.; Condon, R. E.; et al., "Effect of Preoperative Neomycin-Erythromycin Intestinal Preparation on the Incidence of Infectious Complications Following Colon Surgery, " *Ann. Surg.* 178:453, 1964.

259. Notman, M. R., remarks, a part of "Elective Hysterectomy: Pro and Con, " N. *Eng. J. Med.* 295:266, *1976.*

260. Office of Management and Budget, Executive Office of the President, Special *Analyses, Budget of the United States Government, Fiscal Year 1977,* Washington, D. C.: U.S. Govt. Printing Office, p. 199, p. 213, 1976.

261. Page, I. H., "Egregious Errors in the Management of Hypertension, "]. *Am. Med. A.* 236:2621, *1976.*

262. Pakter, J., and Nelson, F., "Factors in the Unprecedented Decline in Infant Mortality in New York City, " *Bull. N. Y. A cad, Med.* 50:839, *1974.*

263. Paradise, J, L., "Pittsburgh Tonsillectomy and Adenoidectomy Study: Differences From Earlier Studies and Problems and Execution, " *Arm. Oto. Rhino/, Laryngol.* 84:15, *1975.*

264. Paradise, J. L., "Why T and A Remains Moot, " *Pediatrics* 49:648, *1972.*

265. Paradise, J. L., and Bluestone, C. D., "Toward Rational Indications for Tonsil and Adenoid Surgery, " *Hosp. Practice* 11:79, *1976.*

266. Paradise, J. L., et al., "History of Recurrent Sore Throat as an Indication for Tonsillectomy, " N. *Eng. J. Med.* 298:409, *1978.*

267. Paul, R. H., "Intrapartum Fetal Monitoring: Current Status and the Future, " *Obstet. Gynecol. Surv.* 28:453, 1973.

268. Paul, R. H., and Hen, E. H., "Clinical Fetal Monitoring: A Survey of Current Usage, " *Obstet. Gyn.* 37:779, 1971.

269. Paul, R. H.; Huey, J. R.; and Yeager, C. F., "Clinical Fetal Monitoring: Its Effect on Cesarean Section Rate and Perinatal Mortality—Five-Year Trends, ' *Postgr. Med. 61:160, 1977.*

270. Peltier, H. C., "Clinical Trials of Drugs From the Viewpoint of the Pharmaceutical Industry, " *Clin. Pharmacol. Ther.* 18(52):637, 1975.

271. Perrin, J. M.; Charney, E.; MacWhinney, J. B.; et al., "Sulfisoxizole as Chemoprophylaxis for Recurrent Otitis Media, " N, *Eng. J. Med. 291* :664, *1974.*

272. Perrott, G. S., "The Federal Employees Health Benefits Program, " Washington, D. C.: Dept. of Health, Education, and Welfare, 1971.

273. Perry, S., "Hodgkin's Disease: Combined Clinical Staff Conference at the National Institutes of Health, " *Ann. Intern.* Med. 67:424, 1967.

274. Phillips, L. A., *Comparative Evaluation of the Effect of a High Yield Criteria List Upon Skull Radiography Utilization in Emergency Rooms*, Washington, D. C.: Dept. of Health, Education, and Welfare, Bureau of Radiological Health, 1977.

275. Policy Research Incorporated, *The Final Report: A Comprehensive Study of the Ethical, Legal, and* Social *Implications of Advances in Biomedical and Behavioral Research and Technology,* prepared for the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Feb. 25, 1977.

276. Popper, K. R., *The Logic of Scientific Discovery, New* York: Basic Books, Inc., 1959.

277. Prahalad, C. K., and Abernathy, W. J., "A Strategy Approach to the Management of Technology in the Health System, " in *The Management* of *Health Care,* edited by W.J. Abernathy, et al., Cambridge, Mass.: Ballinger Publishing Co., 1974.

278. The President's Commission on Federal Statistics, *Federal Statistics* (Vol. II), Washington, D. C.: U.S. Govt. Printing Office, 1971.

279. Preston, T. A., *Coronary Artery Surgery: A Critical Review, New* York: Raven Press, 1977.

280. Quilligan, E. J., and Freeman, R. K., "The Status of Fetal Monitoring in Decision Making in Patient Management, " in Controversy *in Obstetrics and Gynecology* (Vol. II), edited by D.E. Reid and C.D. Christian, Philadelphia: W.B. Saunders Co., 1974.

281. Quilligan, E. J., and Paul, R. H., "Fetal Monitoring: Is It Worth It?" *Obstet. Gynecol.* 45:96, 1975.

282. Raskin, A.; Gershon, S.; Crook, T. H.; et al., "The Effects of Hyperbaric and Normobaric Oxygen on Cognitive Impairment in the Elderly, " *Arch. G. Psyc.* 35:50, 1978.

283. Renou, P., et al., "Controlled Trial of Fetal Intensive Care, " *Am. J. Obstet. Gynecol.* 126:470, 1976.

284. "Report of the Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents, " *J. Am. Med. A,* 231 :583, 1975.

285. "Report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure: A Cooperative Study, " *J. Am. Med. A.* 237:255, *1977.*

286. Rhoads, G. A.; Kagan, A.; and Yano, K., "Usefulness of Community Surveillance for the Ascertainment of Coronary Heart Disease and Stroke, " *Int. J. Epidemiol.* 4:265, *1975.*

287. Richmond, J. B., testimony before the Subcommittee on Health and Scientific Research, Committee on Human Resources, U.S. Senate, Feb. 7, 1978.

288. Roberts, F., and Shopfner, C. E., "Plain Skull Roentgenograms in Children With Head Trauma, " *Am, J. Roentgenol. Radium Ther. Nuc. Med.* 114:30, *1972.*

289. Rogers, E. M., and Shoemaker, F. F., *Communication* of *Innovations: A Cross-Cultural Approach, New* York: The Free Press, 1971.

290. Roll, G. F., Of *Politics and Drug Regulation,* Rochester, N. Y.: Center for the Study of Drug Development, 1977.

291. Rosenberg, I. L., et al., "Preparation of the Intestine in Patients Undergoing Major Large-Bowel Surgery, Mainly for Neoplasms of the Colon and Rectum, " *Brit. J. Surg.* 58:266, *1971.*

292. Ross, R. S., "Ischemic Heart Disease: An Overview, " *Am. J. Cardiol.* 36:486, *1975.*

293. Ruffin, J. M.; Grizzle, J. E.; Hightower, N. C.; et al., "A Cooperative Double-Blind Evaluation of Gastric 'Freezing' in the Treatment of Duodenal Ulcer, " *N. Eng. J. Med.* 281:16, 1969.

294. Russell, L. B., "Making Rational Decisions About Medical Technology, " presented at the American Medical Association's National Commission of the Cost of Medical Care, Chicago, Ill., Nov. 23, 1976.

295. Sackett, D. L., "Screening for Early Detection of Disease: To What Purpose?" *Bull. N. Y. Acad.* Med. 51:39, 1975.

296. Sandmire, H. F., et al., "Experience With 40,000 Papanicolaou Smears, " *Obstet. Gyn.* 48:56, 1976.

297. Schein, P. S.; Davis, R. D.; Carter, S.; et al., "Commentary: The Evaluation of Anticancer Drugs in Dogs and Monkeys for the Prediction of Qualitative Toxicities in Man, " *Clin. Pharmacol & Ther.* 11(1):3, *1970.*

298. Schmidt, A. M., testimony before the Subcommittee on Health, Committee on Labor and Public Welfare, U.S. Senate, Sept. 25, 1974.

299, Schoenbaum, S. C.; Hyde, J. N., Jr.; Bartoshesky, L.; et al., "Benefit-Cost Analysis of Rubella Vaccination Policy, " N. *Eng. J. Med.* 294:306, 1976.

300. Schor, S., and Karten, L., "Statistical Evaluation of Medical Manuscripts, " J. Am. *Med. A.* 195:1123, *1966.*

301. Schroeder, S. A.; Schliftman, A.; and Piemme, T. E., "Variation Among Physicians in Use of Laboratory Tests: Relation to Quality of Care, " Med. Care 12:709, 1974.

302. Schwartzman, D., "Pharmaceutical R&D Expenditures and Rates of Return, " in *Drug Development and Marketing*, edited by R. Helms, Washington, D. C.: American Enterprise Institute, 1975.

303. Seeff, L. B.; Zimmerman, H. J.; Wright, E. C.; et al., "A Randomized, Double-Blind Controlled Trial of the Efficacy of Immune Serum Globulin for the Prevention of Post-Transfusion Hepatitis, " *Gastroenty*. 72:111, 1977.

304. Seidman, H., *Cancer of the Breast: Statistical and Epidemiological Data, New* York: American Cancer Society, 1972.

305. Seidman, H.; Silverberg, E.; and Holleb, A. I., "A Statistical Comparison of Black and White Populations, " *Ca—A Cancer Journal for Clinicians* 26:2, *1976*.

306. Selden, R.; Neill, W. A.; Ritzman, L. W., et al., "Medical vs. Surgical Therapy for Acute Coronary Insufficiency," N. *Eng. J.* Med. 293:1329, 1975.

307. Sellwood, R. A.; Burn, J. J.; Waterworth, P. M.; et al., "A Second Clinical Trial To Compare Two Methods for Preoperative Preparation of the Large Bowel, " *Brit. J. Surg*. 56:610, *1969*.

308. Seybolt, J. F., and Johnson, W. D., "Cervical Cyto-Diagnostic Problems: A Survey, " *Am. J. Obstet. Gynecol*. 109:1089, 1971.

309. Shafer, D. M., "Vitrectomy," N. *Eng. J. Med.* 295:836, *1976*.

31o. Shapiro, S., "Current Observations From a Test of the Efficacy of Breast Cancer Screening and Their Implications, " *Cancer* (in press).

311. Shapiro, S., "End Result Measurements of Quality of Medical Care, " *Milbank Mere. Fund Q.* 45:7, *1967*.

312. Shapiro, S., quoted in ref. *383*, p. *2*.

313. Shapiro, S.; Strax, P.; and Venet, L., "Periodic Breast Cancer Screening in Reducing Mortality From Breast Cancer," *J. Am. Med. A.* 215:1777, *1971*.

314. Shenker, L.; Post, R. C.; and Seiler, J. S., "Routine Electronic Monitoring of Fetal Heart Rate and Uterine Activity During Labor, " *Obstet. Gyn.* 46:185, 1975.

315. Shepherd, J. A., "Acute Appendicitis: A Historical Survey, " *Lancet* 2:299, 1954.

316. Shryock, R. H., *The Development of Modern Medicine, New* York: Hafner Publishing Co., 1969.

317. Shurin, P. A., "Antibacterial Therapy and Middle Ear Effusion," *Ann. Oto. Rhinol. Laryngol.* 85(2) (Suppl. 25):250, 1976.

318. Simpson, N. E., et al., "Prenatal Diagnosis of Genetic Disease in Canada: Report of a Collaborative Study, " *Can. Med. Assn. J.* 115:739, *1976*.

319. Smits, H. L., Health Care Financing Administration, personal communication, July 17, 1978.

320. Sowton, E.; Hendrix, G.; and Roy, P., "Ten-Year Survey of Treatment With Implanted Cardiac Pacemaker, " *Brit. Med. J.* 3:155, *1974*.

321. Spriggs, A. I., "Follow-Up of Untreated Carcinoma-in-Situ of Cervix Uteri, " *Lancet* 2:599, *1971*.

322. Staquet, M., cd., *The Design of Clinical Trials in Cancer Therapy*, Mt. Kisco, N. Y.: Futura Publishing Co., Inc., *1972*.

323. Stein, Z.; Susser, M.; and Guterman, A., "Screening Programme for Prevention of Down's Syndrome, " *Lancet*, 1:305, 1973.

324. Stern, E., et al, " 'Pap' Testing and Hysterectomy Prevalence: A Survey of Communities With High and Low Cervical Cancer Rates, " *Am. J. Epidem*. 106:296, *1977*.

325. Stern, E., and Neely, P. M., "Carcinoma and Dysplasia of the Cervix: A Comparison of Rates for New and Retiring Population, " *Acta Cytol*. 7:357, *1963*.

326. Stone, H. H., presentation to the Southeastern Surgical Congress, Bal Harbour, Fla., 1977.

327. Stookey, R. A., and Rosen, M. G., "The Status of Fetal Monitoring in Decision Making in Patient Management: Fetal Monitoring During Labor, " in *Controversy in Obstetrics and Gynecology*, Philadelphia: W.B. Saunders Co., 1974.

328. Stool, S. E., and Mast, W. R., "Tonsillectomies and Adenoidectomies: Are They Really Necessary ?," *Bol. Assoc. Med. P.R.* 65:71, *1973*.

329. Strum, S., and Rappaport, H., "The Persistence of Hodgkin's Disease in Long-Term Survivors," *Am. J. Med.* 51:222, *1971*.

330. "A Survey of Opportunities and Need in Research on Digestive Diseases, " *Gastroenty.* 69:1165, *1975.*

331. Swartz, H. A., and Reichling, B. A., "The Risks of Mammograms,"]. *Am. Med. A.* 237:965, *1977.*

332. Swazey, J., "The Efficacy of Medical Procedures: Controls by Physicians, " background paper for the Office of Technology Assessment Study of Efficacy of Medical Procedures, January 1977.

333. Swazey, J. P., and Fox, R. C., "The Clinical Moratorium: A Case Study of Mitral Valve Surgery, " in *ExperimentationWith Human Subjects,* edited by P.A. Freund. New York: George Braziller, 1969.

334. Swazey, J. P.; Klerman, G. L.; and Neville, R., "Regulatory Models for Therapeutic Innovations: Surgery and Drugs, " prepared for the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1976.

335. Tancredi, L. R., "The Ethics Quagmire and Random Clinical Trials, " *Inquiry* XII:171, 1975.

336. Tchilinguirian, N. G., "Fetal Monitoring in High-Risk Pregnancy, " *Clin. Obstet. Gynecol.* 16:329, 1973.

337. Thomas, L., *The Lives of a Cell, New* York: Viking Press, 1974.

338. Thompson, H. W.; McFee, J. G.; Haverkamp, A. D.; et al., "Factors Contributing to Improved Maternal Care and Fetal Outcome in a Medium-Sized City-County Hospital, " *Am. J. Obstet. Gynecol.* 116: 229, 1973.

339. Thompson, L. W.; Davis, G. C.; Obrist, W.; et al., "Effects of Hyperbaric Oxygen on Behavioral and Physiological Measures in Elderly Demented Patients,"}. *Gerontol.* 31:23, 1976.

340. Thorn, J. B.; Russell, E. M.; MacGregor, J. E.; et al., "Costs of Detecting and Treating Cancer of the Uterine Cervix in Northeast Scotland in 1971, " *Lancet* 1:674, 1975.

341. Thrall, R. M., and Cardus, D., "Benefit-Cost and Cost-Effectiveness Analyses in Rehabilitation Research Programs, " *Met. Inf. Med.* 13:147, *1974.*

342. Tukey, J. W., "Some Thoughts on Clinical Trials, Especially Problems of Multiplicity, " *Science* 198:679, *1977.*

343. Turchin, V. F., *The Phenomenon* of *Science, New* York: Columbia University Press, 1977.

344. Tutera, G., and Newman, R. L., "Fetal Monitoring: Its Effect on the Perinatal Mortality and Cesarean Section Rates and Its Complications, " *Am. J. Obstet. Gynecol.* 122:750, 1975.

345. U.S. Congress, Congressional Budget Office, *Catastrophic Health Insurance,* Washington, D. C.: U.S. Govt. Printing Office, 1977.

346. U.S. Congress, Congressional Research Office, *Medical Malpractice: A Survey* of *Associated Problems and Proposed Remedies,* Washington, D. C.: Library of Congress, 1975.

347. U.S. Congress, House. Committee on Interstate and Foreign Commerce, *A Discursive Dictionary of Health Care,* Washington, D. C.: U.S. Govt. Printing Office, 1976.

348. U.S. Congress, House, Committee on Interstate and Foreign Commerce, *Getting Ready for National Health Insurance: Unnecessary Surgery,* hearings before the Subcommittee on Oversight and Investigations, Washington, D. C.: U.S. Govt. Printing Office, 1975.

349. U.S. Congress, House, Committee on Interstate and Foreign Commerce, "Health Services Research, Health Statistics, and Health Care Technology Act, " Washington, D. C.: U.S Govt. Printing Office, 1978.

35o. U.S. Congress, House, Committee on Interstate and Foreign Commerce, *investigation* of *the National Institutes of Health,* Washington, D. C.: U.S. Govt. Printing Office, 1976.

351. U.S. Congress, House, Committee on Interstate and Foreign Commerce, *Medical Device Amendments* of 2975, hearings before the Subcommittee on Health and the Environment, Washington, D. C.: U.S. Govt. Printing Office, 1975.

352. U.S. Congress, House, Committee on Interstate and Foreign Commerce, "Medical Device Amendments of 1976, " report by the Committee, Washington, D. C.: U.S. Govt. Printing Office, 1976.

353. U.S. Congress, Office of Technology Assessment, Cancer *Testing Technology and Saccharin,* Washington, D. C.: U.S. Govt. Printing Office, 1977.

354. U.S. Congress, Office of Technology Assessment, *Development* of *Medical Technology: Opportunities for Assessment,* Washington, D. C.: U.S. Govt. Printing Office, 1976.

355. U.S. Congress, Office of Technology Assessment, *Policy implications of the Computed Tomography (CT) Scanner,* Washington, D. C.: U.S. Govt. Printing Office, 1978.

356. U.S. Congress, Senate, Committee on Human Resources, "The National Institutes of Health Care Research Act of 1978, " report by the Committee, S. Rept. 95-839, Washington, D. C.: U.S. Govt. Printing Office, 1978.

357. U.S. Congress, Senate, Committee on Human Resources, *Oversight of Biomedical and Behavioral Research in the United States, 1977,* hearings before the Subcommittee on Health and Scientific Research, Washington, D. C.: U.S. Govt. Printing Office, *1977.*

358. U.S. Department of Commerce, Bureau of the Census, *Vital Statistics of the United States* 1950: *Part II. Natality and Mortality Data for the United States Tabulated by Place of Residence,* p. 19, Washington, D. C.: U.S. Govt. Printing Office, 1953.

359. U.S. Department of Health, Education, and Welfare, Alcohol, Drug Abuse, and Mental Health Administration, "Charge to the ADAMHA Workgroup on Treatment Assessment Research (TAR)," memorandum from ADAMHA Administrator to Institute Directors, Mar. 3, 1978.

360. U.S. Department of Health, Education, and Welfare, Food and Drug Administration, *FDA Annual Report: 1977,* Washington, D. C.: HEW, 1978.

361. U.S. Department of Health, Education, and Welfare, Food and Drug Administration, Bureau of Radiological Health, *Environmental Assessment Report: Performance Standards for Diagnostic X-Ray Systems and Their Major Components, and Related interpretive Policy Concerning the Assembly and Reassembly of Diagnostic X-Ray Components,* Washington, D. C.: HEW, 1974.

362. U.S. Department of Health, Education, and Welfare, Food and Drug Administration, Bureau of Radiological Health, *Population Exposure to X-Rays: U.S. 1970,* Washington, D. C.: HEW, 1973.

363. U.S. Department of Health, Education, and Welfare, Food and Drug Administration, *Code of Federal Regulations, Tit/e 21,* Washington, D. C.: U.S. Govt. Printing Office, 1977.

364. U.S. Department of Health, Education, and Welfare, Food and Drug Administration, "Drug Application Evaluation: The IND/NDA Process" (draft manual), Washington, D. C.: HEW, Dec. *20, 1976.*

365. U.S. Department of Health, Education, and Welfare, Food and Drug Administration, *Medical Device Classification Procedures,* Part *II,* Washington, D. C.: HEW, 1974.

366. U.S. Department of Health, Education, and Welfare, Health Resources Administration, "The Computer-Assisted EKG, from Laboratory to Community, " HEW Pub. No. (HRA) 74-3104, Washington, D. C.: HEW, Sept. 1973.

367. U.S. Department of Health, Education, and Welfare, Health Resources Administration, "Computer-Based Patient Monitoring Systems, " NCHSR Research Report Series, HEW Pub. No. (HRA) 76-3143, Washington, D. C.: HEW, n.d.

368. U.S. Department of Health, Education, and Welfare, Health Resources Administration, "NCHSR Research Bibliography (July 1, 1976 to June 30, 1977), " NCHSR Research Management Series, HEW Pub. No. (HRA) 78-3202, Washington, D. C.: HEW, n.d.

369. U.S. Department of Health, Education, and Welfare, "Health Technology Management at the Department of Health, Education, and Welfare, " final Phase I Report for the Secretary, prepared by the Office of the Assistant Secretary for Planning and Evaluation and the Office of the Assistant Secretary for Health, 1977.

370. U.S. Department of Health, Education, and Welfare, National Center for Health Services Research, *Computer Applications in Health Care,* draft report, Apr. 26, 1978.

371. U.S. Department of Health, Education, and Welfare, National Center for Health Services Research, "Optimal Electrocardiography, " NCHSR Research Summary Series, HEW Pub. No. (PHS) 78-3193, Washington, D, C.: HEW, n.d.

372. U.S. Department of Health, Education, and Welfare, National Center for Health Statistics, "Characteristics of Females Ever Having a Pap Smear and Interval Since Last Pap Smear, United States, 1973, " *Monthly Vital Statistics Report, Health inter-uiew Survey, Provisional Data* (HRA) 76-1120, Vol. 24, No. 7, Suppl., October 1975.

373. U.S. Department of Health, Education, and Welfare, National Center for Health Statistics, "The National Ambulatory Medical Care Survey: 1973 Summary, " in *Vital and Health Statistics,* Series 13, No. 21, Washington, D. C.: U.S. Govt. Printing Office, 1975.

374. U.S. Department of Health, Education, and Welfare, National Center for Health Statistics, "Surgical Operations in Short-Stay Hospitals: United States, 1973, " in *Vital and Health Statistics,* Series 13, No. 21, Washington, D. C.: U.S. Govt. Printing Office, 1976.

375. U.S. Department of Health, Education, and Welfare, National Center for Health Statistics, "Utilization of Short-Stay Hospitals by Diagnosis: United States, 1965, " in *Vital and Health Statistics*, Series 13, No. 6, Washington, D. C.: U.S. Govt. Printing Office, 1970.

376. U.S. Department of Health, Education, and Welfare, National Center for Health Statistics, "Utilization of Short-Stay Hospitals: United States, 1974, " in *Vital and Health Statistics*, Series 13, No. 26, Washington, D. C.: U.S. Govt. Printing Office, 1976.

377. U.S. Department of Health, Education, and Welfare, National Institute for Occupational Safety and Health, "NIOSH Fact Sheet, " November 1977.

378. U.S. Department of Health, Education, and Welfare, National Institutes of Health, "Issue Paper—NIH Support of Clinical Trials, " May 16, 1975.

379. U.S. Department of Health, Education, and Welfare, National Institutes of Health, *The National Institutes of Health Forward Plan: Fiscal Years 1977-1981*, Apr. 30, 1975.

380. U.S. Department of Health, Education, and Welfare, National Institutes of Health, *NIH Inventory of Clinical Trials: Fiscal Year 1975* (Vol. I and II), Washington, D. C.: U.S. Govt. Printing Office, 1977.

381. U.S. Department of Health, Education, and Welfare, National Institutes of Health, *The Responsibilities of NIH at the Health Research/Health Care Interface*, draft report of the Office of the Director, Feb. 14, 1977.

382. U.S. Department of Health, Education, and Welfare, National Institutes of Health, "Thirty-Third Meeting, Advisory Committee to the Director, NIH, " briefing book, Dec. 2-3, 1976.

383. U.S. Department of Health, Education, and Welfare, National Institutes of Health, National Cancer Institute, "Final Reports of National Cancer Institute Ad Hoc Working Groups on Mammography Screening for Breast Cancer and a Summary Report of Their Joint Findings and Recommendations, " HEW Pub. No. (NIH) 77-1400, Washington, D. C.: HEW, 1977.

384. U.S. Department of Health, Education, and Welfare, National Institutes of Health, National Cancer Institute, "NIH/NCI Consensus Development Meeting on Breast Cancer Screening—Background Statement," Bethesda, Md., Sept. 14-16, 1977.

385. U.S. Department of Health, Education, and Welfare, National Institutes of Health, National Cancer Institute, "NIH/NCI Consensus Development Meeting on Breast Cancer Screening—Issues and Recommendations: Sept. 14-16, 1977, " manuscript, Bethesda, Md., Oct. 18, 1977.

386. U.S. Department of Health, Education, and Welfare, National Institutes of Health, National Cancer Institute, "Statement on X-Ray Mammography in Screening for Breast Cancer, " press release, Bethesda, Md.: NIH, January 1977.

387. U.S. Department of Health, Education, and Welfare, National Institutes of Health, National Heart and Lung Institute, "Clinical Trials Briefing Document, " manuscript, Nov. 10, 1975.

388. U.S. Department of Health, Education, and Welfare, National Institutes of Health, National Institute of Allergy and Infectious Disease, memorandum to the Office of Technology Assessment, Jan. 19, 1977.

389. U.S. Department of Health, Education, and Welfare, Public Health Service, *Vital Statistics of the U.S. 1960: Mortality* (Part A, Vol. II), Washington, D. C.: U.S. Govt. Printing Office, 1963.

390. U.S. Department of Health, Education, and Welfare, Public Health Service, *Vital Statistics of the U.S. 1965: Mortality* (Part A, Vol. II), Washington, D. C.: U.S. Govt. Printing Office, 1967.

391. U.S. Department of Health, Education, and Welfare, Public Health Service, *Vital* Statistics of the *U.S.* 1973: *Mortality* (Part A, Vol. II), Washington, D. C.: U.S. Govt. Printing Office, 1975.

392. U.S. Department of Health, Education, and Welfare, *Report of the Presidents Biomedical Research Panel*, HEW Pub. No. OS 76-500, Washington, D. C.: Apr. *30, 1976.*

393. U.S. Energy Research and Development Administration, ERDA Task Force on Nuclear Medicine, "Nuclear Medicine Research: An Evaluation of the ERDA Program, " Washington, D. C.: ERDA, August 1976.

394. Utterback, J. M., "Innovation in Industry and the Diffusion of Technology, " Science 183(4125):620, *1974.*

395. "Vein Graft Patency and Intimal Proliferation After Aorto-Coronary Bypass.: Early and Long Term Angiopathologic Correlations, " *Am.J.Cardiol.* 38:856, 1976.

396. Veterans Administration, *Medical Research in the Veterans Administration* (printed for the use of the Committee on Veterans Affairs, U.S. House of Representatives), Washington, D. C.: U.S. Govt. Printing Office, 1976.

397. Veterans Administration, Ad Hoc Interdisciplinary Advisory Committee on Antimicrobial Drug Usage, "Prophylaxis in Surgery,"]. *Am. Med. A. 237: 1003, 1977.*

398. "The Veterans Administration Cooperative Randomized Study of Surgery for Coronary Arterial Occlusive Disease, " *Circulation 54(* Suppl. *3)*, December *1976.*

399. "Veterans Administration Cooperative Study Group on Antihypertensive Agents, " *J. Am. Med. A.* 202:1028, 1967; and 213:1143, 1970.

400. Veterans Administration, Cooperative Urological Research Group, "Treatment and Survival of Patients With Cancer of the Prostate, " *Surg. Gynecol. Obstet.* p. 1011, May 1967.

401. Vinciguerra, V., et al., "A Combination Chemotherapy for Resistant Hodgkin's Disease,"}. *Am. Med. A.* 237:33, *1977.*

402. Visscher, M.B., *Ethical Constraints and Imperatives in Medical Research,* Springfield, Ill.: Charles C. Thomas, 1975.

403. Wakely, C., and Childs, P., "Appendicitis," *Br. Med. J.* 2:1347, *1950.*

404. Waldman, S., "The Effect of Changing Technology on Hospital Costs, " *Research and Statistics Note No. 4,* Washington, D. C.: Social Security Administration, Office of Research and Statistics, Feb. *28, 1972.*

405. Ward, G. W., National Institutes of Health, memorandum from the National High Blood Pressure Education Program to the Office of Technology Assessment, Jan. 18, 1977.

406. Warden, W., and Lasagna, L., *Regulation and Drug Development,* Washington, D. C.: American Enterprise Institute, 1975.

407. Warner, K.E. "A 'Desperation-Reaction' Model of Medical Diffusion, " *Health Serv. Res.* 10:369, 1975.

408. Warner, K. E., "Treatment Decision Making in Catastrophic Illness, " *Med.* Care 15(1):19, 1977.

409. Washington, J. A.; Dearing, W. H.; Judd, E. S., et al., "Effect of Preoperative Antibiotic Regimen on Development of Infection After Intestinal Surgery: Prospective, Randomized, Double-Blind Study, " *Ann. Surg.* 189:567, 1974.

410. Waters, W.E. "Controlled Clinical Trial of Ergotamine Tartrate, " *Brit. Med. J.* p. *325,* May 1970.

411. Watkins, R. N.; Duncan, E. A.; and Andenes, J. L., "Appendectomy Experience in Prepaid Group Practice, " *N. Eng. J. Med.* 293:995, *1975.*

412. Weinstein, A. S., et al., *Final Report to the National Science Foundation on Product Liability: A Study of the Interaction of Law and Technology,* NSF Pub. No. NSF/RA-770025, Washington, D. C.: September 1977.

413. Weinstein, M., "Allocation of Subjects in Medical Experiments, " N. *Eng. J.* Med. 291:1278, 1974.

414. Weinstein, M. A.; Alfidi, R. J.; and Guchesneau, P. M., "Computed Tomography, Six-Skull Roentgenography, " *Am. J. Roentgenol.* (in press).

415. Weinstein, M. C.; Pliskin, J. S.; and Stason, W. B., "Coronary Artery Bypass Surgery: Decision and Policy Analysis, " in *Costs, Risks, and Benefits of Surgery,* edited by J.P. Bunker, et al., London: Oxford University Press, 1977.

416, Weiss, W., and Boucot, K. R., "The Philadelphia Pulmonary Neoplasm Research Project: Early Roentgenographic Appearance of Bronchogenic Carcinoma, " *Arch. Intern. Med.* 134:306, *1974.*

417. Welch, C. E., "Abdominal Surgery, " N. Eng. *J. Med.* 293:903, *1975.*

418. Wells, C. K., and Feinstein, A. R., "Routine Radiographic Measurement and Prognostic Importance of Rate of Growth (Auxometry) in Patients With Lung Cancer" (abstract), *Clin. Res. 25:266A, 1977.*

419. Wennberg, J. E., "Changing Patterns of Risk, Medical Care and Perinatal Mortality in Vermont, " paper presented before the American Public Health Association, Oct. 15-19, 1978.

420. Wennberg, J., opening remarks at the 25th Annual National Health Forum, New York, N. Y., Mar. 23, 1977.

421. Wennberg, J., and Gittelsohn, A., "Health Care Delivery in Maine: I. Patterns of Use of Common Surgical Procedures, "]. *Maine Med. Assoc.* 66:123, 1975.

422. Wennberg, J., and Gittelsohn, A., "Small Area Variations in Health Care Delivery, " *Science* 182:1102, 1973.

423. Whalan, D. J., "The Ethics and Morality of Clinical Trials in Man, " *Med. J. Aust.* 1:491, 1975.

424. White, J. J.; Santillana, M.; and Hailer, J. A., "Intensive In-Hospital Observation: A Safe Way To Decrease Unnecessary Appendectomy, " *Am. Surg.* 41:793, 1975.

425. White, K. L., "Contemporary Epidemiology, " *Int. J. Epidemiol.* 3:295, *1974.*

426. White, K. L., "International Comparisons of Health Services Systems, " *Milbank Mere.* Fund Q. 46:117, 1968.

427. Wickstrom, J., *Clinical Orthopedics* and *Related Research,* Philadelphia: J.B. Lippincott, 1966.

428. Williamson, J., "Correctable Deficiencies in Contemporary Quality Assurance" (manuscript), Jan. 20, 1977.

429. Williamson, J. W.; Alexander, M.; and Miller, G., "Continuing Education and Patient Care Research: Physician Response to Screening Test Results,"]. *Am. Med. A.* 201:118, *1967.*

430. Willis, A. M., "The Mortality in Important Surgical Disease, Especially Appendicitis, " *Surg. Gynecol. Obstet.* 42:318, *1926.*

431. Wing, J. L., "The Ethics of Clinical Trials,"]. *Med. Ethics* I:174, 1975.

432. Wolf, J., et al., "Controlled Study of Survival of Patients With Clinically Inoperable Lung Cancer Treated With Radiation Therapy, " *Am. J. Med.* 40:360, 1966.

433. Wolfe, S. M., and Warner, R., "Mammography: A Case for Informed Consent" (mimeograph), Washington, D. C.: Health Research Group, 1976.

434. "Workshop on Tonsillectomy and Adenoidectomy" (supported by the National Institute of Neurological Diseases and Stroke, National Institutes of Health), *Ann. Oto. Rhinol. Laryngol. 84(* Suppl *19): 1975.*

435. World Health Organization, *Statistical Indicators for the Planning and Evaluation of Public Health Programs: Fourteenth Report of the Expert Committee on Health Statistics,* WHO Technical Report Series, No. 472. Geneva, 1971.

436. World Health Organization, "Vital Statistics and Causes of Death, " in *World Health Statistics Annual 1973-1976 (Vol. 1),* Geneva, 197.

437. Worthington, N. L., "Expenditures for Hospital Care and Physicians' Services: Factors Affecting Annual Changes, " *Soc.* Sec. *Bull.* 38:3, 1975.

U.S. GOVERNMENT PRINTING OFFICE : 1978 O - 33-406