TA101

Introduction to Health Care Technology Assessment

Clifford S. Goodman
The Lewin Group
January 1998

Table of Contents

ACKNOWLEDGMENTS

TA 101: Introduction to Health Care Technology Assessment, is derived from an evolving set of seminars and other presentations that I have given on health care technology assessment since the mid-1980s. *TA 101* was first assembled as a single document for a conference, *Technology Assessment: A Tool for Technology Management and Improved Patient Outcomes*, held in January 1995 in Washington, DC. The conference was sponsored by the VA Health Services Research and Development Service and its Management Decision and Research Center, and the Association for Health Services Research (AHSR) [http://www.ahsr.org]. As she did for so many AHSR efforts, the late Alice Hersh, who was Executive Vice President and Chief Executive Officer of AHSR, provided the impetus and considerable motivation for that conference.

The substance of *TA 101* draws from the work of the many people whose publications are cited in the references and reading lists. In particular, I would like to acknowledge the influence of David Banta, Robert Brook, Thomas Chalmers, David Eddy, Bryan Luce, and Frederick Mosteller on my understanding of the field and appreciation for the importance of involving others in it. I wish to thank Earl Steinberg, Kevin Marvelle, and Allan Meyers for their thoughtful and expert reviews of earlier versions of *TA 101*. I also thank the many conference participants who gave useful suggestions for improving the conference version of *TA 101*.

The updating, expansion, and preparation of *TA 101* for distribution and viewing via the World Wide Web was funded by the National Library of Medicine. I wish to acknowledge the support and guidance of Catherine Selden, the project officer in NLM's National Information Center on Health Services Research and Health Care Technology.

Clifford S. Goodman
The Lewin Group
January 1998

ABOUT THE AUTHOR

Clifford Goodman is a senior manager in the medical technology practice of The Lewin Group, a health care policy and management consulting firm located near Washington, DC. His work primarily concerns evaluation and related policy analyses of health care technology, including drugs, medical devices, procedures, and systems in health care.

Dr. Goodman's methodological experience involves technology assessment, outcomes research, cost effectiveness analysis, decision analytic modeling, and studies pertaining to technological innovation, third party payment, and diffusion. He has worked in such diverse technological areas as heart assist devices, percutaneous transluminal coronary angioplasty, hypertension management, diagnostic testing, substance abuse treatment, computed tomography, positron emission tomography, and

medical informatics. His work on databases in health care technology assessment and health services research contributed to the development of the HealthSTAR and HSRProj databases of the National Library of Medicine.

Dr. Goodman was a visiting researcher at the Swedish Council on Technology Assessment in Health Care and a visiting professor at Uppsala University (1990-91). As a National Research Council Fellow and later as director of the Council on Health Care Technology, he managed a series of technology assessment projects at the Institute of Medicine of the U.S. National Academy of Sciences (1982-90). With the Medicus Systems Corp., he provided management consultation to federally funded health centers in medically underserved areas in the U.S. (1978-80).

Dr. Goodman is a board member of the International Society of Technology Assessment in Health Care (ISTAHC) [http://www.istahc.org] and of the Health Care Technology Assessment Division of the International Federation for Medical and Biological Engineering (IFMBE), and a Fellow of the American Institute for Medical and Biological Engineering (AIMBE). He did his undergraduate work at Cornell University, received a master's degree from the Georgia Institute of Technology, and earned his doctorate from the Wharton School of the University of Pennsylvania.

INTRODUCTION

Technological innovation has yielded truly remarkable advances in health care during the last three decades. In just the last several years, breakthroughs in biotechnology, biomaterials, surgical techniques, and computer technology have helped to improve health care delivery and patient outcomes.
The proliferation of health care technology has accompanied burgeoning health care costs, and the former has been cited as "culprit" for the latter, although the nature and strength of this relationship are not clear (Health Care Technology Institute 1994; Newhouse 1992). Certainly, few patients or clinicians are willing to forego access to state-of-the-art health care technology. Particularly in the U.S., adoption and use of technology has been stimulated by patient and physician incentives to seek any potential health benefit regardless of cost, and by third-party payment, malpractice concerns, provider competition and effective marketing of technologies.

In this era of increasing cost pressures, market-based restructuring of health care, proposals for further health reforms, and continued inadequate access to care for tens of millions in the U.S., technology remains the substance of health care. Culprit or not, technology can be managed. Indeed, the flow of technology is increasingly modulated by a widening group of policymakers in the health care sector. Health product makers, regulators, clinicians, patients, hospital managers, payers, government leaders and others increasingly demand well-founded information to support decisions about whether or how to develop technology, to allow it on the market, to acquire it, to use it, to pay for its use and more. The growth and development of **health care technology assessment** (HCTA) in government and the private sector reflect this demand.

Health care technology assessment methods are evolving and its applications are increasingly diverse. This text introduces certain fundamental aspects and issues of a dynamic field of inquiry. Broader

participation of people with multiple disciplines and different roles in health care is enriching the field. The heightened demand for health care technology assessment, in particular from the private sector and from those public institutions that are shifting to more competitive arenas, is pushing the field to evolve keener processes and market-specific products. Like the information required to conduct most assessments, the body of knowledge about health care technology assessment cannot be found in one place and is not static. Practitioners and users of health care technology assessment should not only monitor changes in the field, but they should contribute to its development.

Origins of Technology Assessment

Technology assessment (TA) arose in the mid-1960s from an appreciation of the critical role of technology in modern society and its potential for unintended, and sometimes harmful, consequences. Experience with the side effects of a multitude of chemical, industrial and agricultural processes, and such services as transportation, health and resource management contributed to this understanding. Early assessments concerned such topics as offshore oil drilling, pesticides, automobile pollution, nuclear power plants, supersonic airplanes and the artificial heart. TA was conceived as a way to identify the desirable first-order, intended effects of technologies as well as the higher-order, unintended social, economic and environmental effects (Brooks and Bowers 1970).

The term technology assessment was introduced in 1965 during deliberations of the Committee on Science and Astronautics of the U.S. House of Representatives. Congressman Emilio Daddario emphasized that the purpose of TA was to serve policymaking:

[T]echnical information needed by policymakers is frequently not available, or not in the right form. A policymaker cannot judge the merits or consequences of a technological program within a strictly technical context. He has to consider social, economic, and legal implications of any course of action (U.S. Congress, House of Representatives 1967).

Congress commissioned independent studies by the National Academy of Sciences, the National Academy of Engineering (NAE), and the Legislative Reference Service of the Library of Congress that significantly influenced the development and application of TA. These studies and further congressional hearings led the National Science Foundation to establish a TA program and, in 1972, Congress to authorize the congressional Office of Technology Assessment (OTA), which was founded in 1973, became operational in 1974 and established its health program in 1975.

Many observers were concerned that TA would be a means by which government would impede the development and use of technology. However, this was not the intent of Congress or of the agencies that conducted the original TAs. In 1969, an NAE report to Congress emphasized that:

Technology assessment would aid the Congress to become more effective in assuring that broad public as well as private interests are fully considered while enabling technology to make the maximum contribution to our society's welfare (National Academy of Engineering 1969).

With somewhat different aims, private industry used TA to aid in competing in the marketplace, for improving understanding of the future business environment, and for producing options for decision makers.

TA methodology drew upon a variety of analytical, evaluative and planning techniques. Among these were systems analysis, cost-benefit analysis, consensus methods (e.g., Delphi method), engineering

feasibility studies, clinical trials, market research, technological forecasting, and others. TA practitioners and policymakers recognized that TA is evolving, flexible, and should be tailored to the task (U.S. Congress, Office of Technology Assessment 1977).

Early Health Care Technology Assessment

Health care technologies had been studied for safety, effectiveness, cost, and other concerns long before the advent of HCTA. Development of TA as a systematic inquiry in the 1960s and 1970s coincided with the introduction of health care technologies that prompted widespread public interest in matters that transcended their immediate health effects. Health care technologies were among the topics of early TAs. Multiphasic health screening was one of three topics of "experimental" TAs conducted by the NAE at the request of Congress (National Academy of Engineering 1969). In response to a request by the National Science Foundation to further develop the TA concept in the area of biomedical technologies, the National Research Council conducted TAs on *in vitro* fertilization, predetermination of the sex of children, retardation of aging and modifying human behavior by neurosurgical, electrical or pharmaceutical means (National Research Council 1975). The OTA issued a report on drug bioequivalence in 1974, and the OTA Health Program issued its first formal report in 1976.

Since its early years, HCTA has been fueled in part by emergence and diffusion of technologies that have evoked social, ethical, legal, and political concerns. Among these technologies are contraceptives, organ transplantation, artificial organs, life-sustaining technologies for critically or terminally ill patients, and, more recently, genetic testing and genetic therapy. These technologies have challenged certain societal institutions, codes, and other norms regarding fundamental aspects of human life such as parenthood, heredity, birth, bodily sovereignty, freedom and control of human behavior, and death (National Research Council 1975).

Despite the comprehensive approach originally intended for TA, its practitioners recognized early on that "partial TAs" may be preferable in circumstances where selected impacts are of particular interest or where necessitated by resource constraints (U.S. Congress, Office of Technology Assessment 1977). In practice, relatively few TAs encompass the full range of possible technological impacts; most focus on certain sets of impacts or concerns, depending upon who conducts the assessment and what data and other resources are available for the assessment. Although there have been important instances of comprehensive HCTAs, partial ones have been typical (Goodman 1992).

Box 1 shows various definitions of TA and HCTA.

Box 1

Some Definitions of TA and HCTA

We shall use the term assessment of a medical technology to denote any process of examining and reporting properties of a medical technology used in health care, such as safety, efficacy, feasibility, and indications for use, cost, and cost-effectiveness, as well as social, economic, and ethical consequences, whether intended or unintended (Institute of Medicine 1985).

Technology assessment (TA) is a category of policy studies, intended to provide decision makers with information about the possible impacts and consequences of a new technology or a significant change in an old technology. It is concerned with both direct and indirect or secondary consequences, both benefits and disbenefits, and with mapping the uncertainties involved in any

government or private use or transfer of a technology. TA provides decision makers with an ordered set of analyzed policy options, and an understanding of their implications for the economy, the environment, and the social, political, and legal processes and institutions of society (Coates & Jarratt 1992).

Technology assessment is a form of policy research that examines short- and long-term social consequences (for example, societal, economic, ethical, legal) of the application of technology. The goal of technology assessment is to provide policy-makers with information on policy alternatives (Banta and Luce 1993).

The term health technology assessment (HTA) is used to describe the assessment of the costs, effectiveness and broader impact of all methods used by health professionals to promote health, prevent and treat disease and improve rehabilitation and long term care (Department of Health, National Health Service, U.K. 1994).

Health technology assessment ... is a structured analysis of a health technology, a set of related technologies, or a technology-related issue that is performed for the purpose of providing input to a policy decision (U.S. Congress, Office of Technology Assessment 1994).

Health technology assessment is the evaluation of medical technologies -- including procedures, equipment and drugs. An assessment requires an interdisciplinary approach which encompasses analyses of safety, costs, effectiveness, efficacy, ethics, and quality of life measures (Canadian Coordinating Office for Health Technology Assessment, 1995).

[HTA] is a multidisciplinary field of policy analysis. It studies the medical, social, ethical, and economic implications of development, diffusion, and use of health technology.

FUNDAMENTAL CONCEPTS AND ISSUES

Health Care Technology Assessment

Health care technology assessment is the systematic evaluation of properties, effects and/or other impacts of health care technology. The main purpose of HCTA is to inform technology-related policymaking in health care, where policymaking is used in the broad sense to include decisions made at, e.g., institutional, regional, national and international levels. HCTA may address the direct and intended consequences of technologies as well as their indirect and unintended consequences. HCTA is conducted by interdisciplinary groups using explicit analytical frameworks drawing from a variety of methods.

Purposes of HCTA

HCTA can be used in many ways to support policymaking. Among these ways are to:

- advise a regulatory agency such as the Food and Drug Administration (FDA) about whether or not to permit the commercial use (e.g., marketing) of a drug, device or other technology
- help health care payers and providers to determine which technologies should be included in health benefits plans, addressing coverage (whether or not to pay) and/or reimbursement (how much to pay) policies
- advise clinicians, providers and patients about the proper use of health care interventions for particular health problems (e.g., practice guidelines, disease management programs)
- help managers of hospitals, health care networks and other health care organizations to make decisions regarding technology acquisition and management
- advise government health department officials about undertaking public health programs (e.g., vaccination, screening and environmental protection programs)
- support health care product company decisions about product development and marketing
- set voluntary or mandatory standards regarding the manufacture, use, maintenance, reuse and other aspects of health care technologies, components and materials
- advise investors and companies concerning venture capital funding, acquisitions and divestitures, and other transactions concerning health care product and service companies
- advise state and national leaders about policies concerning technological innovation, research and development, regulation, payment and delivery of health care.

HCTA contributes in many ways to the knowledge base for improving the quality of health care, especially to support development and updating of a wide spectrum of standards, guidelines and other health care policies. For example, such standards-setting organizations as the American National Standards Institute and the American Society for Testing and Materials coordinate development of voluntary national consensus standards for the manufacture, use and reuse of health devices and their materials and components. The Joint Commission on Accreditation of Healthcare Organizations and the National Committee for Quality Assurance set standards for measuring quality of care and services of hospitals, managed care organizations, long-term care facilities, hospices, ambulatory care centers, and other health care institutions. Medical professions organizations (e.g., American College of Cardiology, American College of Physicians, American College of Radiology, American Medical Association), government agencies (e.g., Agency for Health Care Policy and Research) and special panels (e.g., U.S. Preventive Services Task Force) develop **clinical practice guidelines,** standards and other statements regarding the appropriate use of technologies. Increasingly, organizations such as these are explicitly linking HCTA findings to standards setting.

As noted above, HCTA can be used to inform decision making by clinicians and patients. In recent years, the term **evidence-based medicine** has arisen to describe this application of HCTA. This has prompted the publication of various resources, including: a guide to evidence-based medicine (Sackett et al. 1997); a new journal, *Evidence-Based Medicine* (a joint product of the American College of Physicians and the BMJ Publishing Group), which is a digest of articles selected from a selection of articles from international medical journals using scientific criteria; and a Centre for Evidence-Based Medicine [http://cebm.jr2.ox.ac.uk] in the NHS Research and Development (UK).

Basic HCTA Orientations

The impetus of a HCTA is not necessarily a technology. Three basic orientations to HCTA are as follows:

- *Technology-oriented* assessments are intended to determine the characteristics or impacts of particular technologies. For example, a government agency may want to determine the clinical, economic, professional, scientific and/or industrial impacts of magnetic resonance imaging (MRI).
- *Problem-oriented* assessments focus on solutions or strategies for managing a particular problem for which alternative or complementary technologies might be used. For example, clinicians and providers concerned with the problem of diagnosis of dementia may call for the development of clinical practice guidelines involving some combination or sequence of neurological examination, computed tomography (CT), MRI and/or positron-emission tomography (PET).
- *Project-oriented* assessments focus on a local placement or use of a technology in a particular institution, program or other designated project. For example, this may arise when a particular hospital must decide whether or not to purchase an MRI unit, considering the facilities, personnel, and other resources needed to install and operate an MRI unit; the hospital's financial status; local market potential for MRI services; competitive factors; etc.

These basic assessment orientations can overlap and complement one another. Certainly, all three types could draw upon the same body of scientific evidence and other information. A technology-oriented assessment may address the range of problems for which the technology might be used and how appropriate the technology might be for different types of local settings (e.g., inpatient versus outpatient). A problem-oriented assessment, examining the effects or other impacts of alternative technologies on a given problem, may incorporate multiple, focused (i.e., on the problem at hand) technology-oriented assessments. A project-oriented assessment would consider the range of impacts of a technology or its alternatives in a given setting, as well as the role or usefulness of that technology for various problems. Although the information used in a project-oriented assessment by a particular hospital may include findings of pertinent technology- and problem-oriented assessments, local data collection and analysis may be required to determine what is sensible for that hospital. Thus, most good HCTAs will blend some aspects of all three basic orientations.

Scope of Health Care Technology

Three ways to describe health care technology include its material nature, its purpose, and its stage of diffusion.

Material Nature

In the broad sense, **technology** is the practical application of knowledge. Accordingly, health care technology means more than health care hardware. Most health care technologies fall into the following broad categories:

- *drugs:* e.g., aspirin, beta-blockers, penicillin, vaccines, blood products
- *devices, equipment and supplies:* e.g., cardiac pacemakers, CT scanners, surgical gloves, diagnostic test kits
- *medical and surgical procedures:* e.g., psychotherapy, coronary angiography, gall bladder removal
- *support systems:* e.g., electronic patient record systems, telemedicine systems, blood banks, clinical laboratories
- *organizational and managerial systems:* e.g., prospective payment using diagnosis-related groups, alternative health care delivery configurations (types of managed care systems, settings of care, etc.), clinical pathways, total quality management programs

Purpose

Technologies can also be grouped according to their health care purpose, i.e.:

- *prevention*: intended to protect against disease by preventing it from occurring, reducing the risk of its occurrence, or limiting its extent or sequelae (e.g., immunization, hospital infection control program, fluoridated water supply)
- *screening*: intended to detect a disease, abnormality, or associated risk factors in asymptomatic people (e.g., Pap smear, tuberculin test, mammography, serum cholesterol testing)
- *diagnosis*: intended to identify the cause and nature or extent of disease in a person with clinical signs or symptoms (e.g., electrocardiogram, serological test for typhoid, x-ray for possible broken bone)
- *treatment*: designed to improve or maintain health status, avoid further deterioration, or provide palliation (e.g., coronary artery bypass graft surgery, psychotherapy, drugs for cancer pain)
- *rehabilitation*: intended to restore, maintain or improve a physically or mentally disabled person's function and well-being (e.g., exercise program for post-stroke patients, assistive device for severe speech impairment, incontinence aid)

Not all technologies fall neatly into single categories. Many tests and other technologies used for diagnosis also are used for screening. (The probability that a patient has a disease or other health condition is greatly affected by whether these technologies are used for screening asymptomatic patients or diagnosing symptomatic patients.) Some technologies are used for diagnosis as well as treatment, e.g., coronary angiography to diagnose heart disease and to monitor coronary angioplasty. Implantable cardioverter defibrillators detect potentially life-threatening heart arrhythmias and deliver electrical pulses to restore normal heart rhythm. Electronic patient record systems can support all technological purposes.

Certain "boundary-crossing" technologies that combine characteristics of drugs, devices or other major categories of technology (e.g., spermicidal condoms, implantable drug infusion pumps and bioartificial organs that combine natural tissues and artificial components) may pose jurisdictional or administrative challenges to regulatory agencies, payers, standards-setting bodies and other organizations (Goodman 1993). Examples of technologies whose boundary-crossing attributes have

complicated regulatory approval and coverage decisions in recent years are gallstone lithotripters (used with stone-dissolving drugs) (Zeman et al. 1994) and PET (used with radiopharmaceuticals) (Coleman 1992).

Stage of Diffusion

Technologies may be assessed at different stages of diffusion and maturity. In general, health care technologies may be described as being:

- *future:* in a conceptual stage, anticipated, or in the earliest stages of development;
- *experimental:* undergoing bench or laboratory testing using animals or other models;
- *investigational:* undergoing initial clinical evaluation (i.e., in humans);
- *established:* considered by providers to be a standard approach to a particular condition and diffused into general use; or
- *obsolete/outmoded/abandoned:* superseded by other technologies or demonstrated to be ineffective or harmful.

Often, these stages are not clearly delineated, and technologies do not necessarily mature through them in a linear fashion. A technology may be established for certain patient applications and investigational for others. A technology that was once considered obsolete may return to established use for a better defined or entirely different clinical purpose. Many technologies undergo multiple incremental innovations after their initial acceptance into general practice (Gelijns and Rosenberg 1994; Reiser 1994).

Timing of Assessment

There is no single correct time to conduct a HCTA. HCTA is conducted to meet the needs of a variety of policymakers seeking assessment information throughout the lifecycles of technologies. Investors, regulators, payers, hospital managers and others tend to make decisions about technologies at particular junctures, and each may subsequently reassess technologies. Indeed, the determination of a technology's stage of diffusion may be the primary purpose of an assessment. For insurers and other payers, technologies that are deemed experimental or investigational are usually excluded from coverage, whereas those that are established or generally accepted are usually eligible for coverage (Newcomer 1990; Reiser 1994). Recently, some payers have begun to provide conditional coverage for selected investigational technologies in order to compile evidence on safety, efficacy/effectiveness, etc., for making more informed coverage policies.

There are tradeoffs inherent in decisions regarding the timing for HCTA. On one hand, the earlier a technology is assessed, the more likely its diffusion can be curtailed if it is unsafe or ineffective (McKinlay 1981). From centuries' old purging and bloodletting to the more contemporary gastric bubble, the list of poorly evaluated technologies that diffused into general practice before being found to be ineffective and/or harmful continues to grow. (See Box 2).

On the other hand, to regard the findings of an early assessment as definitive or final may be misleading. An investigational technology may not yet be perfected; its users may not yet be proficient; its costs may not yet have stabilized; it may not have been applied in enough circumstances to recognize its potential benefits; and its long-term outcomes may not yet be known. Further, the "**moving target problem**" can complicate HCTA (Goodman 1996). By the time a HCTA is

conducted, reviewed and disseminated, its findings may be outdated by changes in a technology, in how it is used, or in its technological alternatives for a given problem. As one technology assessor concluded about the problems of when-to-assess: "It's always too early until, unfortunately, it's suddenly too late!" (Buxton 1987)

More and more, HCTA is considered to be an iterative process rather than a one-time analysis. Assessment programs may revisit topics periodically as changes warrant. To the extent that assessment reports document their information sources, assumptions and processes, assessment programs and decision makers can more readily recognize when it is time for reassessment.

Properties and Impacts Assessed

What is assessed in HCTA? HCTA may involve the investigation of one or more properties, impacts, or other attributes of health care technologies or applications. In general, these include the following.

- Technical properties
- Clinical safety
- Efficacy and/or effectiveness
- Economic attributes or impacts
- Social, legal, ethical and/or political impacts

Technical properties include performance characteristics and conformity with specifications for design, composition, manufacturing, tolerances, reliability, ease of use, maintenance, etc. **Safety** is a judgment of the acceptability of risk (a measure of the probability of an adverse outcome and its severity) associated with using a technology in a given situation, e.g., for a patient with a particular health problem, by a clinician with certain training and/or in a specified treatment setting.

Efficacy and effectiveness both refer to how well a technology works to improve patient health, usually based on changes in one or more pertinent health outcomes or "endpoints" as described below. A technology that works under carefully controlled conditions or with carefully selected patients under the supervision of its developers does not always work as well in other settings or as implemented by other practitioners. In HCTA, **efficacy** refers to the benefit of using a technology for a particular problem under ideal conditions, e.g., within the protocol of a carefully managed **randomized controlled trial**, involving patients meeting narrowly defined criteria, and/or conducted at a "center of excellence." **Effectiveness** refers to the benefit of using a technology for a particular problem under general or routine conditions, e.g., by a physician in a community hospital for a variety of types of patients.

Clinicians, patients, managers and policymakers are increasingly concerned about the practical implications of differences in efficacy and effectiveness. Researchers are delving into **registers**, **databases** (e.g., of claims and administrative data) and other epidemiological and observational data to discern possible associations between the use of technologies and patient outcomes in general or routine practice settings. The **validity** of any findings regarding causal connections between interventions and patient outcomes may be weakened to the extent that these data are not derived from prospective, randomized, controlled studies (U.S. Congress, OTA 1994). As discussed below, some newer prospective trials are designed to incorporate varied groups of patients and settings.

Box 3 shows certain distinctions in efficacy and effectiveness for diagnostic tests. Whereas the relationship between a preventive, therapeutic, or rehabilitative technology and its patient outcomes is

typically direct (though not always easy to measure), the relationship between a technology used for diagnosis or screening and its patient outcomes is typically indirect. This relationship and other concepts concerning assessment of diagnostic and screening technologies are addressed in Appendix A.

Health care technologies can have a wide range of microeconomic and macroeconomic attributes or impacts. Microeconomic concerns include costs, prices, charges and payment levels associated with individual technologies. Other concerns include comparisons of resource requirements and outcomes (or benefits) of technologies for particular applications, such as cost effectiveness, cost utility and cost benefit. (Methods for determining these are described below.)

Examples of macroeconomic impacts of health care technologies are: the impact of new technologies on national health care costs, the effect of technologies on resource allocation among different health programs or among health and other sectors, and the effects of new technologies on outpatient versus inpatient care. Other macroeconomic issues that pertain to health care technologies include the effects of regulatory policies, health care reforms and other policy changes on technological innovation, technological competitiveness, technology transfer and employment.

A variety of technologies raise social and ethical concerns. Such technologies as genetic testing, fertility treatments, transplantation of scarce organs and life-support systems for the critically ill challenge certain legal standards and societal norms. For example, in dialysis and transplantation for patients with end-stage renal disease, ethical concerns arise from patient selection criteria, termination of treatment and dealing with problem patients (e.g., who are non-compliant, hostile, or self-destructive) (Rettig and Levinsky 1991). Ethical questions continue to prompt improvement in informed consent procedures for patients involved in trials of investigative technologies. Allocation of scarce resources to technologies that are expensive, inequitably used or non-curative raises broad social concerns.

The terms "appropriate" and "necessary" often are used to describe whether or not a technology should be used in particular circumstances. For example, the appropriateness of a diagnostic test may depend on its safety and effectiveness compared to alternative available interventions for particular patient indications, clinical settings, and resource constraints. A technology may be considered necessary if withholding it would be deleterious to the patient's health (Hilborne et al. 1991; Kahan et al. 1994; Leape et al. 1991).

The properties, impacts, and other attributes assessed in HCTA pertain across the wide range of technologies noted above. Thus, for example, just as drugs, devices and surgical procedures can be assessed for safety, effectiveness and cost effectiveness, so can hospital infection control programs, computer-based drug-utilization review systems (Soumerai and Lipton 1995) and rural telemedicine networks.

Measuring Health Outcomes

Health outcome variables are used to measure the safety, efficacy and effectiveness of health care technologies. Health outcomes have been measured primarily in terms of changes in mortality (death rate) or morbidity (disease rate). For a cancer treatment, the main outcome of interest may be five-year survival; for treatments of coronary artery disease, the main endpoints may be **incidence** of fatal and nonfatal acute myocardial infarction and recurrence of angina.

In a clinical trial comparing alternative treatments, the effect on health outcomes of one treatment relative to another (e.g., a control treatment) can be expressed using various measures of **treatment effect**. These measures compare the probability of a given health outcome in the treatment group with the probability of the same outcome in a **control group.** Examples are **absolute risk reduction, odds ratio, number needed to treat,** and **effect size.** (See Box 4.)

Health-Related Quality of Life Measures

Although mortality and morbidity are usually the outcomes of greatest concern, they are not the only outcomes of importance to patients nor to others. Many technologies affect patients, family members, providers, employers, and other interested parties in ways that are not reflected in mortality or morbidity rates; this is particularly true for many chronic diseases.

Health-related quality of life (HRQL) measures (or indexes) are increasingly used along with more traditional outcome measures to assess health care technologies, providing a more complete picture of the ways in which health care affects patients. HRQL measures are designed to capture such dimensions as: physical function, social function, cognitive function, anxiety/distress, bodily pain, sleep/rest, energy/fatigue and general health perception. HRQL measures may be disease-specific (e.g., heart disease or arthritis) or general (covering overall health). They may be unidimensional (concerning one aspect such as distress) or multidimensional (Patrick and Deyo 1989). They may provide a single aggregate score or yield a set of scores, each for a particular dimension. HRQL measures are increasingly used by health product companies to differentiate their products from those of competitors, which may have virtually indistinguishable effects on morbidity for particular diseases (e.g., hypertension and depression) but may have different profiles of side effects that affect patients' quality of life (Tanouye 1995).

HRQL measures can be used to determine the effects of a technology on patients, to compare alternative technologies for their effects on patients with a particular problem or disability, or to compare different technologies' respective abilities to improve the quality of life of patients with different problems. Considerable advances have been made in the development and validation of these measures in the last 20 years. Box 5 shows dimensions of general HRQL measures that have been used extensively and that are well validated for certain applications. Box 6 shows aspects of selected disease-specific HRQL measures.

Quality-Adjusted Life Years

A unit of health care outcome that combines gains (or losses) in length of life with quality of life is the **quality-adjusted life year (QALY)**. QALYs represent years of life subsequent to a health care intervention that are weighted or adjusted for the quality of life experienced by the patient during those years (Torrance and Feeny 1989). QALYs provide a common unit for comparing the cost-effectiveness (in particular the cost-utility) of different health care interventions and ranking priorities for allocating scarce health care resources. Other units that are analogous to QALYs include **disability-adjusted life years (DALYs)** (World Development Report 1993) and **healthy-years equivalents** (HYEs) (Johannesson et al. 1993; Mehrez and Gafni 1993).

The scale of quality of life used for QALYs can be based on a HRQL measure such as one of those noted in Box 5 or other methods of eliciting patient **utility** for certain states of life. This dimension is typically standardized to a scale ranging from 0.0 (death) to 1.0 (perfect health); a scale may allow for

ratings below 0.0 for states of disability and distress that some patients consider to be worse than death (Patrick et al. 1994). QALYs can be useful for making comparisons among alternative technologies because they are generic units that can reflect changes brought about by different health care interventions for the same or different health problems. Box 7 shows how QALYs were used to compare the cost utilities of three alternative therapies for end-stage heart disease. Box 8 lists the cost utility of different interventions for different health problems according to the amount of money that must be invested per QALY gained.

Certain methodological aspects and the proposed use of QALYs or similar units in setting health care priorities remain controversial (Dougherty 1994; Gafni and Birch 1993; Gerard and Mooney 1993; La Puma and Lawlor 1990; Mason 1994; Nord 1994; Richardson 1994; Smith 1993). HRQL measures and QALYs are increasingly used in HCTA while substantial work continues in reviewing, refining and validating them.

Box 2
Examples of Health Care Technologies Found to be Ineffective or Harmful After Being Widely Diffused

Colectomy to treat epilepsy
Diethylstilbestrol (DES) to improve pregnancy outcomes
Electronic fetal monitoring during labor without access to fetal scalp sampling
Episiotomy (routine or liberal) for birth
Extracranial-intracranial bypass to reduce risk of ischemic stroke
Gastric bubble for morbid obesity
Gastric freezing for peptic ulcer disease
Hydralazine for chronic heart failure
Intermittent positive pressure breathing
Mammary artery ligation for coronary artery disease
Optic nerve decompression surgery for nonarteritic anterior ischemic optic neuropathy
Quinidine for suppressing recurrences of atrial fibrillation
Radiation therapy for acne
Thalidomide for sedation in pregnant women
Triparanol (MER-29) for cholesterol reduction

Sources include: Coplen et al. 1990; Enkin 1995; Feeny et al. 1986; Grimes 1993; The Ischemic Optic Neuropathy Decompression Trial Research Group 1995; Passamani 1991; U.S. DHHS 1990, 1993.

Box 3
Efficacy vs. Effectiveness for Diagnostic Tests

	EFFICACY	EFFECTIVENESS
PATIENT POPULATION	Homogeneous; patients with coexisting illness often excluded	Heterogeneous; includes all patients who usually have test
PROCEDURES	Standardized	Often variable
TESTING CONDITIONS	Ideal	Conditions of everyday practice
PRACTITIONER	Experts	All users

Adapted from: Institute of Medicine 1989

Box 4
Choice of Treatment Effect Measures Can Give Different Impressions

A study of the effect of breast cancer screening can be used to contrast several treatment effect measures and to show how they can give different impressions about the effectiveness of an intervention (Forrow et al. 1992). In 1988, Andersson et al. (1988) reported the results of a large RCT that was conducted to determine the effect of mammographic screening on mortality from breast cancer. The trial involved more than 42,000 women who were over 45 years old. Half of the women were invited to have mammographic screening and were treated as needed. The other women (control group) were not invited for screening.

The report of this trial states that "Overall, women in the study group aged ≥55 had a 20% reduction in mortality from breast cancer." Although this statement is true, calculation of other types of treatment effect measures provides important additional information. The table below shows the number of women aged ≥55 and breast cancer deaths in the screened group and control group, respectively. Based on these figures, four treatment effect measures are calculated.

For example, the *absolute risk reduction* is the difference in the rate of adverse events between the screened group and the control group. In this trial, the absolute risk reduction of 0.0007 means that the absolute effect of screening was to reduce the incidence of breast cancer mortality by seven deaths per 10,000 women screened, or 0.07%.

Group	Number of Patients	Deaths from breast cancer	Probability of death from breast cancer	Absolute risk reduction[1]	Relative risk reduction[2]	Odds ratio[3]	Number needed to screen[4]
Screened	13,107	35	$P_c = 0.0027$	0.0007	21%	0.79	1,429
Control	13,113	44	$P_c = 0.0034$				

Women in the intervention group were invited to attend mammographic screening at intervals of 18-24 months. Five rounds of screening were completed. Breast cancer was treated according to stage at diagnosis. Mean follow-up was 8.8 years.

1. Absolute risk reduction: $P_c - P_s$
2. **Relative risk reduction**: $(P_c - P_s) \div P_c$
3. Odds ratio: $[P_s \div (1 - P_s)] \div [P_c \div (1 - P_c)]$
4. Number needed to screen: $1 \div (P_c - P_s)$

Source of number of patients and deaths from breast cancer: Andersson et al. 1988.

Box 5
Dimensions of Selected General Health-Related Quality of Life Indexes

Sickness Impact Profile (Bergner et al. 1981; de Bruin et al. 1992)

- body care and movement
- ambulation
- mobility
- sleep and rest
- home management
- recreation and pastimes
- emotional behavior
- alertness behavior
- communication
- social interaction
- work
- eating

Nottingham Health Profile (Doll et al. 1993; Jenkinson et al. 1988)

- physical mobility
- pain
- sleep
- energy
- social isolation
- emotional reactions

Quality of Well-Being Scale (Kaplan and Anderson 1988; Kaplan et al. 1989)

- symptom-problem complex
- mobility
- physical activity
- social activity

Functional Independence Measure (Bunch and Dvonch 1994; Linacre et al. 1994)

- self-care
- sphincter control
- mobility
- locomotion
- communication
- social cognition

Short Form (SF)-36 (McHorney et al. 1994; Ware and Sherbourne 1992)

- physical functioning
- role limitations due to physical problems
- social functioning
- bodily pain
- general mental health
- role limitations due to emotional problems
- vitality
- general health perceptions

EuroQol Descriptive System (Essink-Bot et al. 1993; EuroQol Group 1990)

- mobility
- self-care
- usual activities
- pain/discomfort
- anxiety/depression

Katz Activities of Daily Living (Katz et al. 1970; Lazaridis et al. 1994)

- bathing
- dressing
- toileting
- mobility
- continence
- eating

Box 6
Selected Disease-Specific Health-Related Quality of Life Indexes

New York Heart Association Functional Classification (O'Brien et al. 1993; van den Broek et al. 1992)

Class I: Patients with cardiac disease but without resulting limitations of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, dyspnoea, or anginal pain.
Class II: Patients with cardiac disease resulting in slight limitation of physical activity. They are comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnoea or anginal pain.
Class III: Patients with cardiac disease resulting in marked limitation of physical activity. They are comfortable at rest. Less than ordinary physical activity causes fatigue, palpitation, dyspnoea or anginal pain.
Class IV: Patients with cardiac disease resulting in inability to carry on any physical activity without discomfort. Symptoms of cardiac insufficiency or of anginal syndrome may be present even at rest. If any physical activity is undertaken, discomfort is increased.

Arthritis Impact Measurement Scales (Kazis et al. 1989; Meenan et al. 1992)

- mobility
- walking and bending
- hand and finger function
- arm function
- self care
- household tasks
- social activities
- support from family and friends
- arthritis pain
- work
- level of tension

- mood

Visual Functioning (VF)-14 Index (Steinberg et al. 1994)

- reading small print, such as labels on medicine bottles, a telephone book, or food labels
- reading a newspaper or book
- reading a large-print book or newspaper or the numbers on a telephone
- recognizing people when they are close to you
- seeing steps, stairs, or curbs
- reading traffic, street, or store signs
- doing fine handwork such as sewing, knitting, crocheting, or carpentry
- writing checks or filling out forms
- playing games such as bingo, dominos, card games, or mahjong
- taking part in sports such as bowling, handball, tennis, or golf
- cooking
- watching television
- daytime driving
- nighttime driving

Box 7
Cost-Utilities for Alternative Therapies for End-Stage Heart Disease

Therapy	Life years gained(yr)	Mean utility	QALY (yr)	Aggregate cost($)	Cost per QALY($/yr)
Conventional medical treatment	0.50	0.06	0.03	$28,500	$950,000
Heart Transplantation	11.30	0.75	8.45	$298,200	$35,290
Total artificial heart	4.42	0.65	2.88	$327,600	$113,750

Notes: Costs and outcomes discounted at three percent per year; 20-year horizon. Mean utilities derived using time-tradeoff method on scale for which 1.0 was well, 0.0 was death, and states worse than death were valued between 0.0 and -1.0.

This table indicates that, although the cost of conventional medical treatment is the lowest, its cost per QALY is the highest, as the life-years gained and the patient utility of those years are low compared to the alternatives. The costs of heart transplantation and total artificial heart are of similar magnitude, but the cost per QALY is much lower for heart transplantation, as the life-years gained and the patient utility of those years are higher compared to the total artificial heart.

Source: Hogness and VanAntwerp 1991.

Box 8
Cost per QALY for Selected Health Care Technologies

	Cost per QALY (£ 1990)
Cholesterol testing and diet therapy (all 40-69 yrs)	220
Neurosurgery for head injury	240
General practitioner advice to stop smoking	270
Neurosurgery for subarachnoid hemorrhage	490
Antihypertensive therapy to prevent stroke (45-64 yrs)	940
Pacemaker implantation	1,100
Hip replacement	1,180
Valve replacement for aortic stenosis	1,140
Cholesterol testing and treatment	1,480
Coronary artery bypass graft surgery (left main disease, severe angina)	2,090
Kidney transplant	4,710
Breast cancer screening	5,780
Heart transplantation	7,840
Cholesterol testing and treatment (incremental) (all 25-39 yrs)	14,150
Home hemodialysis	17,260
Coronary artery bypass graft surgery (one-vessel disease, moderate angina)	18,830
Continuous ambulatory peritoneal dialysis	19,870
Hospital hemodialysis	21,970
Erythropoietin for dialysis anemia (with 10% reduction in mortality)	54,380
Neurosurgery for malignant intracranial tumors	107,780
Erythropoietin for dialysis anemia (with no increase in survival)	126,290

This table ranks selected procedures for a variety of health problems according to their cost utility, (i.e., the amount of money that must be spent on each procedure to gain one more QALY). There were some methodological differences in determining costs and QALYs among the studies from which these results were derived. Nonetheless, giving considerable latitude to these figures, the range in the magnitude of investment required to yield the next QALY for these treatments is great. This type of "bucks for the bang" (here, British pounds for the QALY) analysis helps to illustrate implicit choices made in allocating scarce health care resources, and suggests how decision makers might move toward reallocating those resources if societal gain in net health benefits (e.g., as measured using QALYs) is used as an allocation criterion.

Source: Maynard A. "Developing the Health Care Market." Economic Journal 1991.

Types of Organizations That Conduct HCTA

The types of organizations that undertake some form of HCTA include:

- regulatory agencies
- government and private sector payers

- health professions organizations
- standards-setting bodies
- hospitals, managed care organizations, and other providers
- patient and consumer organizations
- government policy research agencies
- private sector assessment/policy research organizations
- academic centers
- biomedical research agencies
- health product companies
- venture capital groups and other investors

As described throughout this monograph, the purposes, scope, methods and other characteristics of HCTAs that are conducted or sponsored by these organizations vary widely. Examples of these organizations are noted in this document. As in other fields, professional societies and organizational consortia exist in HCTA. At the international level, the International Society of Technology Assessment in Health Care (ISTAHC) [http://www.istahc.org], formed in 1985, now has about 1,200 members from more than 40 countries. The International Network of Agencies for Health Technology Assessment (INAHTA), formed in 1993, is a network of about 23 nonprofit organizations (government and private sector) that share assessment information and engage in related collaborative activities [http://www.sbu.se/sbu-site/links/inahta/index.html]. Examples of other professional organizations whose interests include areas related to HCTA include: the Association for Health Services Research (AHSR) [http://www.ahsr.org], the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) [http://www.ispor.org], and the Society for Medical Decision Making (SMDM) [http://polaris.nemc.org/SMDM]. (The ISTAHC web site provides numerous links to organizations active in HCTA.)

Expertise for Conducting HCTA

Given the variety of impacts addressed and the range of methods that may be used in an assessment, multiple types of experts are needed in HCTA. Depending upon the topic and scope of assessment, these may include a selection of the following:

- physicians, nurses, dentists, and other clinicians
- managers of hospitals, HMOs and other health care organizations
- radiology technicians, laboratory technicians and other allied health professionals
- biomedical and clinical engineers
- pharmacologists
- patients and patient affairs representatives
- epidemiologists
- biostatisticians
- economists
- lawyers
- social scientists
- ethicists
- decision scientists
- computer scientists/programmers
- librarians/information specialists

Certain individuals have expertise in more than one area. The set of participants in an assessment

depends upon its purpose, available resources and other factors. For example, the standing members of a hospital technology assessment committee might include: the chief executive officer, chief financial officer, physician chief of staff, director of nursing, director of planning, materials manager and director of biomedical engineering (Taylor 1994). Physician specialists and marketing, legal, patient affairs and additional analytical support staff could be involved as appropriate.

Selected Issues in HCTA

Quality of Care and HCTA

The relationship between HCTA and quality of care is often poorly understood. Although a thorough discussion of this subject is not possible here, the following are some definitions and fundamental relationships concerning these concepts.

Quality of care is a measure or indicator of the degree to which health care is expected to increase the likelihood of desired health outcomes and is consistent with standards of health care. HCTA and *quality assurance* are distinct yet interdependent processes that contribute to quality of care.

HCTA generates findings that add to our knowledge about the relationship between health care interventions and health care outcomes. This knowledge can be used to develop and revise health care standards, e.g., manufacturing standards, clinical laboratory standards, practice guidelines, and other agreed upon criteria, practices and policies regarding the performance of health care.

The purpose of quality assurance activities is to ensure that the best available knowledge concerning the use of health care to improve health outcomes is properly used. It involves the implementation of health care standards, including activities to correct, reduce variations in, or otherwise improve health care practices relative to these standards. Continuous quality improvement (CQI) and total quality management (TQM) (Gann 1994; Kritchevsky and Simmons 1991; Wakefield and Wakefield 1993) are among the contemporary systematic approaches to quality assurance that are being adapted for hospitals and other health care institutions. Such approaches include, for example, the identification of *best practices* and the use of **benchmarking** to develop improved **clinical pathways** or **disease management** for medical and surgical procedures, administrative operations, etc. (Barnes et al. 1994; Berkey 1994; Lenz et al. 1994).

Quality assurance involves a measurement and monitoring function, (i.e., *quality assessment*). **Quality assessment** is, primarily, an administrative means for determining how well health care is delivered in comparison with applicable standards or acceptable bounds of care. These standards or bounds may be grouped according to the structure of care (institutional, professional and physical characteristics), the process of care (content or nature of the health care delivered) and the outcomes of care (health status and well-being of patients) (Donabedian 1988). In detecting these differences, quality assessment can also call attention to the need for further HCTA or other investigations.

In summary, HCTA contributes knowledge used to set standards for health care, and quality assurance is used to determine the extent to which health care providers adhere to these standards (Lohr 1990; Lohr and Rettig 1988).

Outcomes Research and HCTA

In principle, outcomes research concerns any inquiry into the health benefits of using a technology for a particular problem under general or routine conditions. In practice, the term **outcomes research** has been used interchangeably with the term **effectiveness research** since the late 1980s to refer to a constellation of methods and characteristics that overlap considerably with HCTA. It has received increased attention in the U.S., particularly in the form of grants made to Patient Outcomes Research Teams (PORTs) by the Medical Treatment Effectiveness Program of the Agency for Health Care Policy and Research (AHCPR) [http://www.ahcpr.gov]. The attention given to outcomes/effectiveness research by government and, increasingly, the private sector (Lewin Group 1997) reflects greater demand for data on patient and provider experience with technologies beyond what can be learned from the limited number of carefully circumscribed trials (e.g., premarketing clinical trials for new drugs and devices).

Outcomes/effectiveness research has emphasized health problem-oriented assessments of care delivered in general or routine settings; interdisciplinary teams; a wide range of patient outcomes including mortality, morbidity, adverse events and HRQL measures; the use of nonexperimental data (e.g., from epidemiological studies and administrative data sets); variations in practice patterns and their relationship to patient outcomes; and patient roles in clinical decisionmaking. Currently, the scope of outcomes/effectiveness research is being expanded to include collection of experimental data on effectiveness, e.g., from large, simple trials conducted in general practice settings (described below).

Decentralization of HCTA

Although technology assessment may have originated as a centralized function conducted by federal government agencies or other national- or regional-level organizations, HCTA is increasingly a decentralized activity conducted by a great variety of organizations in the public and private sectors that make technology-related policy decisions (Rettig 1997). As noted above, a HCTA done from a particular perspective may not serve the technology-related policymaking needs of other perspectives. Even for the same technology or clinical problem, the needs of members of Congress, the FDA, a health product company, a hospital, a large managed care organization, a physician specialist, or others vary. These needs are heightened with increased economic responsibilities and pressures on these different parties.

The growth in decentralized HCTA activity has arisen not from a reduction in the level of centralized activity (e.g., drug and device regulation at the FDA, National Institutes of Health (NIH) consensus development conferences and the OTA heath program), but an expansion in activities elsewhere, particularly in the private sector. Increasingly, large health care providers and major health care product companies are establishing units devoted to "technology assessment," "pharmacoeconomics," "clinical effectiveness," "health outcomes research," and related areas. More insurance companies and other private sector payers have established formal programs to assess new procedures and other technologies in support of payment decisions. The number and magnitude of private firms and university centers involved in HCTA is increasing. Technology assessment committees are now common among medical specialty and subspecialty societies. Hospital networks, managed care organizations and other large health care providers in the private sector are establishing HCTA programs to support acquisition and management of pharmaceuticals, equipment and other technologies (Haynes et al. 1992), and other technology-related needs throughout their systems.

Tracking changes in the locus and magnitude of HCTA is confounded by a broadening connotation of

the term. Rather than referring only to the comprehensive inquiries involving broad societal impacts envisioned for the field in the 1960s, HCTA is now used to refer to almost any evaluation or analysis pertaining to health care technology. Much of the expansion of HCTA concerns meeting focused, immediate needs such as a coverage decision for a particular procedure, determination of the cost-effectiveness of a new device, or an equipment purchase decision. Another shift in locus concerns professional responsibility. Whereas technology-related decisionmaking in health care organizations was largely the responsibility of physicians, it is increasingly shared or redistributed among a wider spectrum of managers and other professionals.

Among the disadvantages of decentralized HCTA are lack of coordination, duplication of effort and inefficient resource allocation. Although multiple federal agencies are each involved in HCTA and related areas such as clinical trials, effectiveness research, clinical practice guidelines, cost-effectiveness analysis, methods development and registries of clinical trials, there is minimal coordination of setting priorities for these important and expensive activities or engaging in complementary or cooperative efforts. A 1994 OTA report (U.S. Congress, OTA 1994) provides numerous detailed options for better coordination among federal agencies, including designating single lead agencies for certain functions or requiring collaboration among agencies, e.g., requiring NIH institutes to give high priority to funding large clinical trials on topics identified by guideline panels, PORT findings, or advisory bodies at AHCPR. Other options concern specific cooperative assessment efforts with the private sector.

Certain changes in the health care market are prompting greater balance between centralized and decentralized HCTA efforts. Hospital networks, large managed care systems and other large health care systems such as that of the Department of Veterans Affairs (VA) are increasingly seeking to build economies of scale and buying leverage. With HCTA units that are centralized yet responsive to needs of individual facilities, these large organizations can consolidate their HCTA efforts and support system-wide acquisition of drugs, equipment and services. As health care providers and payers realize the level of costs required to conduct well-designed assessments using more rigorous methods, they become more willing to purchase assessment reports generated by other credible assessment organizations.

Barriers to HCTA

Although the general trend in health care is toward wider and improved HCTA, several countervailing forces to HCTA remain. Foremost, particularly in the U.S., has been a "**technological imperative**" comprising an abiding fascination with technology, the expectation that new is better, and the inclination to use a technology that has potential for some benefit, however marginal or unsubstantiated. Another is the sway of prestigious proponents of a technology in the absence of credible evidence. A third impediment is the inertia of medical practice, e.g., in the form of gravitation toward long-standing practice routines, conservative payment policies and quickly outdated education. This is complemented by lack of opportunities for, or encouragement of, scientific inquiry and skepticism in clinical education. Ever more effective marketing and promotions, including short courses sponsored by medical product companies to train physicians in procedures using these products, can divert attention from key concerns of HCTA. Another obstacle is the paltry level of investment, by government and industry sources in HCTA and related evaluations of what works in health care. In the U.S., the amount is lost in the rounding error of national health care expenditures. Finally, the impression persists in some quarters that the goal of HCTA is to limit the innovation and diffusion of health care technology.

HCTA and Underused Technologies

HCTA used in regulatory, payment, and utilization decisions poses market hurdles to health care technology. When used properly, HCTA can reduce or eliminate the use of technologies that are not safe and effective, or whose cost is too high relative to their benefits. As discussed above, HCTA can also be used to remove technologies from the market that are harmful or ineffective. Less attention is given to the ability of HCTA to identify technologies that are underused, and to help determine why they are underused. Box 9 shows examples of health care technologies for which good evidence exists of effectiveness or cost-effectiveness, but that are used significantly less than is indicated, even where they are affordable. Although this list applies primarily to the U.S., many of these technologies are underused elsewhere in North America, western Europe, and other of the wealthier countries.
The reasons that worthy technologies are underused are diverse, and include the following.

- Inadequate information dissemination
- Limited coverage and reimbursement
- Inappropriate concerns about improper use (e.g., pain therapy)
- Inconvenience and misperceptions on the part of clinicians or patients
- Professional conflict/turf battles
- Clinical inertia
- Insufficient supply (e.g., organ transplantation)
- Disproportionate concerns about adverse effects (e.g., warfarin to reduce risk of stroke)
- Concerns about patient **compliance** (e.g., polypharmacy for HIV/AIDS)
- Fear of stigma (treatment of depression)

Box 9
Underused Health Care Technologies (U.S.)

- ACE inhibitors for treatment of heart failure
- ACE inhibitors for prevention of renal deterioration in insulin-dependent diabetics
- Ambulation aids (canes, crutches, walkers)
- Antibiotics for gastrointestinal ulcers
- Beta blockers for survivors of acute myocardial infarction
- Cholesterol-lowering drugs for patients at risk of coronary artery disease
- Cochlear implants for severe-to-profound deafness
- Colorectal cancer screening
- Corticosteroid inhalants for treating asthma
- Corticosteroid therapy for fetuses at risk of preterm delivery
- Depression diagnosis and treatment
- Diabetic retinopathy screening
- Hepatitis B virus vaccination of infants
- Implantable cardioverter-defibrillators for survivors of cardiac arrest
- Incontinence diagnosis and treatment
- Intraocular pressure screening for glaucoma
- Oral rehydration therapy for dehydrated children
- Organ transplantation
- Pain management
- Polypharmacy (with protease inhibitors) for HIV/AIDS
- Pneumococcal vaccine for high risk patients

- Prenatal care
- Thrombolytic therapy for acute myocardial infarction
- Thrombolytic therapy for ischemic stroke
- Warfarin to prevent strokes due to atrial fibrillation

TEN BASIC STEPS OF HCTA

As described above, there is great variation in the scope, selection of methods and level of detail in the

practice of HCTA. Nevertheless, most HCTA activity involves some form of the following basic steps.

1. Identify assessment topics
2. Specify the assessment problem
3. Determine locus of assessment
4. Retrieve evidence
5. Collect new primary data (as appropriate)
6. Interpret evidence
7. Synthesize/consolidate evidence
8. Formulate findings and recommendations
9. Disseminate findings and recommendations
10. Monitor impact

Of course, not all assessment programs conduct all of these steps, or conduct them in sequence. Many programs use evidence from available sources only and do not collect primary data. Some assessment efforts involve multiple cycles of retrieving/collecting, interpreting and synthesizing evidence before completing an assessment. For example, to gain FDA approval for new drugs, pharmaceutical companies typically sponsor several iterations of new data collection: preclinical testing in the laboratory and in animals and **phase I, II and III studies** in humans; additional phase IV postmarketing studies may be a condition of approval. Depending upon the circumstances of a HCTA, the dissemination of findings and recommendations and monitoring of impact may not be parts of the HCTA itself, although they may be important responsibilities of the sponsoring program or parent organization.

Step One: Identify Assessment Topics

Organizations that conduct or sponsor HCTAs have only limited resources for this activity. With the great variety of potential assessment topics, how does an organization determine what it should assess? This section considers how assessment programs identify candidate assessment topics, and how programs can set priorities among these topics.

Identify Candidate Topics

To a large extent, assessment topics are determined, or at least bounded, by the mission or purpose of an organization. For instance, under statute, assessment topics of the OTA, an analytical support agency of Congress, are identified in requests of congressional committees. Third-party payers generally assess technologies on a reactive basis; a new medical or surgical procedure that is not recognized by payers as being standard or established may become a candidate for assessment. For the Health Care Financing Administration (HCFA), which administers the Medicare program, assessment topics arise in the form of requests for coverage policy determinations that cannot be resolved at the local level. These requests originate with regional Medicare contractors, beneficiaries, physicians, health product companies, professional associations and government entities. HCFA may request assistance in the form of an assessment by the Office of Health Technology Assessment (OHTA) in AHCPR.

The FDA [http://www.fda.gov] is systematically required to assess all new drugs and to assess health devices according to specific provisions made for particular classes of devices. For a new drug, a company normally files an **Investigational New Drug Application (IND)** with the FDA to begin to test the drug in people; later, following successful completion of necessary clinical trials, the company

files a **New Drug Application (NDA)** to seek FDA approval to market the drug. For certain medical devices (i.e., new "Class III" devices that sustain or support life, are implanted in the body, or present a potential risk of illness or injury), the **Investigational Device Exemption (IDE)** and **Premarketing Approval (PMA) Application** are analogous to the IND and NDA, respectively. The FDA is notified about many other devices when a company files a "510(k)" application seeking market approval based on a device's "substantial equivalence" to another device that has already received FDA marketing approval.

The Clinical Efficacy Assessment Program of the American College of Physicians (ACP) surveys ACP members, asking them to rank topics according to their relevance to medical practice. The American Medical Association's Diagnostic and Therapeutic Technology Assessment Program (DATTA) receives requests from clinicians and surveys program subscribers and other interested groups for candidate topics. The topics undertaken by ECRI's technology assessment service are identified by request of the service's subscribers, including payers, providers, and others.

Some topic identification involves broader scanning of technology information sources. For example, market research firms and other organizations that provide investment advice use databases on new drugs and devices, medical/technical journals and other publications, and company press releases and annual reports; they also send representatives to gather intelligence about new technologies at medical specialty meetings.

Setting Assessment Priorities

Some assessment programs have explicit procedures for setting priorities; others set priorities only in an ad hoc or vague way. Given very limited resources for assessment and increasing accountability of assessment programs to their respective stakeholders, it is important to articulate how assessment topics are chosen.

Most assessment programs have criteria for topic selection, although these criteria are not always explicit. Is it most important to focus on costly health problems and technologies? What about health problems that affect large numbers of people, or health problems that are life-threatening? What about technologies that cause great public controversy? Examples of selection criteria that might be used in setting assessment priorities are:

- high burden of morbidity and/or mortality
- large number of patients affected
- high unit or aggregate cost of a technology or health problem
- substantial variations in practice
- potential to improve health benefit/patient outcomes
- potential to reduce health risks
- sufficient research findings available upon which to base assessment
- scientific controversy or great interest among health professionals
- public or political demand
- need to make regulatory decision
- need to make payment decision
- available findings not well disseminated or adopted by practitioners

The priorities of a given assessment program may comprise technologies, health problems, projects, or some combination of these.

A systematic priority-setting process might include the following steps (Donaldson and Sox 1992; Lara and Goodman 1990):

1. Select criteria to be used in priority setting.
2. Assign relative weights to the criteria.
3. Identify candidate topics for assessment (e.g., as described above).
4. If the list of candidate topics is large, reduce it by eliminating those topics that would clearly not rank highly according to the priority setting criteria.
5. Obtain data for rating the topics according to the criteria.
6. For each topic, assign a score for each criterion.
7. Calculate a priority score for each topic.
8. Rank the topics according to their priority scores.
9. Review the priority topics to ensure that assessment of these would be consistent with the organizational purpose.

Processes for soliciting candidate assessment topics and ranking assessment priorities range from being highly subjective (e.g., informal opinion of a small group of experts) to quantitative (e.g., using a mathematical formula) (Donaldson and Sox 1992; Eddy 1989; Phelps and Mooney 1992). (See Box 10.) Of course, there is no single correct way to set priorities. Reviewing the process by which an assessment program sets its priorities, including the implicit and explicit criteria it uses in determining whether or not to undertake an assessment, can help to clarify its role and rationalize its operations.

Box 10
A Quantitative Model for Priority Setting

A 1992 report by the Institute of Medicine provided recommendations for priority setting to the Office of Health Technology Assessment of the Agency for Health Care Policy and Research. Seven criteria were identified:

- **prevalence** of a health condition
- burden of illness
- cost
- variation in rates of use
- potential of results to change health outcomes
- potential of results to change costs
- potential of results to inform ethical, legal, or social issues

The report offered the following formula for calculating a priority score for each candidate topic.
Priority Score = $W_1 \ln S_1 + W_2 \ln S_2 + \ldots W_7 \ln S_7$
where:

W is the relative weight of each of seven priority-setting criteria

S is the score of a given candidate topic for a criterion

ln is the natural logarithm of the criterion scores.

Candidate topics would then be ranked according to their priority score.

Source: Donaldson and Sox 1992.

Step Two: Specify the Assessment Problem

One of the most important aspects of an assessment is to specify clearly the problem(s) or question(s) to be addressed; this will affect all subsequent aspects of the assessment. A group conducting an assessment should have an explicit understanding of the purpose of the assessment and who the intended users of the assessment are to be. This understanding might not be established at the outset of the assessment; it may take more probing, discussion and clarification.

The intended users or target groups of the assessment should affect its content and presentation and dissemination of results. Health care professionals, patients, politicians, researchers, hospital managers, company executives and others have different interests and levels of expertise. They tend to have different concerns about the effects or impacts of health care technologies (health outcomes, costs, social and political effects, etc.). They also have different needs regarding the scientific or technical level of reports, the presentation of evidence and findings, and the format (e.g., length and appearance) of reports. (See Step Nine. Disseminate Findings and Recommendations.)

When the assessment problem and intended users have been specified, they should be reviewed by the leaders of the requesting agency or sponsors of the assessment. The discussion of the problem by the assessment group may have clarified or focused the problem in a way that differs from the original request. This clarification may prompt a reconsideration or restatement of the problem before the assessment proceeds.

Problem Elements

There is no single correct way to state an assessment problem. In general, an assessment problem could entail specifying at least the following elements: health care problem(s); patient population(s); technology(ies); practitioners or users; setting(s) of care; and properties (or impacts or health outcomes) to be assessed.

For example, a basic specification of one assessment problem would be:

- health care problem: management of moderate hypertension
- patient population: males and females, age >60 years, diastolic blood pressure 90-114 mm Hg, systolic blood pressure <240 mm Hg, no other serious health problems
- technologies: pharmacologic and nonpharmacologic treatments
- practitioners: primary care providers
- setting of care: outpatient care, self care
- properties, impacts, or outcomes: safety (including side-effects), efficacy, effectiveness and cost-effectiveness (especially cost-utility)

Step Three: Determine Locus of Assessment

The nature of an assessment problem will affect the determination of the most appropriate organization to conduct the assessment. Certainly, a comprehensive assessment addressing multiple attributes of a technology can be very resource intensive, requiring considerable and diverse expertise, data sources, and other resources.

Some organizations, such as major insurance companies, HMOs, and integrated health systems, may have internal HCTA programs. For example, in a large hospital or health plan, this might include a core staff and a multidisciplinary HCTA committee representing major clinical departments, nursing, pharmacy, allied health, biomedical engineering. This committee might interact with other committees such as pharmacy and therapeutics, strategic planning, and capital planning (University HealthSystem Consortium 1996).

Other organizations may rely more on assessment reports acquired from organizations that have devoted functions or otherwise specialize in HCTA, such as the American Medical Association DATTA Program, Agency for Health Care Policy and Research, Blue Cross and Blue Shield Association Technology Evaluation Center, Cochrane Collaboration, ECRI, Hayes Inc., MetaWorks Inc., TEMINEX (The HMO Group), and University HealthSystem Consortium [http://uhc.edu]. Depending upon the source HCTA organization, these assessment reports may be available at no cost, on a subscription basis, or for a specific price per report.

Health care decision makers can "make or buy" all or certain portions of HCTAs. Determination of the responsibility for sponsoring or conducting an assessment depends upon the nature of the problem, financial resources available, expertise of available personnel, time constraints, and other factors. For any assessment problem, an organization must determine the extent to which it will devote its resources to conducting the assessment itself or purchasing it from other sources. Some health care organizations commission selected components of an assessment, such as evidence retrieval and synthesis, and perform the other steps in-house.

Factors that influence the "make or buy" decision include the following (Goodman et al. 1996).

Is an existing assessment available? If an existing assessment is available, does it address the specific assessment problem of interest? Does it have a compatible perspective? How recently was it conducted? Is the methodology used sufficiently credible? Is the report worth its price?

If an existing assessment needs to be updated or is not available, do people in the organization have the time and expertise to perform the required data collection and analyses? If a synthesis of existing information is needed, does the organization have database searching capabilities and staff to review and interpret the literature? If new data are needed, does the organization have the requisite resources and expertise?

What methodology will be used? If, for example, a consensus development approach is preferred, does that consensus need to incorporate and reflect the opinions of the organization's own clinicians? Will local clinicians accept the results and report recommendations if they do not participate in the assessment?

Step Four: Retrieve Available Evidence

Information and Data Sources for HCTA

One of the great challenges in HCTA is to assemble the evidence -- the data, literature and other information -- that is relevant to a particular assessment. For very new technologies, this information may be sparse and difficult to find; for many technologies, it can be profuse, scattered and of widely varying quality. Literature searching and related evidence retrieval are integral to successful HCTA,

and the time and resources required for these activities should be carefully considered in planning any HCTA (Auston et al. 1994; Goodman 1993).

Types of Sources

Available information sources cover different, though often overlapping, sectors of health care information. Although some are devoted to health care topics, others cover the sciences more broadly. Multiple sources should be searched to increase the likelihood of retrieving relevant reports. The variety of types of sources that may be useful for HCTA include:

- computer databases of published literature
- computer databases of clinical and administrative data
- printed indexes and directories
- government reports and monographs
- reference lists in available studies, reviews and meta-analyses
- special inventories/registries of reports
- health newsletters and newspapers
- company reports and press releases
- World Wide Web sites
- colleagues and other investigators

Of course, the Internet is an extraordinarily broad and readily accesible medium that provides access to many of the information sources noted above.

Worldwide, publicly available computer databases for health care and biomedical literature number in the hundreds. Among these are various general types. For example, **bibliographic databases** have indexed citations for journal articles and other publications. **Factual databases** provide information in the form of guidelines for available from vendors in printed and electronic form.

Examples of clinical trial registries in particular clinical areas are the *AIDSTRIALS* database and the *PDQ* database on cancer treatment; both are available on *MEDLARS*. The NIH and VA maintain cross-topic registries of controlled clinical trials sponsored by those respective agencies (U.S. Congress, OTA 1994).

The Cochrane Collaboration [http://hiru.mcmaster.ca/cochrane/default.htm] is an international organization that prepares, maintains and disseminates systematic reviews of RCTs (and other evidence when appropriate) of treatments for many clinical conditions. Nearly 500 systematic reviews have been produced by some 40 Cochrane review groups in such areas as acute respiratory infections, breast cancer, diabetes, hypertension, infectious diseases, and pregnancy and childbirth. The Cochrane Collaboration produces the Cochrane Library, which includes the Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effectiveness (DARE) [http://www.york.ac.uk/inst/crd/info], Cochrane Controlled Trials Register, and Cochrane Review Methodology Database (Bero and Rennie 1995; Sheldon and Chalmers 1994)).

The great variety of databases of primary data include those on vital statistics, services utilization, payment claims, etc. (Antczak-Bouckoms et al. 1991). The Centers for Disease Control and Prevention provides CDC WONDER [http://wonder.cdc.gov/rchtml/Convert/data/AdHoc.html], an extensive online system of data sets and full-text sources on mortality, hospital discharges, cancer incidence, and other public health data (Friede 1993). NLM is developing a database of data sets that

is a categorical directory of discharge summaries, epidemiological surveys, disease registries, and other primary data sources (Frawley 1994). The World Health Organization database WHOSIS [http://www.who.ch/whosis.htm#databases] provides numerous sources of data and information on health statistics and epidemiology. Box 12 shows examples of widely used primary databases.

Most health sciences libraries have access to the major databases. In addition to the database producers, many databases are made available by vendors (companies that sell database services) such as BRS, CompuServe and DIALOG. These vendors acquire the rights to use the databases, often making changes in the format or packaging of the information.

Gray Literature

Much valuable information is available beyond the traditional sources. This "gray" or "fugitive" literature is found in industry and government monographs, regulatory documents, professional association reports and guidelines, market research reports, policy and research institute studies, spot publications of special panels and commissions, conference proceedings and other sources. Although the **gray literature** can be timely and cover aspects of technologies that are not addressed in mainstream sources, it is usually not subject to **peer review**, and must be scrutinized accordingly. Other useful literature sources are reference lists in literature reviews and meta-analyses (Goodman 1993) described below.

Publication Bias

Various forms of **bias** can affect the validity of HCTA. One reason for searching multiple types of sources for relevant reports is to overcome **publication bias**. Studies of the composition of the biomedical research literature have found imbalances in the publication of legitimate studies (Chalmers et al. 1990). For instance, positive studies (that find statistically significant treatment effects) are more likely than negative studies (that find no treatment effects) to be published in peer-reviewed journals (Dickersin and Min 1993;Dickersin 1997). A study sponsored by a health product company may be less likely to be submitted for publication if the findings are not favorable to the interests of the company. (Certain organizations with which health product companies contract to assess their products have policies of undertaking such work only if they have the right to submit reports of their studies for publication.) Making a special effort to identify studies outside of the mainstream journal literature helps to overcome publication bias.

Help for Searchers

Given the great number of databases and the variety in their scope, means of access, controlled vocabularies and search commands, it is advisable to consult health information specialists when planning literature searches. Improved indexing, textword searching, user-friendly interfaces, more powerful personal computers and other advances in medical informatics are helping assessors to retrieve valuable information.

In recent years, the NLM has undertaken to improve its **MeSH** (*Me*dical *S*ubject *H*eadings) **controlled vocabulary** (used to index and search literature in *MEDLARS* databases) in the related fields of HCTA and health services research. In cooperation with the Cochrane Collaboration and others, NLM has improved the indexing of citations in MEDLINE and other databases to improve identification of RCTs (Dickersin et al. 1994). More guidance is available for searching strategies for HCTA (Goodman 1993, Sackett et al. 1997).

A key NLM project with immediate and long-term implications for improving access to the literature is the Unified Medical Language System (UMLS). The purpose of UMLS is to help health professionals and researchers to retrieve and integrate electronic health information from a variety of sources, irrespective of the variations in the way similar concepts are expressed in different sources. UMLS can help users link information from patient record systems, bibliographic databases, factual databases, and expert systems. The UMLS Metathesaurus links multiple controlled vocabularies used in health care (e.g., ICD-9-CM, CPT, SNOMED, and DSM-IIIR), so that the terms familiar to users can be mapped to MeSH (or other vocabularies) to facilitate searching. UMLS is an 'electronic Rosetta Stone' involving 600,000 names for more than 250,000 concepts in 30 different vocabularies. It includes French, German, Spanish, and Portugese translations of MeSH. Users can browse and use UMLS vocabulary linkages in *Grateful Med.*

Computer databases are becoming increasingly user-friendly so that people who are not experts in information retrieval can use them more effectively and efficiently. Searchers of *MEDLINE* and certain other *MEDLARS* databases can use the *Grateful Med* software program, which functions as an interface between the searcher and the regular *MEDLARS* search commands, enabling non-experts to search with good results. Grateful Med is available in several formats, including Grateful Med for Windows [http://wwwindex.nlm.nih.gov/databases/gmwin.html], a DOS version, a Macintosh version, and Internet Grateful Med. In 1997, NLM introduced free MEDLINE searching via Internet Grateful Med [http://igm.nlm.nih.gov] and PubMed [http://ncbi.nlm.nih.gov/PubMed].

The search for pertinent existing evidence is normally one of the first major tasks of an assessment, and should be planned accordingly. Costs associated with evidence searches can be significant, coming in the form of staff time, computer access and acquisition of literature, data tapes and other documentation. Although access to *MEDLARS* and other public-source databases is generally inexpensive, using some specialized scientific and business databases can be more costly. Database vendors offer a variety of packages of databases and pricing algorithms for these. HCTA programs of such organizations as ECRI, the Blue Cross and Blue Shield Association and the American Medical Association sell their reports on a subscription basis. Some market research monographs and other reports oriented for health product companies, investors and other business interests are priced in the thousands of dollars.

Box 11
Some Computer Databases Useful for HCTA

National Library of Medicine databases (some of these are also available through vendors)
AIDSDRUGS: descriptions of substances used in AIDS-related trials
AIDSLINE: citations for recent AIDS literature
AIDS TRIALS: AIDS-related clinical trials
BIOETHICS: citations for bioethics literature
CANCERLIT: citations for journal articles in cancer
CATLINE: citations for books, monographs in biomedicine
DIRLINE: directory of organizations
HealthSTAR: citations for planning/administration health services research/technology assessment literature
HSRProj: ongoing health services research projects
HSTAT: full text of U.S. clinical practice guidelines, consensus development reports, technology assessment reports, etc.

MEDLINE: citations for biomedical journal articles
PREMEDLINE: basic citation information and abstracts of articles before they are indexed and placed into *MEDLINE*
PDQ: cancer treatment, supportive care, screening, prevention, clinical trials
PubMed: World Wide Web access to MEDLINE, integrated molecular biology databases, links to participating online journals and related databases

Cochrane Library
Cochrane Database of Systematic Reviews: systematic reviews of controlled trials on hundreds of clinical topics
Database of Abstracts of Reviews of Effectiveness (DARE): structured abstracts of systematic reviews from around the world, critically appraised by NHS Centre for Reviews and Dissemination
Cochrane Controlled Trials Register: bibliography of controlled trials included sources outside peer-reviewed journal literature
Cochrane Review Methodology Database: bibliography of sources on research synthesis

Others
EMBASE (Excerpta Medica data base): citations for biomedical journal articles (Elsevier)
EStoC (Elsevier Science Table of Contents): tables of contents of journal articles in science (Elsevier)
SCISEARCH: citations for scientific journal articles (Institute for Scientific Information)
Current Contents: tables of contents of journals in science, social sciences, and other fields (Institute for Scientific Information)

Box 12
Examples of Computer Databases of Primary Data (U.S.)

Public sector
CDC Wonder: public health and epidemiology (Centers for Disease Control)
Medicare Enrollment Database (HCFA)
Medicare National Claims History Repository (HCFA)
National Ambulatory Medical Care Survey (NCHS)
National Death Index (NCHS)
National Health Interview Survey: health services use, acute and chronic conditions and other health status, health promotion and disease prevention, etc. (NCHS)
National Health and Nutrition Examination Survey (NHANES): nutritional status, cardiovascular disease, diabetes, overweight, hypertension, etc.; functional status (NCHS)
National Hospital Discharge Survey (NCHS)
National Medical Expenditure Survey (NMES): health services use and expenditures, third party payer coverage, health status, etc. (AHCPR)
National Vital Statistics System: demographic, infant and maternal health, family data, pregnancy outcomes, cause of death (NCHS)
Patient Treatment File: sociodemographic, diagnostic, surgical, episode information for inpatient and extended care (Department of Veterans Affairs)

Private sector
American Hospital Association Tapes: costs, discharges, outpatient visits, hospital characteristics (American Hospital Association)
ARAMIS (Arthritis, Rheumatism, and Aging Medical Information System): disease registry (American Rheumatism Association)

Duke DataBank for Cardiovascular Disease (Duke University Medical Center)
HCIA Clinical Pathways Database: hospital drugs, imaging, laboratory, other services (HCIA)
HCIA Projected Inpatient Database: hospital discharge data (HCIA)
HELP System Data Base: clinical records (LDS Hospital, Salt Lake City, Utah)
MEDSTAT Market Scan Data Base: (MEDSTAT Systems)

Step Five: Collect New Primary Data

Compiling evidence for an assessment may entail collection of new primary data. In principle, an assessment program may determine that existing evidence is insufficient for meeting the desired policy needs, and that new studies are needed to generate data for particular aspects of the assessment. Indeed, this can be an iterative process. Once available, the new data are interpreted and synthesized with the existing body of evidence.

It is beyond the scope of this monograph to describe the planning, design and conduct of clinical trials, epidemiological studies and other investigations for collecting new primary data. There is a considerable and evolving literature on these subjects. Further, there is a newer, growing literature on priority setting and efficient resource allocation for clinical trials, and cost-effective design of clinical trials (Detsky 1990; Thornquist et al. 1993).

This section notes the range of responsibilities that assessment programs have in primary data collection, the types and relative rigor of primary data collection methods, and certain recent developments in improving and adapting these methods.

Varying Responsibilities for Primary Data Collection

In practice, the ability of most assessment programs to undertake new primary data collection, particularly clinical trials, is limited by such factors as programs' financial constraints, time constraints, proscription of conducting or sponsoring clinical studies and other aspects of the roles or missions of the programs.

In the U.S., major units of the NIH such as the National Cancer Institute and the National Heart, Lung and Blood Institute sponsor and conduct biomedical research, including clinical trials. Elsewhere at NIH, the Office of Medical Applications of Research coordinates the NIH Consensus Development Program but does not collect primary clinical data, although it occasionally surveys physician specialists and other groups for which the NIH assessment reports are targeted. The Cooperative Studies Program of the VA Medical Research Service coordinates more than 50 ongoing multicenter clinical trials at hospitals in the VA system. Similarly, the VA's Cooperative Studies in Health Services Program collects primary data on delivery, quality and cost of health care in the VA system. The FDA does not typically conduct **primary studies** related to the marketing of new drugs and devices; rather, the FDA reviews primary data from studies sponsored or conducted by the companies that make these technologies.

Organizations such as the congressional OTA, the AHCPR OHTA, the nonprofit private Institute of Medicine (part of the National Academy of Sciences) and health professions societies such as the American College of Physicians and American College of Cardiology work almost exclusively from existing data and do not do clinical research.

Third-party payers generally do not sponsor clinical studies, but increasingly analyze claims data and

other administrative data. Payers have supported trials of new technologies indirectly by paying for care associated with clinical studies of those technologies, or by paying unintentionally for uncovered new procedures that were coded as covered procedures. As noted above, some payers are providing conditional coverage for certain investigational technologies in selected settings in order to compile data that can be used to make more informed coverage decisions (Peters and Rogers 1994).

Whether or not an assessment involves collection of new primary data, the assessment report should note what new primary studies should be undertaken to meet future assessment needs.

Methods for Primary Data Collection

Certain attributes of primary studies produce better evidence than others. In general, the following attributes are associated with stronger evidence.

- Prospective studies are superior to retrospective ones.
- Controlled studies are superior to uncontrolled ones.
- Randomized studies are superior to nonrandomized ones.
- Large studies (i.e., involving enough patients to detect with acceptable confidence levels any true treatment effects) are superior to small ones.
- Contemporaneous controls are superior to historical ones.
- Blinded studies (patients, clinicians, analysts) are superior to unblinded ones.

Basic types of methods for generating new data on the effects of health care technology in humans include the following:

- large randomized controlled trial (RCT)
- small RCT
- nonrandomized trial with contemporaneous controls
- nonrandomized trial with **historical control**s
- **cohort study**
- **case-control study**
- **cross-sectional study**
- surveillance (e.g., using **databases, registers**, or surveys)
- **series** of consecutive cases
- single case report

These methods are listed in rough order of most to least scientifically rigorous for internal validity, (i.e., for accurately representing the causal relationship between an intervention and an outcome in the particular circumstances of a study). (Types of methodological validity are described below.) This ordering of methods assumes that each study is properly designed and conducted. This list is representative; there are other variations of these methodologic designs and some investigators use different terminology for certain methods. The demand for studies of higher methodological rigor is increasing among health care technology regulators, payers, providers and other policymakers.

Methods of Cost Analysis

Studies of costs and related economic implications comprise a major group of methods used in HCTA. These studies can involve attributes of either or both of primary data collection and synthetic methods.

Interest in cost analyses has accompanied concerns about rising health care costs, pressures on health care policymakers to allocate resources, and the need for health product makers and other technology advocates to demonstrate the economic benefits of their technologies. This interest is reflected in a considerable increase in the number of reports of cost analyses in the literature and further refinement of methods.

There is a variety of approaches to cost analysis, the suitability of any of which depends upon the purpose of an assessment and the availability of data and other resources. It is rarely possible or necessary to identify and quantify all costs and all benefits (or outcomes), and the units used to quantify these may differ. Main types of cost analysis include the following.

- *cost-of-illness analysis:* a determination of the economic impact of an illness or condition, e.g., smoking, arthritis or bedsores, including associated treatment costs;
- *cost-minimization analysis:* a determination of the least costly among alternative interventions that are assumed to produce equivalent outcomes;
- *cost-effectiveness analysis (CEA):* a comparison of costs in monetary units with outcomes in quantitative non-monetary units, e.g., reduced mortality or morbidity;
- *cost-utility analysis (CUA):* a form of cost-effectiveness analysis that compares costs in monetary units with outcomes in terms of their utility, usually to the patient, measured, e.g., in QALYs; and
- *cost-benefit analysis (CBA):* compares costs and benefits, both of which are quantified in common monetary units.

Cost-minimization analysis, CEA and CUA necessarily involve comparisons of alternative interventions. A technology cannot be simply cost effective, though it may be cost effective compared to something else. Although CBA typically involves comparisons of alternative technologies, this is not necessary. In principle, a technology is cost beneficial as long as its costs are outweighed by the monetary value of its benefits or outcomes. (See discussion of CBA following).

Cost-Benefit Analysis: Two Basic Approaches

Two basic approaches for cost-benefit analysis (CBA) are *net benefits* and *benefit/cost ratio.* The basic formulation for net benefits is:

$$V = \sum_{t=0}^{T} \frac{B_t - C_t}{(1+r)^t}$$

The formulation for benefit/cost ratio is:

$$V = \sum_{t=0}^{T} \frac{B_t}{(1+r)^t} \div \sum_{t=0}^{T} \frac{C_t}{(1+r)^t}$$

For either formulation:

V is the present value of a technology;
the time-frame of analysis extends from t = 0 (the present) to some future t = T;

$\sum$ means summation across all time periods t;

B is the level of benefit (or outcome) for each time period t;

C is the level of costs for each time period t;

r is the discount rate with respect to the present (e.g., 5 percent per year).

Net benefits indicate the absolute amount of money saved or lost due to a technology. In the net benefits formulation, a technology is cost-beneficial if $V > 0$. Benefit/cost ratios indicate the amount of benefits (or outcomes) that can be realized per unit expenditure on a technology. In the benefit/cost ratio formulation of CBA, a technology is cost-beneficial if $V > 1$.

The general formulation for benefit/cost ratio applies to cost-effectiveness analysis (CEA) and cost-utility analysis (CUA), although the criterion of $V > 1$ is not relevant. The net benefits formulation does not apply to CEA or CUA because these analyses use different units for costs and benefits (outcomes).

The choice between a net benefits approach or a benefit/cost approach for a CBA can affect findings. The approach selected may depend upon such factors as whether costs must be limited to a certain level, whether the intent is to maximize the absolute level of benefits, whether the intent is to maximize the benefit/cost ratio regardless of the absolute level of costs, etc. Indeed, under certain circumstances these two basic approaches may yield different preferences among alternative technologies, as shown below.

Technology	Benefits (B)	Costs (C)	Approach			
			Net Benefit B - C	Prefer	Benefit/Cost Ratio B ¸ C	Prefer
A	1,000	800	200	ü	1.25	
B	200	100	100		2.00	ü

Because it measures costs and outcomes in monetary terms, CBA enables comparison of disparate technologies, e.g., coronary artery bypass graft surgery and screening for breast cancer. A drawback of CBA is the difficulty of assigning monetary values to all pertinent outcomes, including changes in the length or quality of human life. CEA avoids this limitation by using more direct or natural units of outcomes such as lives saved or strokes averted. As such, CEA can only compare technologies whose outcomes are measured in the same units. In CUA, estimates of utility are assigned to health outcomes, enabling comparisons of disparate technologies. (See discussion of QALYs and CBAs, above, and Box 8). In HCTA, CEA and, increasingly, CUA are used more than CBA.

The approaches to accounting for costs and outcomes in cost analyses can vary in a number of important respects, some of which are addressed briefly below. These should be carefully considered by assessors, as well as the policymakers who intend to make use of assessment findings.

The *perspective* of a cost analysis refers to the standpoint at which costs and benefits are realized. For

instance, the perspective of an analysis may be that of overall society, a third party payer, a health product company, a physician, or the patient or consumer. Clearly, costs and benefits are not realized in the same way from each of these perspectives. Many analysts favor using the broad perspective of society and identifying all costs and all benefits accordingly. However, what is cost effective from the perspective of a national agency (if it is assumed to have a societal perspective) may not be what is cost effective from the standpoint of a hospital manager or a patient.

Depending upon the perspective taken, cost analyses should identify **direct costs** of health care (physician fees, hospital fees, payments for drugs and equipment, etc.) and other types of direct costs such as for care provided by family members and transportation to and from the site of care. Analyses should account for **indirect costs** due to morbidity (e.g., the cost of time lost from work and decreased productivity) and mortality (premature death). **Intangible costs** of pain, suffering, and grief are very difficult to measure and are often omitted from cost analyses. Many analyses use readily available hospital or physician *charges* rather than true *costs*, whose determination may require special analyses of resource consumption. However, charges (as well as actual payments) tend to reflect provider cost shifting and other factors that decrease the validity of using charges to represent the true costs of providing care.

Interpretation of cost analyses must consider that the *time-frame* (or time horizon) of a study is likely to affect the findings regarding the relative magnitude of the costs and outcomes of a health care intervention. Costs and outcomes usually do not accrue in steady streams over time. Comparisons of costs and outcomes after one year may yield much different findings than comparisons made after 5, 10, or 25 years. For example, a recent HCFA analysis showed that the annualized expenditure by its Medicare End-Stage Renal Disease Program for a dialysis patient was $32,000. Although patients with functioning transplanted kidneys required a first-year expenditure of $56,000, they cost Medicare only an average of $6,400 in succeeding years. On average, estimated cumulative dialysis and transplantation costs reach a break-even point in about three years, after which transplantation provides a net financial gain compared to dialysis (Rettig and Levinsky 1991).

Assessments should make clear whether *average costs* or **marginal costs** are being used in the analysis. Whereas average cost analysis considers the total costs and outcomes of an intervention, marginal cost analysis considers how outcomes change with changes in costs, which may provide more information about how to use resources efficiently. Marginal cost analysis may reveal that, beyond a certain level of spending, the additional benefits are no longer worth the additional costs. (See Box 13.) It may be that some combination of interventions may be more cost-effective than using one in all instances. Cost analyses should account for the effect of the passage of time on the value of costs and outcomes. Costs and outcomes that occur in the future usually have less present value than costs and outcomes realized today. Thus, costs and outcomes should be *discounted* relative to their present value (e.g., at a rate of five percent per year). Analysis should also correct for the effects of *inflation* (which is different from discounting), such as when costs or cost-effectiveness for one year are compared to another year. For various reasons, there may be uncertainty associated with the estimates of certain costs, outcomes and other variables used in a cost analysis. Therefore, **sensitivity analysis** should be performed to determine if plausible variations in the estimates of these variables affect the results of the analysis. Given the different ways in which costs and outcomes may be determined, all studies should make clear their methodology in the respects noted above.

The validity of a cost-related study depends upon the sources of the data for costs and outcomes. Increased attention is being given to collection of cost data in more rigorous, prospective studies, particularly RCTs. The closer integration of economic and clinical studies raises important

methodological issues (Poe, 1995). In order to promote more rational diffusion of new technologies, it would be desirable to generate reliable cost and outcomes data during the early part of a technology's lifecycle, such as during RCTs required prior to marketing approval. An RCT would be expected to yield the most reliable data concerning efficacy of an intervention; however, the care given in an RCT and the costs of providing it may be atypical compared to more general settings. Other aspects of trial design, **sample size**, choice of outcome measures, identification and tabulation of costs, burden on investigators of data collection and related matters affect the usefulness of clinical trial data for meaningful economic studies.

In practice, there is wide variation in economic study methodologies (Elixhauser 1993). Although some variation is unavoidable, many differences in perspective, accounting for direct and indirect costs, time frames, **discounting** and other aspects are often arbitrary, result from lack of expertise, and may reflect biases on the part of investigators or sponsors. This diminishes comparability and transferability of study results as well as credibility of findings. National and international groups have developed and revised voluntary standards for conducting and reporting economic studies, particularly for pharmaceuticals (CCOHTA 1994; Drummond and Jefferson 1996; Drummond 1993; Gold et al. 1997; Joint Strategy Group 1994; Langley 1996; Ontario 1995; Rovira 1994).

Box 13
Average Cost Analysis vs. Marginal Cost Analysis

The importance of determining marginal costs is apparent in the analysis by Neuhauser and Lewicki of a proposed protocol of sequential stool guaiac testing for colon cancer. Here, average cost figures obscure a steep rise in marginal costs of testing because the high detection rate from the initial tests is averaged over subsequent tests that contribute little to the detection rate. This type of analysis helps to demonstrate how it is possible to spend steeply increasing health care resources for diminishing returns in health benefits.

Cancer screening and detection costs with sequential guaiac tests

No. of tests	No. of cancers detected	Additional cancers detected	Total cost ($) of diagnosis	Additional ($) cost of diagnosis	Average cost ($) per cancer detected	Marginal cost ($) per cancer detected
1	65.9469	77,511	77,511	1,175	1,175	
2	71.4424	5.4956	107,690	30,179	1,507	5,492
3	71.9004	0.4580	130,199	22,509	1,810	49,150
4	71.9385	0.0382	148,116	17,917	2,059	469,534
5	71.9417	0.0032	163,141	15,024	2,268	4,724,695
6	71.9420	0.0003	176,331	13,190	2,451	47,107,214

This analysis assumed that there were 72 true cancer cases per 10,000 population. The testing protocol provided six stool guaiac tests per person to detect colon cancer. If any one of the six tests was positive, a barium-enema test was performed, which was assumed to yield no false-positive and no false-negative results. Other assumptions: the true-positive cancer detection rate of any single guaiac test was 91.667%; the false-positive rate of any single guaiac test was 36.508%; the cost of the first stool guaiac test was $4 and each subsequent guaiac test was $1; the cost of a barium-enema was $100. The marginal cost per case detected depends on the population screened and the sensitivity of the test used.

Source (permission requested): Neuhauser and Lewicki 1975.

RCTs Versus Observational Studies: Tradeoffs of Validity

Although primary study investigators and assessors would prefer to have methods that are both internally and externally valid, they often find that study design attributes that increase one type of validity jeopardize the other. A well designed and conducted RCT is widely considered to be the best approach for ensuring **internal validity**, as it gives investigators the most control over factors that could confound the relationship between the use of a technology and health outcomes. However, for the reasons that a good RCT has high internal validity, its **external validity** may be limited.

Most RCTs are designed to investigate the effects of an intervention in specific types of patients so that the relationship between the intervention and outcomes is less likely to be confounded by patient variations. However, findings from an RCT involving a narrowly defined patient group may not be applicable for the same intervention given to other types of patients. As noted above, RCTs often involve special protocols of care and testing that may not be characteristic of general care, and are often conducted in university medical centers or other special settings that do not represent the general or routine settings in which most health care is provided.

Findings of some large **observational studies** (e.g., from large cross-sectional studies or registries) have external validity to the extent that they can provide insights into the types of outcomes that are experienced by different patient groups in different circumstances. However, these weaker designs are subject to biases that threaten internal validity, diminishing the certainty with which particular outcomes can be attributed to an intervention. Interesting or promising findings from weaker studies can raise hypotheses that can be tested using stronger studies. The recent use of "large, simple trials" (discussed below) is an attempt to bridge the strengths of RCTs and observational studies.

Developments in Primary Data Collection Studies

Applications of primary data collection methods evolve. In particular, investigators have made progress in trying to combine some of the desirable attributes of RCTs and observational studies. For example, while retaining the methodological strengths of prospective, randomized design, "**large, simple trials**" use large numbers of patients, more flexible patient entry criteria and multiple study sites to improve external validity and gain effectiveness data. Also, fewer types of data may be collected for each patient, easing participation by patients and clinicians (Buring 1994; Ellenberg 1992; Peto et al. 1993; Yusuf et al. 1990). Examples of these approaches include the recent large, simple trials supported by NIH; certain large, multicenter RCTs coordinated by the VA Cooperative Studies Program, and "firm trials" involving random assignment of patients and providers to alternative teams to evaluate organizational and administrative interventions (Cebul 1991).

Biomedical research organizations such as NIH and regulatory agencies such as FDA are permitting certain mid-trial changes in clinical trial protocols such as drug dosage modifications and patient cross-overs to alternative treatment groups to reflect the most recent scientific findings. Selected use of **surrogate endpoints**, especially biologic markers, is employed where these might indicate technological effectiveness years before meaningful differences in "hard" endpoints of mortality or morbidity could be detected. Streamlining or combining clinical trial phases and "parallel track" availability of technologies to patients outside of ongoing formal RCTs are intended to speed regulatory approval and make technologies available to patients who are ineligible for RCT protocols

but have exhausted other treatments. Although many of these adaptations have been instituted for RCTs involving new drug treatments for AIDS, cancer and other life-threatening conditions, their use in trials of other technologies is expected (Merigan 1990). Another important type of development in primary data collection is the incorporation of contemporaneous cost data collection in prospective clinical trials. For example, health care product companies increasingly are using such data in product promotion and to help secure favorable payment decisions.

Types of Methodological Validity

Methodological designs vary in their ability to produce valid results. Validity concerns whether what we are measuring is what we intend to measure. Understanding different aspects of validity helps in comparing alternative methodological designs and interpreting the results of studies using those designs.

Internal validity refers to the extent to which the findings of a study accurately represent the causal relationship between an intervention and an outcome in the particular circumstances of an investigation. This includes the extent to which a study minimizes any systematic or non-random error in the data.

External validity refers to the extent to which the findings obtained from an investigation conducted under particular circumstances can be generalized to other circumstances. To the extent that the circumstances of a particular investigation (e.g., patient characteristics or the manner of delivering a treatment) differ from the circumstances of interest, the external validity of the findings of that investigation may be questioned.

Other concepts related to measurement validity are the following. *Face validity* is the ability of a measure to represent reasonably a construct (i.e., a concept or domain of interest) as judged by someone with expertise in the health problem and interventions of interest. *Content validity* refers to the degree to which a measure covers that range of meanings or dimensions of a construct. In health care, an outcome measure is often used as a marker or surrogate for a disease of interest. For example, how well do lowered T-4 cell (a type of white blood cell) counts represent the severity of AIDS? How well does performance on an exercise treadmill represent cardiovascular fitness?

Construct validity is the ability of a measure to correlate with other accepted measures of the construct of interest, and to discriminate between groups known to differ according the varible. *Convergent validity* refers to the extent to which two different measures that are intended to measure the same construct do indeed yield similar results. *Discriminate validity*, opposite convergent validity, concerns whether different measures that are intended to measure different constructs do indeed fail to correlate with each other. *Concurrent validity* refers to the ability of a measure to accurately differentiate between different groups at the time the measure is applied, or the correlation of one measure with another at the same point in time. *Predictive validity* refers to the ability to use differences in a measure to predict future events or outcomes.

Step Six: Interpret Evidence

A challenge to any HCTA is to derive substantial findings from scientific evidence drawn from different types of studies of varying quality. Assessors should use a systematic approach to critically appraise the quality of the available studies (Eddy 1992; Goodman 1993).

Interpreting evidence requires knowledge of investigative methods and statistics. Assessment groups should include members who are knowledgeable in these areas. Some assessment programs assign content experts and evidence evaluation experts to prepare background papers that present and appraise the available evidence for use by assessment groups. In general, evidence interpretation involves classification of studies, grading the evidence and selection of studies to use.

Classification of Studies

As described above, certain attributes of primary study designs produce better evidence than others. A useful step in reviewing evidence is to classify it by basic methodologic type and other study characteristics.

Evidence tables provide a useful way to summarize and display important qualities about available studies. The information summarized in evidence tables may include attributes of study design (e.g., **randomization**, control, **blinding**, patient characteristics (e.g., number, age, gender), patient outcomes (e.g., mortality, morbidity, HRQL) and derived summary statistics (e.g., ***P* values, confidence intervals**). The tabular format enables reviewers to compare systematically the key attributes of studies and to provide an overall picture of the amount and quality of the available evidence.

Box 14 is an evidence table of selected study characteristics and outcomes of double-blind **placebo**-controlled RCTs of aspirin for patients after myocardial infarction. Box 15 shows a count of articles published during 1980-1990, arranged by type of study, about the clinical use of percutaneous transluminal coronary angioplasty. This evidence table does not provide details about individual studies, but it does show the distribution of different types of methods for evaluating this technology.

Grading the Evidence

"Grading" evidence that is gathered in a comprehensive literature search according to its methodological rigor is increasingly becoming a standard part of HCTA (Evidence-Based 1992). It can take various forms, each of which involves structured, critical appraisal of the evidence against formal criteria.

Box 16 shows a basic evidence-grading scheme used by the U.S. Preventive Services Task Force. This scheme grades evidence in a manner that favors certain attributes of stronger studies for primary data, beginning with properly-designed RCTs. Box 17 shows an application of another evidence-grading scheme used by a panel convened by the AHCPR to develop clinical practice guidelines for management of cancer pain. In this scheme, the strongest evidence is a meta-analysis (a quantitative consolidation of studies, described below) of multiple, well-designed controlled studies; the weakest evidence includes case reports and clinical examples (anecdotes). Assessment groups can use hierarchies such as these to begin sorting the various types of available studies or to summarize a body of evidence. Still, more information may be needed to compare studies in a useful way. For example, this type of grading scheme does not account for instances where two or more good studies have conflicting results (Eddy 1992).

As noted above, it is not only the basic type of a study (e.g., RCT or case-control study) that affects the quality of the evidence, but the way in which the study was designed and conducted. There are systematic ways to evaluate the quality of individual studies. For example, there are detailed

quantitative grading schemes for rating the relative methodological quality of RCTs (Chalmers et al. 1981). A simple checklist for RCTs is shown in Box 18. The ability of analysts to determine the internal and external validity of a published study and to otherwise interpret its quality depends on how thoroughly and clearly the information about its study's design, conduct, statistical analysis, and other aspects are reported. The inadequate quality of a high proportion of published reports of RCTs, even in leading journals, is well documented (Freiman et al. 1978; Moher et al. 1994). Several national and international groups of researchers and medical journal editors have developed standards for reporting of RCTs and other studies (Begg et al. 1996; International Committee of Medical Journal Editors 1997). Also, the trend of more journals to require structured abstracts will assist analysts in identifying and screening reports of RCTs and other studies.

Many primary studies of health care technologies involve small, non-randomized series of consecutive cases or single case reports, and therefore have methodological limitations that make it difficult to establish the efficacy (or other attributes) of the technologies with sound scientific validity. To some extent, these methodological shortcomings are unavoidable given the nature of the technologies being evaluated, or are otherwise beyond the control of the investigators. In the instance of determining the efficacy of a new drug, the methodological standard is a large, prospective, double-blind, placebo-controlled RCT. These methodological attributes increase the chances of detecting any real treatment effect of the new drug, control for patient characteristics that might influence any treatment effect, and reduce opportunities for investigator or patient bias to affect results.

Although their contributions to methodological validity are generally well recognized, it is not possible to apply all of these attributes for trials of certain types of technologies or for certain clinical indications or settings. Further, these attributes are controversial in certain instances. Patient and/or investigator blinding is impractical or impossible for many medical devices. For clinical trials of technologies for rare patient problems (e.g., "orphan" drugs and devices), it may be difficult to recruit numbers of patients large enough to detect convincing treatment effects.

Considerable controversy exists over the appropriate use of placebo controls in clinical trials. Among the issues are: (1) appropriateness of using a placebo in a trial of a new therapy when a therapy judged to be effective already exists, (2) statistical requirements for discerning what may be smaller differences in outcomes between a new therapy and an existing one compared to differences in outcomes between a new therapy and a placebo, and (3) concerns about comparing a new treatment to an existing therapy that, except during the trial itself, may be unavailable in a given setting (e.g., a developing country) because of its cost or other economic or social constraints (Rothman and Michels 1994; Varmus and Satcher 1997).

Notwithstanding the limitations inherent in clinical study of many technologies, the methodological rigor used in many primary studies falls short of what it could be. From clinicians to hospital financial officers to congressional aides, those who affect technology policy in all sectors of health care are becoming more sophisticated in demanding and interpreting the strength of scientifically-based findings for health care technologies.

Decide How to Use Studies

After reviewing the individual studies for their methodological quality, an assessment group may judge that it would not be appropriate to consider all studies equally important. The group may decide that studies of higher quality should influence their findings more than studies of lesser quality.

Experts in evidence interpretation do not agree on the proper approach for deciding how to use studies of differing quality. According to some experts, the results of studies that do not have randomized controls are subject to such great bias that they should not be included for determining the effects of an intervention. Others say that studies from nonrandomized prospective studies, observational studies and other weaker designs should be used, but given less weight or adjusted for their biases (Chalmers et al. 1989).

There are several basic approaches to deciding how to use the individual studies in an assessment. These are: use all studies as reported; decide whether to include or exclude each study as reported; weight studies according to their relative quality; and make adjustments to the results of studies to compensate for their biases. Each approach has strengths and weaknesses, as well as differing technical requirements. It is important for the assessment group to document in its report the criteria or procedures by which it chose to make use of study results for use in the assessment.

Conflict of Interest and Bias

Interpreting the literature for an assessment should include consideration of the existence of potential conflicts of interest that may have affected the conduct of a study or presentation of results. For study investigators, conflicts of interest may arise from having a financial interest (through, e.g., salary support, ongoing consultancy, owning stock, owning a related patent) in a health product company (or one of its competitors) that may be affected by the results of a study or being an innovator of a technology under study. Peer-reviewed journals increasingly require disclosure of information pertaining to financial interests of investigators and the source of funding of studies (Conflict of interest 1993; Kassirer and Angell 1993). Information about investigators, sponsorship of a study, or other factors that suggests the potential for conflict of interest should be considered when interpreting the evidence. Studies that are subject to potential conflicts of interest may have to be discounted or dropped from the body of evidence accordingly. HCTA programs should take active measures to ensure that members of assessment teams and panels do not have potential conflicts of interest.

The existence of potential conflicts of interest and bias on the part of individual members of an assessment team or panel should be carefully considered. A conflict of interest may be any financial or other interest that conflicts with one's service on an assessment group because it could impair that person's objectivity or could create an unfair advantage. Bias may entail views or inclinations that are intellectually motivated or that would be expected to arise from having a given organizational or professional affiliation. HCTA programs should take active measures to minimize or balance bias among panel members.

Box 14
Evidence Table for Double-Blind Placebo-Controlled Randomized Trials of Aspirin in Patients After Myocardial Infarction

Trial,	No. patients	Age range Male		Months from qualifying event	Daily dose	Average follow-up	Mortality	Cardiac
Year	randomized	(mean)	%	to trial entry	ASA* (mg)	(years)	(%)† Sum. stat.‡	(%) Sur
AMIS	ASA: 2,267	30-69	89	2 - 60	1,000	3.2	10.8 Z=1.27	8.7 Z=(
1980	plac:2,257	(54.8)					9.7	8.0
Breddin	ASA: 317	45-70	78	1 - 1.4	1,500	2.0	8.5 Z=-0.79	1.6 -
1980	plac: 309						10.4	3.2

CDPR 1976	ASA: 758 plac: 771	62%≥55yrs 61%≥55yrs	100	74% ≥ 60 77% ≥ 60	972	1.8	5.8 Z=-1.90 8.3	5.4 Z=- 7.8
Elwood 1974	ASA: 615 plac: 624	57%≥55yrs 54%≥55yrs (55.0)	100	76% ≤ 3	300	1.0	7.6 not sig. 9.8	- -
Elwood 1979	ASA: 832 plac: 850	(56.0)	85	50% ≤ 0.25	900	1.0	12.8 not sig. 14.8 at *P*<0.05	- -
PARIS 1980	ASA: 810 plac: 406	30-74	87	2 - 60	972	3.4	10.5 Z=-1.21 12.8	8.0 Z=- 10.1

*ASA: aspirin (acetylsalicylic acid); plac: placebo

†% of mortality, cardiac death, and nonfatal myocardial infarction based on number of patients randomized.

‡Sum. stat.: summary statistic. *Z* is a statistical test that can be used to determine whether the difference in proportions or r treatment group and a control group are statistically significant. For a two-tailed test, *Z* values of ±1.96 and ±2.58 are appr values of 0.05 and 0.01.

Sources: Aspirin Myocardial Infarction Study Research Group 1980; Breddin et al. 1980; The Coronary Drug Project Rese and Sweetnam 1979; Elwood et al. 1974; Elwood 1983; The Persantine-Aspirin Reinfarction Study Research Group 1980.

Box 15

Distribution of Research Articles on PTCA

By Year of Publication and Method Used to Collect or Review Data

Article Class	1980	81	82	83	84	85	86	87	88	89	90	Total
Prospective RCT	0	0	0	0	1	1	2	4	2	1	2	13
Prospective non-RCT	0	0	1	3	4	5	5	6	11	8	3	46
Prospective registry	0	0	2	4	13	2	1	1	3	4	7	37
Case-control & adjusted cohort	0	0	1	2	0	0	2	2	4	5	2	18
Observational	1	1	1	3	12	12	12	27	25	29	8	131
Survey	0	0	0	0	0	0	0	0	1	0	1	2
Editorial	0	0	0	1	2	3	1	4	2	4	5	22
Review	0	0	0	2	3	4	4	5	16	14	11	59
Cross-sectional	0	0	0	0	0	0	0	2	1	0	0	3
Decision analysis	0	0	0	0	0	0	0	0	0	0	1	1
Total	1	1	5	15	35	27	27	51	65	65	40	332

Articles were retrieved using *MEDLINE* searches.

Source:Hilborne et al., 1991.

Box 16
Grading the Evidence

I: Evidence obtained from at least one properly-designed randomized controlled trial.
II-1: Evidence obtained from well designed controlled trials without randomization.
II-2: Evidence obtained from well designed cohort or case-controlled analytic studies, preferably from more than one center or research group.
II-3: Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s) could also be regarded as this type of evidence.
III: Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.
Source: U.S. Preventive Services Task Force 1996.

Box 17
Scientific Evidence for Pain Reduction in Adults

Pharmacologic Interventions		Type of Evidence		Strength and Consistency of Evidence
		Cancer Patients	Other Patients	
Acetaminophen and NSAIDS	Oral (alone) Oral (adjunct to opioid Rectal Parenteral (ketorolac)	I II I	I I II,III I	A A B A
Opioids	Oral Controlled release Rectal Transdermal Intramuscular Subcutaneous Intravenous PCA (IV and SC) Epidural and intrathecal Intracerebral ventricular	I I III,IV III,V II II II I II,III II,III,IV	I II II I I I I I,III	A A B B A A A A A A
Local anesthetics	Oral: Topical Systemic Intravenous Epidural and intrathecal Interpleural	II IV IV V	 II I II,III,V	B B C A B
Inhalent analgesia	Nitrous oxide	V	II,III	C

Adjuvant analgesics	Corticosteroids	II	II,III,IV	B
	Anticonvulsant agents	III	II,III,IV	B
	Antidepressants	II,III	I,II,III,IV	A
	Neuroleptics	II	II	C
	Hydroxyzine	III	II	C
	Psychostimulants	II,V	II	D
	Calcitonin	II,III,V		C
	Bisphosphonates	II,III		C

Type of evidence
I. Meta-analysis of multiple, well-designed controlled studies
II. At least one well-designed experimental study
III. Well-designed, quasiexperimental studies such as nonrandomized controlled, single group pre-post, cohort, time series, or matched case-controlled studies.
IV. Well-designed nonexperimental studies, such as comparative and correlational descriptive and case studies
V. Case reports and clinical examples

Strength and consistency of evidence
A. There is evidence of type I or consistent findings from multiple studies of types II, III, or IV.
B. There is evidence of types II, III, or IV, and findings are generally consistent.
C. There is evidence of types II, III, or IV, but findings are inconsistent.
D. There is little or no evidence, or there is type V evidence only.

Excerpted from: U.S. Department of Health and Human Services, Agency for Health Care Policy and Research 1994.

Box 18

Basic Checklist for Reviewing Reports of Randomized Controlled Trials

Did the trial:

Yes No

1. Specify outcome measures (endpoints) prior to the trial?
2. Provide patient inclusion/exclusion criteria?
3. Specify a-level for defining **statistical significance**?
4. Specify b-level (**power**) to detect a treatment effect of a given meaningful magnitude?
5. Make a prior estimate of required sample size (to satisfy levels of a and b)?
6. Use a proper method for random allocation of patients to treatment and control groups?
7. Use blinding (where possible):
   a. in the randomization process?
   b. for patients regarding their treatment?
   c. for observers/care givers regarding treatment?
   d. in collecting outcome data?
8. State the numbers of patients assigned to the respective treatment and control groups?
9. Clearly describe treatment and control (including placebo where applicable)?
10. Account for patient compliance with treatments/regimens?
11. Account for all events used as primary outcomes?

12. Account for patient withdrawals/losses to **follow-up**?
13. Analyze patient withdrawals/losses to follow-up
    a. by intention-to-treat?
    b. by treatment actually received?
14. Account for treatment complications/side effects?
15. Provide test statistics (e.g., *F, t, Z, chi-square*) and *P* values for endpoints?
16. Provide confidence intervals or confidence distributions?
17. Discuss whether power was sufficient for negative trials?
18. Interpret retrospective analyses (post hoc examination of subgroups and additional endpoints not identified prior to trial) appropriately?

Source: Goodman 1993.

Step Seven: Synthesize Evidence

Having considered the merits of individual studies, an assessment group must begin to synthesize or consolidate the available findings. For many topics in HCTA, there is no single definitive primary study, e.g., that settles whether one technology is better than another for a particular clinical situation. Even where definitive primary studies exist, findings from different types of studies must be combined or considered in broader social and economic contexts in order to formulate policies.

Methods used to combine or synthesize data include the following:

- non-quantitative **literature reviews**
- meta-analysis or other quantitative literature syntheses
- decision analysis
- group judgment or "consensus development"

The biases inherent in traditional means of consolidating literature, (i.e., non-quantitative literature reviews and editorials), are well recognized and greater emphasis is given to more structured, quantified and better documented methods. The body of knowledge concerning how to strengthen and apply these synthetic methods has grown substantially in recent years. Considerable work is being done to improve the validity of decision analysis and meta-analysis in particular (Eckman et al. 1992; Eddy 1992; Lau et al. 1992). Experience with the NIH Consensus Development Program, the panels on appropriateness of selected medical and surgical procedures conducted by the RAND Corporation, the clinical practice guidelines activities sponsored by AHCPR and others is adding to the body of knowledge concerning valid group judgment processes.

Three major types of synthetic studies--meta-analysis, decision analysis, and consensus development--are described below. Consensus development is also described in the next section, Step Eight: Formulate Findings and Recommendations.

Meta-analysis

Meta-analysis refers to a group of statistical techniques for combining results of multiple studies to obtain a quantitative estimate of the overall effect of a particular technology (or variable) on a defined outcome (Laird and Mosteller 1990; Thacker 1988). This combination may produce a stronger conclusion than can be provided by any individual study.

The purposes of meta-analysis are to:

- encourage systematic organization of evidence
- increase statistical power for primary end points and general applicability (external validity) of findings
- resolve uncertainty when reports disagree
- provide quantitative estimates of effects (e.g., odds ratios or effect sizes)
- provide insight into the nature of relationships among variables
- call attention to strengths and weaknesses of a body of research in a particular area
- identify needs for new primary data collection

Meta-analysis typically is used for topics that have no definitive studies, including topics for which nondefinitive studies are in some disagreement. Evidence collected for assessments often includes studies that do not have sufficient statistical power (e.g., because of small sample sizes) to detect any true treatment effects. By combining the results of multiple studies, a meta-analysis may have sufficient power to detect a true treatment effect if one exists (or at least narrow the confidence interval about the mean treatment effect).

Meta-analysis has been used for a variety of technologies. Recent examples are: adjuvant chemotherapy for colorectal cancer, surgery for otitis media in children, salt-restriction for high blood pressure, steroids for treatment of chronic obstructive pulmonary disease, treatments for duodenal ulcer and thrombolytic therapy. (See Box 19.) Although meta-analysis has been applied primarily for treatments, meta-analytic techniques are being extended to diagnostic technologies (Hasselblad and Hedges 1995; Irwig et al. 1994; Littenberg and Moses 1993).

Some of the particular techniques used in meta-analysis are: pooling, effect size, variance weighting, Mantel-Haenszel, Peto, DerSimonian and Laird, and **confidence profile method**. The suitability of any of these techniques for a group of studies depends upon the comparability of circumstances of investigation, type of outcome variables used, assumptions about the uniformity of treatment effects and other factors (Eddy 1992).

The basic steps in meta-analysis are the following:

- Specify the problem of interest.
- Specify the criteria for inclusion of studies (e.g., type and quality).
- Identify all studies that meet inclusion criteria.
- Classify study characteristics and findings according to, e.g.: study characteristics (patient types, practice setting, etc.), methodological characteristics (e.g., sample sizes, measurement process), primary results and type of derived summary statistics.
- Statistically combine study findings using common units (e.g., by averaging effect sizes); relate these to study characteristics; perform sensitivity analysis.
- Present results.

The different techniques of meta-analysis have specific rules about whether or not to include certain types of studies and how to combine their results. Some meta-analytic techniques adjust the results of the individual studies to try to account for differences in study design and related biases to their internal and external validity. Special computational tools (e.g., computer software) may be required to make the appropriate adjustments for the various types of biases in a systematic way (Detsky et al.

1992; Eddy et al. 1990).

Meta-analysis can be limited by biased selection of studies, poor quality data, insufficiently comparable studies selected (or available) for a meta-analysis and biased interpretation of findings. These shortcomings, which are shared by unstructured literature reviews and other less rigorous synthesis methods, can be minimized by the systematic approach of meta-analysis. Compared to the other methods of combining evidence, meta-analysis can be time-consuming and requires greater statistical and methodologic skills. However, meta-analysis is a much more explicit and accurate method.

Even though many assessments still tend to rely on overall subjective judgments and similar less rigorous approaches of integrating evidence, there is a clear trend toward learning about and using more meta-analytic approaches. An assessment group that uses the inclusion rules and other stipulations of meta-analysis is likely to conduct a more thorough and credible assessment, even if the group decides not to perform the final statistical consolidation of the results of pertinent studies. More ctice guidelines for specific health problems. For individual patients, decision models can be used to relate the likelihood of potential outcomes of alternative clinical strategies, and/or to identify the clinical strategy that has the greatest utility for a patient.

The basic steps of decision analysis are:

- Develop a model (e.g., a decision tree) that depicts the set of important choices (or decisions) and potential outcomes of these choices. For treatment choices, the outcomes may be health outcomes (health states); for diagnostic choices, the outcomes may be test results (e.g., positive or negative).
- Assign estimates (based on available literature) of the probabilities (or magnitudes) of each potential outcome given its antecedent choices.
- Assign estimates of the value of each outcome to reflect its utility or desirability (e.g., using a health-related quality of life measure, QALYs, etc.).
- Calculate the expected value of the outcomes associated with the particular choice(s) leading to those outcomes. This is typically done by multiplying the set of outcome probabilities by the value of each outcome.
- Identify the choice(s) associated with the greatest expected value. Based on the assumptions of the decision model, this is the most desirable choice, as it provides the highest expected value given the probability and value of its outcomes.
- Conduct a sensitivity analysis of the model to determine if plausible variations in the estimates of probabilities of outcomes or utilities change the relative desirability of the choices. (Sensitivity analysis is used because the estimates of key variables in the model may be based on limited data or simply expert conjecture.)

Box 20 shows an example of a decision tree for alternative therapies for managing recurrent angina following coronary artery bypass graft surgery. A limitation of modeling with decision trees is representing recurrent health states (e.g., recurrent complications of a chronic disease). An alternative approach is to use state-transition models that probabilities of moving from one state of health to another, including remaining in a given state or returning to it after intervening health states. One commonly used type of state-transition modeling is **Markov modeling.**

Consensus Development

In various forms, group judgment or **consensus development** is used to set standards, make

regulatory decisions, make payment policies, make technology acquisition decisions, formulate practice guidelines, define the state-of-the-art and other purposes. The term consensus development can refer to discrete group judgment processes or techniques that contribute to an assessment; or it can refer to particular consensus development programs (e.g., that of NIH) in which some group judgment event, typically a consensus development conference, is the principal means of assessment. In contrast to the quantitative synthesis methods of meta-analysis and decision analysis, group judgment is generally qualitative in nature. It may be unstructured and informal, or it may involve formal group methods such as the **nominal group technique** and **Delphi technique** (Fink et al. 1984; Gallagher et al. 1993; Jairath 1994). Although these processes typically involve face-to-face interaction, some group judgment efforts combine remote, iterative interaction of panelists (as in the formal Delphi technique) with face-to-face meetings. Computer conferencing and related telecommunications approaches also are used.

Virtually all HCTA efforts involve group judgment at some juncture, particularly to formulate findings and recommendations (Gruber 1993; Small et al. 1994). Group judgment also can be used for ranking, such as to set assessment priorities, and rating. For example, RAND uses a two-stage modified Delphi process (first stage, independent; second stage, panel meeting) in which expert panels rate the appropriateness of a procedure (e.g., tympanostomy tubes for children) (Kleinman et al. 1994) for each of many possible patient indications on a scale of 1.0 (extremely inappropriate) to 9.0 (extremely appropriate).

As implied by the discussion of grading the evidence and noted in Box 16, the opinion of an expert committee concerning, e.g., the safety or effectiveness of a particular intervention, does not in itself constitute strong evidence. Where the results of pertinent, rigorous scientific studies exist, these should take precedence. In the absence of strong evidence, and where practical guidance is needed, expert group opinion can be used to infer or extrapolate from the limited available evidence. Where many assessment efforts are deficient is not making clear where the evidence fades and where the expert group opinion begins (Institute of Medicine 1985).

Consensus development programs typically embrace most of the steps of HCTA described here. In these programs, the consensus development conference usually spans at least three of the HCTA steps: interpret evidence, synthesize evidence and formulate findings and recommendations.

Many current consensus development programs are derived from the model of consensus development originated at the U.S. NIH in 1977 as part of an effort to improve the translation of NIH biomedical research findings to clinical practice. Since then, NIH has held more than 100 consensus development conferences, and has modified and experimented with its process. Canada, Denmark, France, Israel, The Netherlands, Sweden and the U.K. are among the countries that have national-level consensus development programs. Although largely based on the NIH model, each has evolved to meet the needs of their respective sponsoring organizations and environments.

The variety in consensus development programs can be described and compared along several main types of characteristics:

- context of the process: e.g., intended audience, topics and impacts addressed, topic selection
- prepanel preparation: e.g., responsibility for planning, evidence preparation, prior drafting of questions and/or recommendations
- panel composition: e.g., panel size, selection, types of expertise/representation, characteristics of chairperson

- consensus conference attributes: e.g., length of conference, public involvement, private panel sessions, definition of consensus, decisionmaking procedures (such as voting), process for handling disagreement, format and dissemination of final product.

Among most programs, preparation for conferences takes approximately one year. Some programs prepare assessment questions and draft a consensus statement prior to the consensus conference; other programs do not. Most programs assemble evidence in advance and share this in some form with the panelists; in some instances, this involves providing structured, quantitative literature reviews with specific instructions to panelists weeks in advance. Programs usually provide for speakers to present the evidence during the consensus conference. Most programs select panels of 9-18 members, including clinicians, scientists and analysts, and lay people, with varying attention to balancing panels for members' known positions on the assessment issues. In most instances, the conference is held over a two-to-three day period, although others have multiple meetings over longer periods of time. Programs generally provide for part or all of the conference to be held in a public forum (McGlynn 1990).

In general, the advantages of consensus development processes are that they:

- provide a focus for assembling experts on an assessment topic
- provide a means for participation of lay people
- are relatively inexpensive and less time-consuming compared to new primary data collection
- provide a good way to call public attention to a HCTA
- increase exposure of participants and the public to relevant evidence
- can prompt face-to-face, evidence-based resolution of opposing viewpoints
- can apply expert judgment in areas where data are insufficient or inconclusive

In general, the disadvantages of consensus development processes are that they:

- do not generate new scientific evidence
- may appear to offer veracity to viewpoints that are not supported by evidence
- may over-emphasize or inhibit viewpoints depending upon the stature or personalities of the participants
- may be structured to force or give the appearance of group agreement when it does not exist
- are difficult to validate

Various reports have made recommendations concerning how to strengthen consensus development programs in particular or in general (Goodman and Baratz 1990; Institute of Medicine 1990; Olsen 1995) A synopsis of these recommendations is shown in Box 21.

Variations in Methodological Approach

There is no standard methodological approach for conducting all HCTAs. It is not always possible to conduct the most rigorous types of studies, and policies often must be made in the absence, or beforecompletion, of definitive studies. Given their varying assessment orientations, resource constraints and other factors, assessment programs tend to rely on different combinations of methods. Such variation notwithstanding, the general trend in HCTA is toward use of the more rigorous methods.

Box 19

Example of Meta-Analysis: Clinical Trials of Intravenous Streptokinase for Acute Myocardial Infarction

Lau et al. conducted two types of meta-analysis of 33 clinical trials of the effect on mortality of using the thrombolytic (i.e., to dissolve blood clots) drug streptokinase for treating myocardial infarction.

Text pending copyright clearance

Box 20

Example of Decision Analysis: Management of Angina After Coronary Artery Bypass Surgery

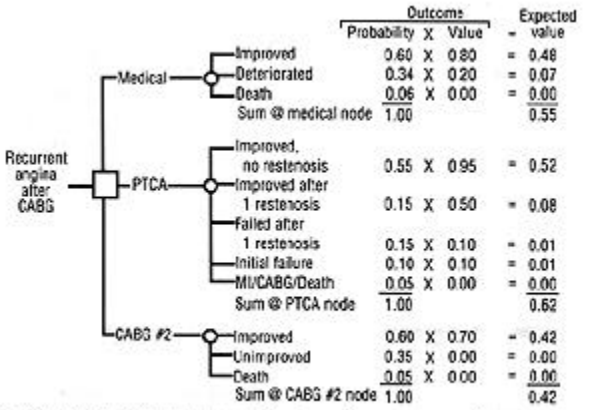

FIG. 2 Decision tree for management of angina after coronary artery bypass surgery.

The decision tree shown above depicts the outcomes for three alternatives for managing recurrent angina among patients who have had previous coronary artery bypass graft surgery (CABG). Each of the three main treatment choices--medical management (drugs and lifestyle changes), percutaneous transluminal coronary angioplasty (PTCA) and repeat CABG--results in a set of potential outcomes. For each outcome there is a probability and a value (ranging from 0.0 to 1.0). For example, given medical management, there is a 60 percent probability of improvement, which is valued at 0.80. According to this model, the most desirable treatment choice is PTCA, as the expected value of its outcomes is 0.62, exceeding that of medical management (0.55) and repeat CABG (0.42).

Source: Mills and Kalan 1991. Copyrighted and reprinted with the permission of Clinical Cardiology Publishing Company, Inc., and/or the Foundation for Advances in Medicine and Science, Inc., Mahwah, NJ 07430-0832, USA

Box 21

Recommendations for Improving Consensus Development Programs

1. Programs, or their sponsoring organizations, should have the ability to disseminate and/or implement their consensus findings and recommendations.
2. For each assessment, programs and/or panels should identify the intended audiences and means for achieving intended impact of consensus reports.
3. Programs should describe their scope of interest and/or responsibility, including their purposes, topics, and technological properties or impacts of concerns for the program in general and for specific assessments.
4. Programs should conduct assessments and provide reports in a timely fashion, including the

timeliness of assessments relative to the selected topics and the timeframe for planning, conducting, and reporting of assessments.

5. Programs should document the procedures and criteria for selecting conference topics and panel members.
6. The topic and scope of each assessment should be specific and manageable, i.e., commensurate with the available evidence, time, and other resources.
7. Panelists should represent the relevant health professions, methodologists such as epidemiologists and biostatisticians, economists, administrators, patient or other consumer representatives, and others who can provide relevant perspectives. Chairpersons should be recognized as objective with regard to consensus topics and skilled in group processes.
8. Programs should compile the available evidence concerning the assessment topics, and provide a systematic compilation or synthesis of this to panelists prior to the conference.
9. Programs should provide basic guidance concerning the interpretation of evidence, to help ensure that all panelists can be involved in this activity, regardless of their formal expertise in this area.
10. The consensus development processes should be structured and documented, including, e.g., advance identification of key questions/issues, operational definition of consensus, systematically organized evidence, opportunity for equitable participation of panelists, and duration and spacing of sessions to facilitate panelists' full and alert participation.
11. Consensus reports should include at least: description of the consensus process used, notations regarding the strength of agreement or assurance of the panel's findings, description of the reasoning used by the panel and the evidential basis for its findings, recommendations for research needed to address unresolved issues and otherwise advance understanding of the topic.
12. Programs should monitor new developments that might justify reassessments.
13. Programs should provide for periodic, independent evaluation of the program and its impacts.

Adapted from: Goodman and Baratz 1990.

Step Eight: Formulate Findings and Recommendations

The terms *findings* and *recommendations* often are used interchangeably; however, they are distinguished for purposes of certain HCTAs. Findings are the results or conclusions of an assessment; recommendations are the suggestions, advice, or counsel that follow from the findings. Recommendations can be made in various forms, such as a set of options, a practice guideline, or a directive. For example, an assessment finding may be that a new technology is acceptably safe and is more particular health condition, although certain patients are strongly averse to one of its side effects. A recommendation could be to offer the choice of the new technologyand the standard technology to patients with the health condition, and to have them decide how to proceed in consultation with their clinician.

Findings and recommendations should be linked to the quality of the evidence. The process of interpreting and synthesizing the evidence helps assessment groups to determine the adequacy of the evidence for addressing aspects of their assessment problems.

Even for those aspects of an assessment problem for which there is little useful evidence, an assessment group may have to provide some type of findings or recommendations. This may involve making inferences from the limited evidence, extrapolations of evidence from one circumstance to another, theory, or other subjective judgments. Whether a recommendation about using a technology in particular circumstances is positive, negative, or equivocal (neutral), users of the assessment should

understand the basis of that recommendation and with what level of confidence it was made. Unfortunately, the recommendations made in many assessment reports do not reflect the relative strength of the evidence upon which they are based. Readers may have the mistaken impression that all of the recommendations in an assessment report are equally valid or authoritative.

One approach to portraying the connection between evidence and recommendations uses different levels of strength of recommendations that are based upon grades of evidence. For instance, the U.S. Preventive Services Task Force assigned levels of strength to recommendations for interventions for breast cancer screening, based on the grades of evidence (from Box 16, as shown in Box 22).

A similar scheme for linking recommendation strength to the quality of evidence is shown in Box 23. This approach incorporates quantitative overviews of evidence (e.g., meta-analyses), where available, in classifying levels of evidence and grading recommendations.

Approaches such as these for grading evidence and recommendations concerning a particular intervention necessarily summarize findings pertaining to the multiple concerns of an assessment problem, e.g., safety (risks), effectiveness and cost effectiveness. The assessment report narrative should make explicit the analyses and reasoning that are used to derive overall recommendations about whether or under what circumstances a particular intervention should be used.

Assessment groups can render more credible judgments by explicitly linking the quality of the available evidence to the strength of their findings and recommendations (Hayward et al. 1995). Existing approaches for doing this are somewhat new and imperfect. Nevertheless, they are being improved and new ones are being developed. Using evidence this way enables readers to better understand the reasoning behind the assessment findings and recommendations. It also provides readers with a more substantive basis upon which to challenge the assessment as appropriate. Further, it helps assessment programs and policy makers to determine if a reassessment is needed as relevant new evidence becomes available.

Box 22

Level of Evidence and Strength of Recommendations for Breast Cancer Screening

This table shows a grading scheme for evidence concerning three types of breast cancer screening interventions for different age groups of women. For each age group, a grade is given to the quality of available evidence from primary studies for the interventions. The recommendations are derived from the level of evidence and the strength (or direction) of the findings. For instance, for women aged 40-49, the quality of the evidence concerning a routine mammogram every 1-2 years with or without an annual clinical breast exam was strong (Level I), though the strength of the recommendation (taking into account the findings of that evidence) was equivocal (Strength of Recommendation C).

Intervention	Age Group (years)	Level of Evidence	Strength of Recommendation

Routine mammogram every 1-2 yr with or without annual clinical breast exam	Women 40-49	I	C
	50-69	I, II-2	A
	70-74	I, II-3	C
	> 75	III	C
Annual clinical breast exam without periodic mammograms	Women 40-49	III	C
	50-59	I	C
	> 60	III	C
Routine breast self-exam		I, II-2, III	C

Level of Evidence:
I: Evidence obtained from at least one properly designed randomized controlled trial.
II-1: Evidence obtained from well designed controlled trials without randomization.
II-2: Evidence obtained from well designed cohort or case-control analytic studies, preferably from more than one center or research group.
II-3: Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s) could also be regarded as this type of evidence.
III: Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.

Strength of recommendations:
A. There is good evidence to support the recommendation that the condition [intervention] be specifically considered in a periodic health examination.
B. There is fair evidence to support the recommendation that the condition [intervention] be specifically considered in a periodic health examination.
C. There is poor evidence regarding the inclusion of the condition [intervention] in a periodic health examination, but recommendations may be made on other grounds.
D. There is fair evidence to support the recommendation that the condition [intervention] be excluded from consideration in a periodic health examination.
E. There is good evidence to support the recommendation that the condition [intervention] be excluded from consideration in a periodic health examination.

The strength of a recommendation depends upon the quality of the evidence regarding whether an intervention does more good than harm.

Adapted from: U.S. Preventive Services Task Force 1996.

Box 23
The Relation Between Levels of Evidence and Grades of Recommendation

If no overview available Level of Evidence	If high-quality overview[1] available Level of evidence	Grade of Recommendation
I: Randomized trials with low false-positive (a) and low false negative (b) errors	Lower limit of CI[2] for treatment effect *exceeds* clinically significant benefit and: I+: Individual study results homogeneous[3] I-: Individual study results heterogeneous[4]	A
II: Randomized trials with high false-positive (a) and high false negative (b) errors	Lower limit of CI for treatment effect *falls below* clinically significant benefit and: II+: Individual study results homogeneous II-: Individual study results heterogeneous	B
III: Nonrandomized concurrent cohort studies IV: Nonrandomized historical cohort studies V: Case series		C

[1] Scientific overview (e.g., meta-analysis) that includes all relevant studies of high quality (Level I or II) chosen in an explicit, unbiased fashion and that provides valid quantitative summaries of the data.

[2] Confidence interval

[3] Overview in which individual studies have consistent findings, e.g., all show beneficial treatment effect, all show no benefit, or all show harm.

[4] Overview in which individual studies have inconsistent findings, e.g., some show beneficial treatment effect, some show no benefit, and some show harm.

Source: Cook et al. 1992

Step Nine: Disseminate Findings and Recommendations

To the analysts and other experts who have participated in a HCTA, the importance of its findings and recommendations may be self-evident. Dissemination of these findings and recommendations, whether for internal use in the same organization or into the national health information mainstream, often is considered as an administrative afterthought.

Worthy HCTA messages get lost because of misidentified and misunderstood audiences, poor packaging, wrong transmission media, bad timing and other factors. Indeed, little is known about how to optimize the dissemination of HCTA findings and recommendations (Goldberg et al. 1994; Mittman and Siu 1992).

Competing for Attention

Dissemination efforts must compete with the burgeoning flow of health-related information being transmitted across diverse channels using increasingly sophisticated means. Advanced communications technologies provide alternative means to transmit more data where and when it can influence decision makers. Marketing, long practiced effectively by health care product companies, offers an evolving, continually researched variety of techniques that are being adapted throughout the health care sector. As the ground shifts in health care organization, delivery and financing, the cast of decision makers constituting the potential users of HCTA changes.

Dissemination Dimensions

Approaches for disseminating reports of HCTAs can be described along three dimensions: target groups (intended audiences), media and implementation techniques or strategies, as shown in Box 24.

The results of the same HCTA may be packaged for dissemination in different formats. As is done by AHCPR, printed versions of clinical practice guidelines can be distributed in booklet form for patients; in a quick reference guide or clinically-oriented handbook version for clinicians; and in a comprehensive report for researchers and policy analysts. Even reaching the same decision maker may require repeated messages and/or multiple media. The style in which an assessment report is written (e.g., an academic, scholarly tone versus a practical, concrete tone) may affect the receptiveness of researchers, practitioners and others (Kahan et al. 1988).

Dissemination Plan

Dissemination should be planned at the outset of an assessment along with other assessment phases or activities; and the costs, time and other resources needed for dissemination should be budgeted accordingly. This does not mean that dissemination plans should be rigid; the nature of the findings and recommendations themselves may affect the choice of target groups and the types of messages to be delivered. Dissemination should be designed to influence behavior of decision makers. This is not always straightforward, as research findings concerning what works for HCTA dissemination strategies do not point to any universally successful approaches.

Of course, the requirements for dissemination depend very much upon the mission or purpose of the organization that sponsored the assessment. Whereas reports of NIH consensus development conferences and AHCPR-sponsored practice guidelines typically are targeted nationally to clinicians, patients and others using multiple media, large third party payers would target their findings primarily to payment administrators using printed policy statements. A regulatory agency such as the FDA that reviews safety and efficacy information for new products supplied by pharmaceutical and health device companies has specific procedures regarding the timing and format of its findings.

Mediating Access

There are many approaches to controlling or enhancing access to assessment reports. Some assessment programs provide their assessments only to paid subscribers or member organizations, or charge fees intended to help recoup the cost of the assessment or provide a profit. (Some market research reports for emerging and new technologies are priced in the thousands of dollars.) While some assessments are made available in open libraries, others are held as proprietary (e.g., company

assessments of new products). Access to assessment literature is also mediated by the capacity of bibliographic organizations (e.g., the NLM and commercial database vendors) to index and abstract the literature, and the availability of such information via online databases and other information services. The wording used by assessment report authors for titles and abstracts can influence the indexing that serves as a key to accessing these reports.

Box 24
Approaches for HCTA Report Dissemination: Three Dimensions

Target groups

- clinicians (individuals, specialty/professional organizations);
- patients/consumers (individuals, organizations);
- provider organizations (hospitals, clinics, managed care organizations);
- third party payers (government, private sector);
- quality assurance and utilization review organizations;
- government policymakers (international, national, state, local);
- biomedical researchers;
- health care product companies;
- news professionals (popular and scientific/professional journalists and editors); and
- educational institutions (schools, continuing professional education programs).

Media

- printed: direct mail, newspapers and popular journals, scientific/professional journals and newsletters, posters, pocket cards;
- electronic: television, radio, video disks, computer databases (online and disk); and
- word of mouth: informal consultation, formal lectures and presentations, focus groups.

Implementation techniques or strategies:

- patient-oriented: mass media campaigns, community based campaigns, interaction with clinicians (including shared decision procedures, interactive video disk), modify insurance coverage (more or less generous benefits, change copayments);
- clinician-oriented: conferences and workshops; continuing professional education; professional curriculum development; opinion leaders; one-on-one educational visits ("academic detailing"); coverage/reimbursement policy; precertification; mandatory second opinion; drug formulary restrictions; feedback (e.g., on laboratory test ordering relative to criteria/guidelines); reminder systems (e.g., as part of computer-based patient record systems); medical audit/peer review; criteria for board certification/recertification, state licensure, Medicare PRO action, specialty designation, professional/specialty society membership; public availability of performance data (e.g., adjusted mortality rates for certain procedures); defense against sanctions and malpractice action; and
- institution-oriented: accreditation, standards (e.g., hospital infection control, clinical laboratories), benchmarking, public availability of performance data.

Step Ten: Monitor Impact

The impacts of HCTAs, from market research reports to RCT reports to expert panel statements, are

variable and inconsistently understood. Whereas some HCTA reports are translated directly into policies with clear and quantifiable impacts, the findings of some "definitive" RCTs and authoritative, well-documented assessment reports go unheeded or are not readily adopted into general practice (Chalmers 1974; Institute of Medicine 1985; U.S. Congress, OTA 1983).

As is the case for the technologies that are the subjects of HCTA, the reports of HCTAs can have intended, direct impacts as well as unintended, indirect ones. Some of the ways in which a HCTA report could make an impact (Banta and Luce 1993) are:

- affect corporate investment decisions
- modify R&D priorities/spending levels
- change regulatory policy
- modify marketing of a technology
- change third-party payment policy
- affect acquisition or adoption of a new technology
- change the rate of use of a technology
- change clinician behavior
- change patient behavior
- change the organization or delivery of care
- reallocate national or regional health care resources

Attributing Impact to HCTA Reports

The impact of a HCTA depends upon the target groups' legal, contractual, or administrative obligation to comply with it (Anderson et al. 1993; Ferguson et al. 1993; Gold et al. 1993). FDA market approvals of new drugs and devices are translated directly into binding policy. Most of the assessments conducted by the AHCPR OHTA are requested by HCFA for use in the Medicare program, although HCFA is not obligated to comply with OHTA findings. The impacts of NIH consensus development conference statements, which are not statements of government policy, are inconsistent and difficult to measure. The ability of NIH statements to change behavior seems to depend upon a variety of factors intrinsic to particular topics, the consensus development process and a multitude of contextual factors (Ferguson 1993).

The task of measuring the impact of HCTA can range from elementary to infeasible. Even if an intended change does occur, it may be difficult or impossible to attribute this change to the HCTA. A national-level assessment that recommends increased use of a particular intervention for a given clinical problem may be followed by a documented change in behavior consistent with that recommendation. However, the recommendation may be made at a time when the desired behavior change is already underway, when third-party payment policy is shifting in favor of the technology, during a strong marketing effort by industry, or close to the time of announcement of the results of a convincing clinical trial. Given widespread and nearly instant communications in health care, it may be difficult to control for factors other than a particular HCTA report that might influence behavior change.

As is the case for attributing changes in patient outcomes to a technological intervention, the ability to demonstrate that the results of a HCTA have an impact depends upon the conditions under which the assessment results were made known and the methodological approach used to determine the impact. Evaluations of the impact of an assessment often are unavoidably observational in nature; however, under some circumstances, quasi-experimental or experimental evaluations are used (Goldberg 1994).

To the extent that impact evaluations are prospective, involve pre- and post-dissemination data collection, and involve directed dissemination to clearly identified groups with well-matched controls (or at least retrospective adjustment for reported exposure to dissemination), they are more likely to detect a causal connection between a HCTA report and behavior change. Even so, generalizing from one experience to others may be impractical, as it is difficult to describe and replicate the conditions of a particular HCTA report dissemination.

Factors Mediating Impact

The factors that can affect the impact of HCTA reports are many. Beyond the particular dissemination techniques used, characteristics of the target groups, the environment and the HCTAs themselves can mediate the impact (Goldberg et al. 1994; Mittman and Siu 1992). Examples are shown in Box 25. Knowledge about these factors can be used prospectively. As noted above, assessment programs should consider how to properly target and modify their dissemination strategies to achieve the desired impact given particular characteristics of organizations, clinicians, environments, etc.

Systematic attempts to document the dissemination processes and impacts of HCTA programs are infrequent (Banta and Luce 1993; Goodman 1988; Institute of Medicine 1985), though a few, notably the NIH Consensus Development Program (Ferguson 1993), have been studied in detail. Like other interventions in health care, HCTA programs may be expected to demonstrate their own cost-effectiveness, (i.e., that the cost-savings and other benefits resulting from a HCTA program outweigh the cost of the program itself).

Box 25
Examples of Factors That Can Affect Impact of HCTA Reports

Target provider organization characteristics

- hospitals: general versus specialized, size, teaching status, patient mix, for-profit vs. non-profit, distribution of payment sources (e.g., fee-for-service vs. capitation), ownership status, financial status, accreditation
- physicians' offices: group practice vs. solo practice, hospital affiliation, teaching affiliation, board certification, distribution of payment sources

Target clinician characteristics

- type of clinician: physician, nurse, dentist, etc.
- specialty
- training
- professional activities/affiliations
- institutional affiliations (e.g., community hospital, university hospital)
- familiarity with and access to recent literature

Environmental characteristics

- urban, suburban, rural
- competitive environment
- economic status
- third-party payment status (e.g., percentage of patients in HMOs, private insurance, etc.)

- state and local laws, regulations
- malpractice activity

Characteristics of HCTA findings/recommendations

- type: research findings, practice guidelines, standards (e.g., equipment acquisition, use, maintenance), appropriateness criteria
- format: printed, word-of-mouth, electronic, etc.
- frequency of message
- required level of compliance (ranging from mandatory to optional)
- locus of decision: general practitioner/primary care physician only, physician specialist only, multiple clinicians, physician with patient input, patient only
- perceived inappropriate rigidity (allowance for discretion for differing circumstances)
- cost of relevant procedure/management of condition
- payment issue(s) addressed: coverage status, payment level
- reputation of sponsoring organization, analysts, expert panel
- overall strength of evidentiary base (e.g., existence of "definitive" clinical trial)
- credibility/rigor of assessment process
- existence or potential for malpractice action
- timeliness of dissemination, especially compared to degree of uncertainty, most recent research findings, or current levels/change rates of utilization of procedure
- existence and nature of other HCTA findings on same topic.

Sources: Goldberg et al. 1994; Mittman and Siu 1992; and others

APPENDIX A. CONCEPTS IN ASSESSMENT OF DIAGNOSTIC TECHNOLOGIES

The relationship between most preventive, therapeutic, and rehabilitative technologies and health outcomes is direct. The relationship between the use of diagnostic and screening technologies and health outcomes is typically indirect; these technologies provide information that may be used to inform providers concerning the use of interventions that may in turn affect health outcomes.

Many tests and other technologies used for diagnosis are also used for screening, and the concepts discussed here for diagnostic technologies pertain as well to screening technologies. A basic difference between screening and diagnosis is that diagnosis is done in symptomatic patients and screening is typically done in asymptomatic patient groups. For a given test used for either screening or diagnosis, this difference has a great effect on the probability that a patient has a disease or other health condition.

The immediate purpose of a diagnostic test is to provide information about the presence (and, less often, the extent) of a disease or other health condition. That is, the diagnostic test should be able to discriminate between patients who have a particular disease and those who do not have the disease (or discriminate among different extents of disease in a given patient).

The technical performance of a diagnostic test depends on a number of factors. Among these are the **precision** and **accuracy** of the test, the *observer variation* in reading the test data, and the

relationship between the disease of interest and the *cutoff level* of the *marker* or *surrogate* used in the diagnostic test to determine the presence or absence of that disease. These factors contribute to the ability of a diagnostic test to detect a disease when it is present and to not detect a disease when it is not present.

The marker for a disease or condition is typically defined as a certain cutoff level of a variable such as blood pressure or glucose level. Disease markers have distributions in nondiseased and diseased populations. For most diseases, these distributions overlap, so that a single cutoff level does not clearly separate nondiseased from diseased people. For instance, in the case of the disease of hypertension, a usual marker for the disease is diastolic blood pressure, the cutoff level of which is often set at 95mm Hg. In fact, some people whose diastolic blood pressure is above 95mm will not be hypertensive (false positives, as noted below), and some people with diastolic blood pressure below 95mm will be hypertensive (false negatives, as noted below). Lowering the cutoff to 90mm will decrease the number of false positives, but increase the number of false negatives.

A diagnostic test can have four basic types of outcomes, as shown in Box 26. A *true positive* diagnostic test result is one that detects a marker when the disease is present. A *true negative* test result is one that does not detect the marker when the disease is absent. A **false positive** test result is one that detects a marker when the disease is absent. A **false negative** test result is one that does not detect a marker when the disease is present.

Operating characteristics of diagnostic tests and procedures are measures of the technical performance of these technologies. These characteristics are based on the probabilities of the four possible types of outcomes of a diagnostic test. The two most commonly used operating characteristics of diagnostic tests are sensitivity and specificity. **Sensitivity** measures the ability of a test to detect disease when it is present. **Specificity** measures the ability of a test to correctly exclude disease in a nondiseased person. One economical way of depicting these operating characteristics for a given diagnostic test is with a **receiver operating characteristic (ROC) curve**, which plots the relationship between the true positive ratio (sensitivity) and false positive ratio (1 - specificity) as a function of the cutoff level of a disease (or condition) marker. ROC curves help to demonstrate how raising or lowering the cutoff point for defining a positive test result affects tradeoffs between correctly identifying people with a disease (true positives) and incorrectly labeling a person as positive who does not have the condition (false positives).

Taken alone, sensitivity and specificity do not reveal the probability that a given patient really has a disease if the test is positive, or the probability that a given patient does not have the disease if the test is negative. These probabilities are captured by two other operating characteristics. **Predictive value positive** is the proportion of those patients with a positive test result who actually have the disease. **Predictive value negative** is the proportion of patients with a negative test result who actually do not have the disease. (See Box 27.) Unlike sensitivity and specificity, predictive value positive and predictive value negative are not constant performance characteristics of a diagnostic test; they change with the prevalence of the disease in the population of interest. For example, if the disease is sufficiently rare in the population, even tests with high sensitivity and high specificity can have low predictive value positive, generating more false-positive than false negative results.

Beyond technical performance of diagnostic technologies, the effect of diagnostic technologies on health outcomes or health-related quality of life is less obvious than for other types of technologies. As health care decision makers increasingly demand to know how health care interventions affect health care outcomes, diagnostic technologies will have to demonstrate their efficacy/effectiveness

accordingly.

The efficacy (or effectiveness) of a diagnostic technology can be determined along a chain of inquiry that leads from technical capacity of a technology to changes in patient health outcomes to cost effectiveness, as follows.

- *Technical capacity*. Does the technology perform reliably and deliver accurate information?
- *Diagnostic accuracy*. Does the technology contribute to making an accurate diagnosis?
- *Diagnostic impact*. Do the diagnostic results influence use of other diagnostic technologies, e.g., does it replace other diagnostic technologies?
- *Therapeutic impact*. Do the diagnostic findings influence the selection and delivery of treatment?
- *Patient outcome*. Does use of the diagnostic technology contribute to improved health of the patient?
- *Cost effectiveness*. Does use of the diagnostic technology improve the cost effectiveness of health care compared to alternative interventions?

If a diagnostic technology is not efficacious at any step along this chain, then it is not likely to be efficacious at any later step. Efficacy at a given step does not imply efficacy at a later step (Feeny et al. 1986; Fineberg et al. 1977; Institute of Medicine 1985). Box 28 shows a hierarchy of studies for assessing diagnostic imaging technologies that is consistent with the chain of inquiry noted above.

For diagnostic technologies that are still prototypes or in other early stages of development, there are little data upon which to base answers to questions such as these. Even so, investigators and advocates of diagnostic technologies should be prepared to describe, at least qualitatively, the ways in which the technology might affect diagnostic accuracy, diagnostic impact, therapeutic impact, patient outcomes and cost effectiveness; how these effects might be measured; approximately what levels of performance would be needed to successfully implement the technology; and how further investigations should be conducted to make these determinations.

Box 26
Possible Outcomes of Diagnostic Tests

Test Result	Disease Status	
	Present	Absent
Positive	True positive	False positive
Negative	False negative	True negative

Box 27
Operating Characteristics of Diagnostic Tests

CHARACTERISTIC	FORMULA	DEFINITION
Sensitivity	True positives / True positives + False negatives	Proportion of people with condition who test positive
Specificity	True Negatives / True negatives + False positives	Proportion of people without condition who test negative
Predictive value positive	True positives / True positives + False positives	Proportion of people with positive test who have condition
Predictive value negative	True negatives / True negatives + False negatives	Proportion of people with negative test who do not have condition

Box 28

Hierarchical Model of Efficacy for Diagnostic Imaging:

Typical Measures of Analysis

Level 1. Technical efficacy
Resolution of line pairs
Modulation transfer function change
Gray-scale range
Amount of mottle
Sharpness

Level 2. Diagnostic accuracy efficacy
Yield of abnormal or normal diagnoses in a case series
Diagnostic accuracy (% correct diagnoses in case series)
Sensitivity and specificity in a defined clinical problem setting
Measures of area under the ROC curve

Level 3. Diagnostic thinking efficacy
Number (%) of cases in a series in which image judged "helpful" to making the diagnosis
Entropy change in differential diagnosis **probability distribution**
Difference in clinicians' subjectively estimated diagnosis probabilities pre- to post-test information
Empirical subjective log-likelihood ratio for test positive and negative in a case series

Level 4. Therapeutic efficacy
Number (%) of times image judged helpful in planning management of patient in a case series
% of times medical procedure avoided due to image information
Number (%) of times therapy planned before imaging changed after imaging information obtained (retrospectively inferred from clinical records)
Number (%) of times clinicians' prospectively stated therapeutic choices changed after test information obtained

Level 5. Patient outcome efficacy

% of patients improved with test compared with/without test
Morbidity (or procedures) avoided after having image information
Change in quality-adjusted life expectancy
Expected value of test information in quality-adjusted life years (QALYs)
Cost per QALY saved with imaging information
Patient utility assessment; e.g., Markov modeling; time trade-off

Level 6. Societal efficacy
Benefit-cost analysis from societal viewpoint
Cost-effectiveness analysis from societal viewpoint

Source: Thornbury and Fryback 1992.

APPENDIX B. SUGGESTED READINGS IN HCTA

Origins and Evolution of Technology Assessment

Arnstein, S. A. "Technology Assessment: Opportunities and Obstacles." IEEE Transactions on Systems, Man, and Cybernetics SMC-7 (1977): 571-82.
Brooks H., and R. Bowers. "The Assessment of Technology." Science 222, no. 2 (1970): 13-20.
Coates, J. F. "Technology Assessment: the Benefits ... the Costs ... the Consequences." The Futurist December (1971): 225-31.
National Academy of Engineering, Committee on Public Engineering Policy. A Study of Technology Assessment. Washington, DC: U.S. Government Printing Office, 1969.
National Research Council, Committee on the Life Sciences and Social Policy. Assessing Biomedical Technologies: An Inquiry into the Nature of the Process. Washington, DC: National Academy of Sciences, 1975.

Porter, A. L. "Technology Assessment." Impact Assessment 13, no. 2 (1995): 135-51.
Rettig, R. A. Health Care in Transition: Technology Assessment in the Private Sector. Santa Monica, Ca.: RAND, 1997.
U.S. Congress, Office of Technology Assessment. Technology Assessment in Business and Government. Summary and Analysis. Washington, DC: U.S. Government Printing Office, 1977.
U.S. Congress, Office of Technology Assessment. Assessing the Efficacy and Safety of Medical Technologies. Washington, DC: U.S. Government Printing Office, 1978.
U.S. Congress, Office of Technology Assessment. Development of Medical Technology: Opportunities for Assessment. Washington, DC: U.S. Government Printing Office, 1976.
U.S. Congress, House of Representatives. Committee on Science and Astronautics. Technology Assessment. Statement of Emilio Q. Daddario, Chairman, Subcommittee on Science Research and Development. 90th Cong., 1st sess., Washington, DC, 1967.

Overview of HCTA

Banta, H. D., and B. R. Luce. Health Care Technology and its Assessment: An International Perspective. New York: Oxford University Press, 1993.
Cochrane, A. L. Effectiveness and Efficiency: Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust, 1972.
Eddy, D. M. A Manual for Assessing Health Practices & Designing Practice Policies: The Explicit Approach. Philadelphia: American College of Physicians, 1992.
Feeny, D., G. Guyatt, and P. Tugwell, eds. Health Care Technology: Effectiveness, Efficiency, and Public Policy. Montreal: Institute for Research on Public Policy, 1986.
Franklin C. "Basic Concepts and Fundamental Issues in Technology Assessment." Intensive Care Medicine 19, no. 2 (1993): 117-21.
Frazier, H. S., and F. Mosteller, eds. Medicine Worth Paying For: Assessing Medical Innovations. Cambridge, Mass.: Harvard University Press, 1995.
Goodman, C. "It's Time to Rethink Health Care Technology Assessment." International Journal of Technology Assessment in Health Care. 8, no. 2 (1992): 335-58.
Hendee, W.R. "Technology Assessment in Medicine: Methods, Status and Trends." Medical Progress Through Technology 17, no. 2 (1991): 69-75.
Institute of Medicine. Assessing Medical Technologies. Washington, DC: National Academy Press, 1985.
"Joint Report of the Council on Scientific Affairs and the Council on Medical Service. Technology Assessment in Medicine." Archives of Internal Medicine 152, no. 1 (1992): 46-50.
Littenberg, B. "Technology Assessment in Medicine." Academic Medicine 67, no. 7 (1992): 424-8.
Luce, B. R., and A. Elixhauser, eds. Standards for Socioeconomic Evaluation of Health Care Products and Services. New York: Springer-Verlag, 1990.
Porter, A. L., F. A. Rossini, S. R. Carpenter, and A. T. Roper. A Guidebook for Technology Assessment and Impact Analysis. New York: North Holland, 1980.
Power, E. J., S. R. Tunis, and J. L. Wagner. "Technology Assessment and Public Health." Annual Review of Public Health 15, (1994): 561-79.
Szczepura, A., Kankaanpää, J., eds. Assessment of Health Care Technologies: Case Studies, Key Concepts and Strategic Issues. New York: John Wiley & Sons, 1996.
U.S. Congress, Office of Technology Assessment. Strategies for Medical Technology Assessment. Washington, DC: U.S. Government Printing Office, 1982.
U.S. Congress, Office of Technology Assessment. Identifying Health Technologies That Work: Searching for Evidence. OTA-H-608. Washington, DC: U.S. Government Printing Office, 1994.

U.S. Congress, Office of Technology Assessment. Health Care Technology Assessment and Its Assessment in Eight Countries. OTA-BP-H-140. Washington, DC: U.S. Government Printing Office, 1995.

Priority Setting

Donaldson, M. S., and H. C. Sox, Jr., eds. Setting Priorities for Health Technology Assessment: A Model Process. Washington, DC: National Academy Press, 1992.
Eddy, D. M. "Selecting Technologies for Assessment." International Journal of Technology Assessment in Health Care 5, no. 4 (1989): 485-501.
Hadorn, D. C. "The Problem of Discrimination in Health Care Priority Setting." Journal of the American Medical Association 268, no. 11 (1992): 1454-9.
Lara, M. E., and C. Goodman, eds. National Priorities for the Assessment of Clinical Conditions and Medical Technologies: Report of a Pilot Study. Washington, DC: National Academy Press, 1990.
Phelps, C. E., and S. T. Parente. "Priority Setting in Medical Technology and Medical Practice Assessment." Medical Care 28, no. 8 (1990): 703-23.
Phelps, C. E., and C. Mooney. "Correction and Update on 'Priority Setting in Medical Technology Assessment.'" Medical Care 30, no. 8 (1992): 744-51.

Randomized Controlled Trials

Buring, J. E., M. A. Jonas, and C. H. Hennekens. "Large and Simple Randomized Trials." In Tools for Evaluating Health Technologies: Five Background Papers, U.S. Congress, Office of Technology Assessment. BP-H-142. Washington, DC: U.S. Government Printing Office, 1995: 67-91.
Chalmers, T. C., P. Celano, H. S. Sacks, and H. Smith, Jr. "Bias in Treatment Assignment in Controlled Clinical Trials." New England Journal of Medicine 309, (1983): 1358-61.
Ellenberg, S. S. "Do Large, Simple Trials Have a Place in the Evaluation of AIDS Therapies?" Oncology 6, no. 4 (1992): 55-9, 63.
Freiman, J. A., T. C. Chalmers, H. Smith, Jr., and R. R. Kuebler. "The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Controlled Trial. Survey of 71 "Negative" Trials." New England Journal of Medicine 299, no. 13 (1978): 690-4.
Gurwitz, J. H., N. F. Col, and J. Avorn. "The Exclusion of the Elderly and Women From Clinical Trials in Acute Myocardial Infarction." Journal of the American Medical Association 268, no. 11 (1992): 1417-22.
Kraemer, H. C., and J. P. Pruyn. "The Evaluation of Different Approaches to Randomized Clinical Trials." Archives of General Psychiatry 47, no. 12 (1990): 1163-9.
Lubsen, J., and J. G. Tijssen. "Large Trials With Simple Protocols: Indications and Contraindications." Controlled Clinical Trials 10, suppl. 4 (1989): 151S-60S.
Merigan, T. C. "You Can Teach an Old Dog New Tricks: How AIDS Trials are Pioneering New Strategies." New England Journal of Medicine 323, no. 19 (1990): 1341-3.
Merkatz, R. B., R. Temple, S. Sobel, K. Feiden, and D. Kessler. "Women in Clinical Trials of New Drugs: A Change in Food and Drug Administration Policy." New England Journal of Medicine 329, no. 4 (1993): 292-6.
Peto, R., R. Collins, and R. Gray. "Large-Scale Randomized Evidence: Large, Simple Trials and Overviews of Trials." Journal of Clinical Epidemiology 48, no. 1 (1995): 23-40.
Pocock, S. J., M. D. Hughes, and R. J. Lee. "Statistical Problems in the Reporting of Clinical Trials."

New England Journal of Medicine 317, no. 7 (1987): 426-32.
Spilker, B. Guide to Clinical Trials. New York: Raven Press, 1991.
Yusuf, S., P. Held, K. K. Teo, and E. R. Toretsky. "Selection of Patients for Randomized Controlled Trials: Implications of Wide or Narrow Eligibility Criteria." Statistics in Medicine 9, no. 1-2 (1990): 73-83.

HCTA of Diagnostic and Screening Technologies

Begg, C.B. "Biases in the Assessment of Diagnostic Tests." Statistics in Medicine 6, no. 4 (1987): 411-23.
Boyko, E. J. "Ruling Out or Ruling In Disease With the Most Sensitive or Specific Diagnostic Test: Short Cut or Wrong Turn?" Medical Decision Making 14, no. 2 (1994): 175-9.
Canadian Task Force on the Periodic Health Examination. "The Periodic Health Examination: 2. 1987 Update." Canadian Medical Association Journal 138, no. 7 (1988): 617-26.
Eddy, D. M. "Screening for Cervical Cancer." Annals of Internal Medicine 113, no. 3 (1990): 214-26.
Goin, J. E., and G. A. Hermann. "The Clinical Efficacy of Diagnostic Imaging Evaluation Studies: Problems, Paradigms, and Prescription." Investigative Radiology 26, no. 5 (1991): 507-11.
Kassirer, J. P. "Our Stubborn Quest for Diagnostic Certainty. A Cause of Excessive Testing." New England Journal of Medicine 321, no. 18 (1989): 1272-3.
Littenberg, B., A. M. Garber, and H. C. Sox, Jr. "Screening for Hypertension." Annals of Internal Medicine 112, no. 3 (1990): 192-202.
Mushlin, A. I., C. Mooney, V. Grow, and C. E. Phelps. "The Value of Diagnostic Information to Patients With Suspected Multiple Sclerosis." Archives of Neurology 51, no. 1 (1994): 67-72.
Riegelman, R. K., and R. P. Hirsch, eds. Studying a Study and Testing a Test. How to Read the Medical Literature. Second Edition. Boston: Little, Brown, 1989.
Royal, H. D. "Technology Assessment: Scientific Challenges." AJR. American Journal of Roentgenology 163, no. 3 (1994): 503-7.
Showstack, J. A., S. A. Schroeder, and H. R. Steinberg. "Evaluating the Costs and Benefits of a Diagnostic Technology." Medical Care 9, no. 5 (1981): 498-509.
Sox, H., S. Stern, D. Owens, and H. L. Abrams. Assessment of Diagnostic Technology in Health Care: Rationale, Methods, Problems, and Directions. Institute of Medicine. Washington, DC: National Academy Press, 1989.
Thornbury, J. R. "Clinical Efficacy of Diagnostic Imaging: Love It or Leave It." AJR. American Journal of Roentgenology 162, no. 1 (1994): 1-8.
Thornbury, J. R., and D. G. Fryback. "Technology Assessment -- An American View." European Journal of Radiology 14, no. 2 (1992): 147-56.
U.S. Preventive Services Task Force. Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions. Baltimore: Williams & Wilkins, 1989.
Wells, P. N., J. A. Garrett, and P. C. Jackson. "Assessment Criteria for Diagnostic Imaging Technologies." Medical Progress Through Technology 17, no. 2 (1991): 93-101.

Evidence in HCTA

Bailar, J. C. and F. Mosteller, eds. Medical Uses of Statistics. Boston, MA: NEJM Books, 1992.
Canadian Task Force on the Periodic Health Examination. The Periodic Health Examination: 2. 1987 Update. Canadian Medical Association Journal 138, no. 7 (1988): 617-26.

Chalmers, I. "Under-Reporting Research is Scientific Misconduct." Journal of the American Medical Association 263, no. 10 (1990): 1405-8.
Chalmers, T. C., C. S. Frank, and D. Reitman. "Minimizing the Three Stages of Publication Bias." Journal of the American Medical Association 263, no. 10 (1990): 1392-5.
Chalmers, T. C., H. Smith, Jr., B. Blackburn, B. Silverman, B. Schroeder, D. Reitman, and A. Ambroz. "A Method for Assessing the Quality of a Randomized Are the Results of the Study Valid?" Journal of the American Medical Association 270, no. 21 (1993): 2598-601.
Guyatt, G. H., D. L. Sackett, and D. J. Cook. "Users' Guides to the Medical Literature. II. How to Use an Article About a Therapy or Prevention. B. What Were the Results and Will They Help Me in Caring for My Patients?" Journal of the American Medical Association 271, no. 1 (1994): 59-63.
Hirsh, J., and B. Haynes. "Transforming Evidence Into Practice: Evidence-based Consensus." Annals of Internal Medicine 118, suppl. 2 (1993): A16-7.
Jaeschke, R., G. Guyatt, and D. L. Sackett. "Users' Guides to the Medical Literature. III. How to Use an Article About a Diagnostic Test. A. Are the Results of the Study Valid?" Journal of the American Medical Association 271, no. 5 (1994): 389-91.
Jaeschke, R., G. H. Guyatt, and D. L. Sackett. "Users' Guides to the Medical Literature. III. How to Use an Article About a Diagnostic Test. B. What Are the Results and Will They Help Me in Caring for My Patients?" Journal of the American Medical Association 271, no. 9 (1994): 703-7.
Laupacis, A., C. D. Naylor, and D. L. Sackett. "How Should the Results of Clinical Trials be Presented to Clinicians?" Annals of Internal Medicine 116, suppl. 3 (1992): A12-4.
Laupacis, A., D. L. Sackett, and R. S. Roberts. "An Assessment of Clinically Useful Measures of the Consequences of Treatment." New England Journal of Medicine 318, no. 26 (1988): 1728-33.
Lomas, J., G. Anderson, M. Enkin, E. Vayda, R. Roberts, and B. MacKinnon. "The Role of Evidence in the Consensus Process: Results from a Canadian Experience." Journal of the American Medical Association 259, no. 20 (1988): 3001-5.
Naylor, C. D., E. Chen, B. Strauss. "Measured Enthusiasm: Does the Method of Reporting Trial Results Alter Perceptions of Therapeutic Effectiveness?" Annals of Internal Medicine 117, no. 11 (1992): 916-21.
Oxman, A. D., D. J. Cook, and G. H. Guyatt. "Users' Guides to the Medical Literature. VI. How to use an Overview. Evidence-Based Medicine Working Group." Journal of the American Medical Association 272, no. 17 (1994): 1367-71.
Oxman, A. D., D. L. Sackett, and G. H. Guyatt. "Users' Guides to the Medical Literature. I. How to Get Started." Journal of the American Medical Association 270, no. 17 (1993): 2093-5.
Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine. New York: Churchill Livingstone, 1997.
Temple, R. "Problems in the Use of Large Data Sets to Assess Effectiveness." International Journal of Technology Assessment in Health Care 6, no. 2 (1990): 211-9.

Cost-Effectiveness and Related Economic Analyses

Bloom, B. S. "Costs, Benefits and Unintended Gastrointestinal Side Effects of Pharmaceutical Therapy." Pharmacoeconomics 1, no. 3 (1992): 175-81.
Doubilet, P., M. C. Weinstein, and B. J. McNeil. "Use and Misuse of the Term "Cost Effective" in Medicine." New England Journal of Medicine 314, no. 4 (1986): 253-6.
Drummond, M. F. "Economic Evaluation and the Rational Diffusion and Use of Health Technology." Health Policy 7, no. 3 (1987): 309-24.
Drummond, M. F. "Allocating Resources." International Journal of Technology Assessment in Health Care 6, no. 1 (1990): 77-92.

Drummond, M. F., and L. Davies. "Economic Analysis Alongside Clinical Trials: Revisiting the Methodological Issues." International Journal of Technology Assessment in Health Care 7, no. 4 (1991): 561-73.
Eddy, D. M. "Clinical Decision Making: From Theory to Practice. Cost-effectiveness Analysis. Is It Up to the Task?" Journal of the American Medical Association 267, no. 24 (1992): 3342-8.
Eisenberg, J. M. "Clinical Economics: A Guide to the Economic Analysis of Clinical Practices." Journal of the American Medical Association 262, no. 20 (1989): 2879-86.
Gagnon, J. P. "A Primer on Pharmacoeconomics." Managed Care Medicine February (1995): 11-8.
Garber, A. M. "Can Technology Assessment Control Health Spending?" Health Affairs 13, no. 3 (1994): 115-26.
Geweke, J., and B. A. Weisbrod. "Clinical Evaluation vs. Economic Evaluation: The Case of a New Drug." Medical Care 20, no. 8 (1982): 821-30.
Laupacis, A., D. Feeny, A. S. Detsky, and P. X. Tugwell. "How Attractive Does a New Technology Have to be to Warrant Adoption and Utilization? Tentative guidelines for using clinical and economic evaluations." Canadian Medical Association Journal 146, no. 4 (1992): 473-81.
Maynard, A. "The Design of Future Cost-benefit Studies." American Heart Journal 119, no. 3 pt. 2 (1990): 761-5.
Pauker, S. G., J. P. Kassirer. "Therapeutic Decision Making: A Cost-Benefit Analysis." New England Journal of Medicine 293, no. 5 (1975): 229-34.
Poe, N. R., and R. I. Griffiths. "Clinical-Economic Trials." In Tools for Evaluating Health Technologies: Five Background Papers. U.S. Congress, Office of Technology Assessment, 125-49. BP-H-142. Washington, DC: U.S. Government Printing Office, 1995.
Russell, L. B. "Opportunity Costs in Modern Medicine." Health Affairs 11, no. 2 (1992): 162-9.
Warner, K. E., and B. R. Luce. Cost-Benefit and Cost-Effectiveness Analysis in Health Care. Ann Arbor, MI: Health Administration Press, 1982.
Weinstein, M. C. "Principles of Cost-effective Resource Allocation in Health Care Organizations." International Journal of Technology Assessment in Health Care 6, no. 1 (1990): 93-103.
Weisbrod, B. A. "A Guide to Benefit-Cost Analysis, As Seen Through a Controlled Experiment in Treating the Mentally Ill." Journal of Health Politics, Policy and Law 7, no. 4 (1983): 808-45.

Meta-Analysis

Chalmers, T. C. "Problems Induced by Meta-analysis." Statistics in Medicine 10, no. 6 (1991): 971-80.
Chalmers, T. C., and J. Lau. "Meta-analytic Stimulus for Changes in Clinical Trials." Statistical Methods in Medical Research 2, no. 2 (1993): 161-72.
Cook, D. J., G. H. Guyatt, G. Ryan, J. Clifton, L. Buckingman, A. Willan, W. McIlroy, and A. D. Oxman. "Should Unpublished Data be Included in Meta-Analyses? Current Convictions and Controversies." Journal of the American Medical Association 269, no. 21 (1993): 2749-53.
Cook, T. D., H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, and F. Mosteller. Meta-Analysis for Explanation: A Casebook. New York: Russell Sage Foundation, 1992.
Detsky, A. S., C. D. Naylor, K. O'Rourke, A. J. McGeer, and K. A. L'Abbé. "Incorporating Variations in the Quality of Individual Randomized Trials into Meta-analysis." Journal of Clinical Epidemiology 45, no. 3 (1992): 255-65.
Dickersin, K., and J. A. Berlin. "Meta-analysis: State-of-the-Science." Epidemiologic Reviews 14, (1992): 154-76.
Eddy, D. M. A Manual for Assessing Health Practices & Designing Practice Policies: The Explicit Approach. Philadelphia: American College of Physicians, 1992.

Eddy, D. M., V. Hasselblad, and R. Shachter. "A Bayesian Method for Synthesizing Evidence: The Confidence Profile Method." International Journal of Technology Assessment in Health Care 6, no. 1 (1990): 31-55.
Hasselblad, V., and D. C. McCrory. "Meta-Analytic Tools for Medical Decision Making: A Practical Guide." Medical Decision Making 15, no. 1 (1995): 81-96.
Hedges, L. V., and S. M. Rowley. "Combining Estimates Across Studies: Meta-analysis of Research." In Clinical Practice Guideline Development: Methodology Perspectives, edited by K. A. McCormick, S. R. Moore and R. A. Siegel, 15-25. Rockville, MD: Agency for Health Care Policy Research, 1994.
Jones, D. R. "Meta-analysis of Observational Epidemiological Studies: A Review." Journal of the Royal Society of Medicine 85, no. 3 (1992): 165-8.
L'Abbé, K. A., A. S. Detsky, and K. O'Rourke. "Meta-analysis in Clinical Research." Annals of Internal Medicine 107, no. 2 (1987): 224-33.
Laird, N. M., and F. Mosteller. "Some Statistical Methods for Combining Experimental Results." International Journal of Technology Assessment in Health Care 6, no. 1 (1990): 5-30.
Lau, J., E. M. Antman, J. Jiminez-Silva, B. Kupelnick, F. Mosteller, and T. C. Chalmers. "Cumulative Meta-analysis of Therapeutic Trials for Myocardial Infarction." New England Journal of Medicine 327, no. 4 (1992): 248-54.
Longnecker, M. P. "Meta-analysis." In Tools for Evaluating Health Technologies: Five Background Papers. U.S. Congress, Office of Technology Assessment, 93-123. BP-H-142. Washington, DC: U.S. Government Printing Office, 1995.
Petitti, D. B. Meta-analysis, Decision Analysis, and Cost-Effectiveness Analysis. Methods for Quantitative Synthesis in Medicine. New York: Oxford University Press, 1994.
Sacks, H. S., J. Berrier, D. Reitman, V. A. Ancona-Berk, and T. C. Chalmers. "Meta-analysis of Randomized Controlled Trials." New England Journal of Medicine 316, no. 8 (1987): 450-5.
Spector, T.D., and S. G. Thompson. "The Potential and Limitations of Meta-analysis." Journal of Epidemiology and Community Health 45, no. 2 (1991): 89-92.
Thacker, S. B. "Meta-analysis: A Quantitative Approach to Research Integration." Journal of the American Medical Association 259, no. 11 (1988): 1685-9.
U.S. General Accounting Office. Cross Design Synthesis: A New Strategy for Medical Effectiveness Research. Gaithersburg, MD, 1992.
Wilson, A., and D. A. Henry. "Meta-Analysis. Part 2: Assessing the Quality of Published Meta-Analyses." Medical Journal of Australia 156, no. 3 (1992): 173-87.

Decision Analysis

Eckman, M. H., H. J. Levine, and S. G. Pauker. "Decision Analytic and Cost-effectiveness Issues Concerning Anticoagulant Prophylaxis in Heart Disease." Chest 102, suppl. 4 (1992): 538-49S.
Fleming, C., J.H. Wasson, P. C. Albertsen, M. J. Barry, and J. E. Wennberg. "A Decision Analysis of Alternative Treatment Strategies for Clinically Localized Prostate Cancer." Journal of the American Medical Association 269, no. 20 (1993): 2650-8.
Hagen, M. D. "Decision Analysis: A Review." Family Medicine 24, no. 5 (1992): 349-54.
McNeil, B. J., and S. G. Pauker. "Decision Analysis for Public Health: Principles and Illustrations." Annual Review of Public Health 5, (1984): 135-61.
Pauker, S. G., and J. P. Kassirer. "Medical Progress: Decision Analysis." New England Journal of Medicine 316, no. 5 (1987): 250-8.
Thornton, J. G., R. J. Lilford, and N. Johnston. "Decision Analysis in Medicine." BMJ 304, no. 6834 (1992): 1099-103.
Zalkind, D. L., and R. H. Shachtman. "A Decision Analysis Approach to the Swine Influenza

Vaccination Decision for an Individual." Medical Care 18, no. 1 (1980): 59-72.

Consensus Development

Andreasen, P. B. "Consensus Conferences in Different Countries: Aims and Perspectives." International Journal of Technology Assessment in Health Care 4, no. 2 (1988): 305-8.
Eddy, D. M. "Clinical Policies and the Quality of Clinical Practice." New England Journal of Medicine 307, no. 6 (1982): 343-7.
Fink, A., J. Kosecoff, M. Chassin, and R. H. Brook. "Consensus Methods: Characteristics and Guidelines for Use." American Journal of Public Health 74, no. 9 (1984): 979-83.
Goodman, C., and S. R. Baratz, eds. Improving Consensus Development for Health Technology Assessment: An International Perspective. Washington, DC: National Academy Press, 1990.
Greer, A. L. "The Two Cultures of Biomedicine: Can There be a Consensus?" Journal of the American Medical Association 258, no. 19 (1987): 2739-40.
Institute of Medicine. Consensus Development at the NIH: Improving the Program. Washington, DC: National Academy Press, 1990.
Kosecoff, J., D. E. Kanouse, W. H. Rogers, L. McCloskey, C. M. Winslow, and R. H. Brook. "Effects of the National Institutes of Health Consensus Development Program on Physician Practice." Journal of the American Medical Association 258, no. 19 (1987): 2708-13.
Lomas, J., G. Anderson, M. Enkin, E. Vayda, R. Roberts, and B. MacKinnon. "The Role of Evidence in the Consensus Process: Results from a Canadian Experience." Journal of the American Medical Association 259, no. 20 (1988): 3001-5.
Perry, S. "The NIH Consensus Development Program a Decade Later." New England Journal of Medicine 317, no. 8 (1987): 485-488.
Veatch, R. M. "Consensus of Expertise: The Role of Consensus of Experts in Formulating Public Policy and Estimating Facts." Journal of Medicine and Philosophy 16, no. 4 (1991): 427-45.
Wortman, P. M., A. Vinokur, and L. Sechrest. "Do Consensus Conferences Work? A Process Evaluation of the NIH Consensus Development Program." Journal of Health Politics, Policy and Law 13, no. 3 (1988): 469-98.

Health-Related Quality of Life Measures

Aaronson, N. K. "Methodologic Issues in Assessing the Quality of Life of Cancer Patients." Cancer 67, suppl. 3 (1991): S844-50.
"Advances in Health Status Assessment: Proceedings of a Conference." Medical Care 30, suppl. 5 (1992): MS1-MS293.
"Advances in Health Status Assessment: Conference Proceedings." Medical Care 27, suppl. 3 (1989): S1-S294.
Bell, M. J., C. Bombardier, and P. Tugwell. "Measurement of Functional Status, Quality of Life, and Utility in Rheumatoid Arthritis." Arthritis and Rheumatism 33, no. 4 (1990): 591-601.
EuroQol© Group. "EuroQol© -- A New Facility for the Measurement of Health Related Quality of Life." Health Policy 16, no. 3 (1990): 199-208.
Fitzpatrick, R., A. Fletcher, S. Gore, D. Jones, D. Spiegelhalter, and D. Cox. "Quality of Life Measures in Health Care. I: Applications and Issues in Assessment." BMJ 305, no. 6861 (1992): 1074-7.
Fletcher, A., S. Gore, D. Jones, R. Fitzpatrick, D. Spiegelhalter, and D. Cox. "Quality of Life

Measures in Health Care. II: Design, Analysis, and Interpretation." BMJ 305, no. 6862 (1992): 1145-8.
Gill, T. M., and A. R. Feinstein. "A Critical Appraisal of the Quality of Quality-of-life Measurements." Journal of the American Medical Association 272, no. 8 (1994): 619-26.
Guyatt, G. H., D. H. Geeny, and D. L. Patrick. "Measuring Health-related Quality of Life." Annals of Internal Medicine 118, no. 8 (1993): 622-9.
Mason JM. Cost-per QALY league tables: their roles in pharmacoeconomic analysis. Pharmacoeconomics 1994;5:472-81.
Mehrez, A., and A. Gafni. "Healthy-years Equivalents Versus Quality-adjusted Life Years: In Pursuit of Progress." Medical Decision Making 13, no. 4 (1993): 287-92.
Mosteller, F., and J. Falotico-Taylor, eds. Quality of Life and Technology Assessment. Washington, DC: National Academy Press, 1989.
Nord, E. "Methods for Quality Adjustment of Life Years." Social Science and Medicine 34, no. 5 (1992): 559-69.
Patrick, D. L., and P. Erickson. Health Status and Health Policy. Quality of Life in Health Care Evaluation and Resource Allocation. New York: Oxford University Press, 1993.
"Proceedings of the International Conference on the Measurement of Quality of Life as an Outcome in Clinical Trials." Controlled Clinical Trials 12, suppl. 4 (1991): 79S-280S.
Revicki, D. A., M. Rothman, and B. Luce. "Health-Related Quality of Life Assessment and the Pharmaceutical Industry." Pharmacoeconomics 1, no. 6 (1992): 394-408.
Spiegelhalter, D., S. Gore, R. Fitzpatrick, A. E. Fletcher, D. R. Jones, and D. R. Cox. "Quality of Life Measures in Health Care. III: Resource Allocation." BMJ 305, no. 6863 (1992): 1205-9.
Walker, S.R., and R. M. Rosser, eds. Quality of Life Assessment: Key Issues for the 1990s. Dordrecht: Kluwer, 1993.
Ware, J. E., Jr. "The Status of Health Assessment 1994." Annual Review of Public Health 15, (1995): 327-54.
Williams, A. "Is the QALY a Technical Solution to a Political Problem? Of Course Not." International Journal of Health Services 21, no. 2 (1991): 365-9.

Ethical, Legal, Social, and Political Issues

Anderson, G. F., M. A. Hall, and E. P. Steinberg. "Medical Technology Assessment and Practice Guidelines: Their Day in Court." American Journal of Public Health 83, no. 11 (1993): 1635-9.
Banta, H. D., and P. B. Andreasen. "The Political Dimension in Health Care Technology Assessment Programs." International Journal of Technology Assessment in Health Care 6, no. 1 (1990): 115-23.
Baum, M., K. Zilka, and J. Houghton. "Ethics of Clinical Research: Lessons for the Future." BMJ 299, no. 6693 (1989): 251-3.
Beauchamp, T. L., and J. F. Childress. Principles of Biomedical Ethics. 4th ed. New York: Oxford University Press, 1993.
Diamond, G. A. "The Importance of Subjective Values to Medical Technology." Journal of Invasive Cardiology 5, no. 1 (1993): 17-22.
Ferguson, J. H., M. Dubinsky, and P. J. Kirsch. "Court-ordered Reimbursement for Unproven Medical Technology." Journal of the American Medical Association 269, no. 16 (1993): 2116-21.
Gillon, R., ed. Medical Technology, Social and Health Care Issues. New York: John Wiley, 1994.
Gold, J. A., M. J. Zaremski, E. R. Lev, and D. H. Shefrin. "Daubert v. Merrell Dow. The Supreme Court Tackles Scientific Evidence in the Courtroom." Journal of the American Medical Association 270, no. 24 (1993): 2964-7.
Lantos, J. "Ethics, Randomization, and Technology Assessment." Cancer 74, suppl. 9 (1994): 2653-6.

Marsden, C. "An Ethical Assessment of Intensive Care." International Journal of Technology Assessment in Health Care 8, no. 3 (1992): 408-18.
Passamani, E. "Clinical Trials: Are They Ethical?" New England Journal of Medicine 324, no. 22 (1991): 1589-92.
Sibbald, W. J., and K. J. Inman. "Problems in Assessing the Technology of Critical Care Medicine." International Journal of Technology Assessment in Health Care 8, no. 3 (1992): 419-43.
Tangri, S. S., and T. E. Levin. "New Medical Technologies: Economic, Ethical, and Political Considerations for Policy." Journal of Social Issues 49, no. 2 (1993): 1-211.

HCTA and Quality of Care

Brook, R. H. "Using Scientific Information to Improve Quality of Health Care." Annals of the New York Academy of Sciences 703, (1993): 74-84.
Donabedian, A. "Quality Assessment and Assurance: Unity of Purpose, Diversity of Means." Inquiry 25, no. 1 (1988): 173-92.
Donabedian, A. "The Quality of Care: How Can It be Assessed?" Journal of the American Medical Association 260, no. 12 (1988): 1743-8.
Lohr, K. N., ed. Medicare: A Strategy for Quality Assurance. Volume I. Institute of Medicine. Washington, DC: National Academy Press, 1990.
Lohr K. N., and R. A. Rettig, eds. Quality of Care and Technology Assessment. Report of a Forum of the Council on Health Care Technology. Washington, DC: National Academy Press, 1988.

Effectiveness/Outcomes Research and HCTA

Brook, R. H., and K. N. Lohr. "Efficacy, Effectiveness, Variations, and Quality: Boundary-crossing Research." Medical Care 23, no. 5 (1985): 710-22.
Ellwood, P. M. "Outcomes Management: A Technology of Patient Experience." New England Journal of Medicine 318, no. 23 (1988): 1549-56.
Fuchs, V. R., and A. M. Garber. "The New Technology Assessment." New England Journal of Medicine 323, no. 10 (1990): 673-7.
Neuhauser, D. "Ernest Amory Codman, M.D., and End Results of Medical Care." International Journal of Technology Assessment in Health Care 6, no. 2 (1990): 307-25.
Roper, W. L., W. Winkenwerder, G. M. Hackbarth, and H. Krakauer. "Effectiveness in Health Care: An Initiative to Evaluate and Improve Medical Practice." New England Journal of Medicine 319, no. 18 (1988): 1197-202.
Schwartz, J. S., and N. Lurie. "Assessment of Medical Outcomes: New Opportunities for Achieving a Long Sought-after Objective." International Journal of Technology Assessment in Health Care 6, no. 2 (1990): 333-9.
U.S. Congress, Office of Technology Assessment. Identifying Health Technologies That Work: Searching for Evidence. OTA-H-608. Washington, DC: U.S. Government Printing Office, 1994.

Searching the HCTA Literature

Antczak-Bouckoms, A., E. Burdick, S. Klawansky, and F. Mosteller. "Using Medical Registries and Data Sets for Technology Assessment." International Journal of Technology Assessment in Health

Care 7, no. 2 (1991): 123-8.
Auston, I., M. A. Cahn, and C. R. Selden. "Literature Search Methods for the Development of Clinical Practice Guidelines." In Clinical Practice Guideline Development: Methodology Perspectives, editied by K. A. McCormick, S. R. Moore, and R. A. Siegel, 123-7. Rockville MD: Agency for Health Care Policy Research, 1994.
Chalmers, I. The Cochrane Collaboration: Preparing, Maintaining, and Disseminating Systematic Reviews of the Effects of Health Care." Annals of the New York Academy of Sciences 703, (1993): 156-63.
Chalmers, T. C., P. Hewett, D. Reitman, and H. S. Sacks. "Selection and Evaluation of Empirical Research in Technology Assessment." International Journal of Technology Assessment in Health Care 5, no. 4 (1989): 521-36.
Dickersin, K., P. Hewitt, L. Mutch, I. Chalmers, and T. C. Chalmers. "Comparison of *MEDLINE* Searching With a Perinatal Trials Database." Controlled Clinical Trials 6 no. 4 (1985): 306-17.
Goodman, C., ed. Medical Technology Assessment Directory. Washington, DC: National Academy Press, 1988.
Goodman, C. Literature Searching and Evidence Interpretation for Assessing Health Care Practices. Stockholm: Swedish Council on Technology Assessment in Health Care, 1993.
Jadad, A. R., and H. J. McQuay. "A High-yield Strategy to Identify Randomized Controlled Trials for Systematic Reviews." Online Journal of Current Clinical Trials 1993. Doc. No. 33.
Larson, E., and R. Sattertwhite. "Searching the Literature: A Professional Imperative." American Journal of Infection Control 17, no. 6 (1989): 359-64.
Ludbrook, J. "How to Find the Surgical Articles You Are Looking For." Australian and New Zealand Journal of Surgery 61, no. 4 (1991): 251-3.
Marcaccio, K. Y., ed. Gale Directory of Databases. Volume 1: Online Databases. London: Gale Research International (Detroit, Washington, DC: Gale Research), 1993.
Marcaccio, K. Y., ed. Gale Directory of Databases. Volume 2: CD-ROM, Diskette, Magnetic Tape, Handheld, and Batch Access Database Products. London: Gale Research International (Detroit, Washington, DC: Gale Research), 1993.
Orenstein, R. M., ed. Fulltext Sources Online. Needham, MA: BiblioData, 1993.
U.S. Department of Health and Human Services, National Library of Medicine. The Basics of Searching MEDLINE. Bethesda, Md., 1989, PB89-146179.

GLOSSARY

Absolute risk reduction: a measure of treatment effect that compares the probability (or mean) of a type of outcome in the control group with that of a treatment group, [i.e.: $P_c - P_t$ (or $\mu_c - \mu_t$)]. For instance, if the results of a trial were that the probability of death in a control group was 25% and the probability of death in a treatment group was 10%, the absolute risk reduction would be (0.25 - 0.10) = 0.15. (See also **number needed to treat, odds ratio,** and **relative risk reduction.**)

Accuracy: the degree to which a measurement (e.g., the mean estimate of a treatment effect) is true or correct. An estimate can be accurate, yet not be precise, if it is based upon an unbiased method that provides observations having great variation (i.e., not close in magnitude to each other). (Contrast with **precision.**)

Alpha (a): the probability of a Type I (false-positive) error. In hypothesis testing, the a-level is the threshold for defining statistical significance. For instance, setting a at a level of 0.05 implies that investigators accept that there is a 5% chance of concluding incorrectly that an intervention is effective

when it has no true effect. The a-level is commonly set at 0.01 or 0.05 or 0.10.

Benchmarking: a quality assurance process in which an organization sets goals and measures its performance in comparison to those of the products, services, and practices of other organizations that are recognized as leaders.

Beta (b): the probability of a Type II (false-negative) error. In hypothesis testing, b is the probability of concluding incorrectly that an intervention is not effective when it has true effect. (1-b) is the **power** to detect an effect of an intervention if one truly exists.

Bias: in general, any factor that distorts the true nature of an event or observation. In clinical investigations, a bias is any systematic factor other than the intervention of interest that affects the magnitude of (i.e., tends to increase or decrease) an observed difference in the outcomes of a treatment group and a control group. Bias diminishes the accuracy (though not necessarily the precision) of an observation. Randomization is a technique used to decrease this form of bias. Bias also refers to a prejudiced or partial viewpoint that would affect someone's interpretation of a problem. Double blinding is a technique used to decrease this type of bias.

Bibliographic database: an indexed computer or printed source of citations of journal articles and other reports in the literature. Bibliographic citations typically include author, title, source, abstract, and/or related information (including full text in some cases). Examples are *MEDLINE* and *EMBASE*.

Blinding: the concealment of group assignment (to either the treatment or control group) from the knowledge of patients and/or investigators in a clinical trial. Blinding eliminates the possibility that knowledge of assignment may affect patient response to treatment or investigator behaviors that may affect outcomes. Blinding is not always practical (e.g. when comparing surgery to drug treatment), but it should be used whenever it is possible and compatible with optimal patient care. A **single-blind** trial is one in which knowledge of group assignment is withheld only from patients; a **double-blind** trial is one in which the knowledge is withheld from patients and investigators.

Case-control study: a retrospective observational study in which investigators identify a group of patients with a specified outcome (cases) and a group of patients without the specified outcome (controls). Investigators then compare the histories of the cases and the controls to determine the extent to which each was exposed to the intervention of interest.

Case study: an uncontrolled (prospective or retrospective) observational study involving an intervention and outcome in a single patient. (Also known as a single case report or anecdote.)

Citation: the record of an article, book, or other report in a bibliographic database that includes summary descriptive information, e.g., authors, title, abstract, source, and indexing terms.

Clinical pathway: a multidisciplinary set of daily prescriptions and outcome targets for managing the overall care of a specific type of patient, e.g., from pre-admission to post-discharge for patients receiving inpatient care. Clinical pathways often are intended to maintain or improve quality of care and decrease costs for patients in particular diagnosis-related groups.

Clinical practice guidelines: a systematically developed statement to assist practitioner and patient decisions about appropriate health care for one or more specific clinical circumstances. The development of clinical practice guidelines can be considered to be a particular type of HCTA; or, it

can be considered to be one of the types of policymaking that is informed or supported by HCTA.

Clinical significance: a conclusion that an intervention has an effect that is of practical meaning to patients and health care providers. Even though an intervention is found to have a statistically significant effect, this effect might not be clinically significant. In a trial with a large number of patients, a small difference between treatment and control groups may be statistically significant but clinically unimportant. In a trial with few patients, an important clinical difference may be observed that does not achieve statistical significance. (A larger trial may be needed to confirm that this is a statistically significant difference.)

Cohort study: an observational study in which outcomes in a group of patients that received an intervention are compared with outcomes in a similar group i.e., the cohort, either contemporary or historical, of patients that did not receive the intervention. In an adjusted- (or matched-) cohort study, investigators identify (or make statistical adjustments to provide) a cohort group that has characteristics (e.g., age, gender, disease severity) that are as similar as possible to the group that experienced the intervention.

Compliance: a measure of the extent to which patients undergo an assigned treatment or regimen, e.g., taking drugs, undergoing a medical or surgical procedure, doing an exercise regimen, or abstaining from smoking.

Concurrent nonrandomized control: a control group that is observed by investigators at the same time as the treatment group, but that was not established using random assignment of patients to control and treatment groups. Differences in the composition of the treatment and control groups may result.

Confidence interval: depicts the range of uncertainty about an estimate of a treatment effect. It is calculated from the observed differences in outcomes of the treatment and control groups and the sample size of a study. The confidence interval is the range of values above and below the point estimate that is likely to include the true value of the treatment effect. The use of confidence intervals assumes that a study provides one sample of observations out of many possible samples that would be derived if the study were repeated many times. Investigators typically use confidence intervals of 90%, 95%, or 99%. For instance, there is a 95% probability that a 95% confidence interval calculated from a particular study includes the true value of a treatment effect. If the interval includes a null treatment effect (usually 0.0 but 1.0 if the treatment effect measure used is an odds ratio or relative risk), the null hypothesis of no true treatment effect cannot be rejected.

Confidence profile method: a type of meta-analysis based on Bayesian statistics for combining results of multiple studies of various design (e.g., RCTs, observational studies, and others) that adjusts the individual studies for their respective methodological biases before combining their results into a probability distribution for the parameter(s) of interst.

Consensus development: various forms of group judgment in which a group (or panel) of experts interacts in assessing an intervention and formulating findings by vote or other process of reaching general agreement. These process may be informal or formal, involving such techniques as the nominal group and Delphi techniques.

Contraindication: a clinical symptom or circumstance indicating that the use of an otherwise advisable intervention would be inappropriate.

Control group: a group of patients that serves as the basis of comparison when assessing the effects of the intervention of interest that is given to the patients in the treatment group. Depending upon the circumstances of the trial, a control group may receive no treatment, a "usual" or "standard" treatment, or a placebo. To make the comparison valid, the composition of the control group should resemble that of the treatment group as closely as possible. (See also **historical control** and **concurrent nonrandomized control**.)

Controlled vocabulary: a system of terms, involving, e.g., definitions, hierarchical structure, and cross-references, that is used to index and retrieve a body of literature in a bibliographic, factual, or other database. An example is the *MeSH* controlled vocabulary used in *MEDLINE* and other *MEDLARS* databases of the NLM.

Cost-benefit analysis: a comparison of alternative interventions in which costs and outcomes are quantified in common monetary units.

Cost-effectiveness analysis: a comparison of alternative interventions in which costs are measured in monetary units and outcomes are measured in non-monetary units, e.g., reduced mortality or morbidity.

Cost-minimization analysis: a determination of the least costly among alternative interventions that are assumed to produce equivalent outcomes.

Cost-utility analysis: a form of cost-effectiveness analysis of alternative interventions in which costs are measured in monetary units and outcomes are measured in terms of their utility, usually to the patient, e.g., using QALYs.

Cost of illness analysis: a determination of the economic impact of an disease or health condition, including treatment costs; this form of study does not address benefits/outcomes.

Crossover bias: occurs when some patients who are assigned to the treatment group in a clinical study do not receive the intervention or receive another intervention, or when some patients in the control group receive the intervention (e.g., outside the trial). If these crossover patients are analyzed with their original groups, this type of bias can "dilute" (diminish) the observed treatment effect.

Crossover design: a clinical trial design in which patients receive, in sequence, the treatment (or the control), and then, after a specified time, switch to the control (or treatment). In this design, patients serve as their own controls, and randomization is used to determine the order in which a patient receives the treatment and control.

Cross-sectional study: a (prospective or retrospective) observational study in which a group is chosen (sometimes as a random sample) from a certain larger population, and the exposures of people in the group to an intervention and outcomes of interest are determined.

Database (or register): any of a wide variety of repositories (often computerized) for observations and related information about a group of patients (e.g., adult males living in Göteborg) or a disease (e.g., hypertension) or an intervention (e.g., antihypertensive drug therapy) or other events or characteristics. Depending upon criteria for inclusion in the database, the observations may have controls. Although these can be useful, a variety of confounding factors (e.g., no randomization and

possible selection bias in the process by which patients or events are recorded) make them relatively weak methods for determining causal relationships between an intervention and an outcome.

Decision analysis: an approach to decision making under conditions of uncertainty that involves modeling of the sequences or pathways of multiple possible strategies (e.g., of diagnosis and treatment for a particular clinical problem) to determine which is optimal. It is based upon available estimates (drawn from the literature or from experts) of the probabilities that certain events and outcomes will occur and the values of the outcomes that would result from each strategy. A decision tree is a graphical representation of the alternate pathways.

Delphi technique: an iterative group judgment technique in which a central source forwards surveys or questionnaires to isolated, anonymous (to each other) participants whose responses are collated/summarized and recirculated to the participants in multiple rounds for further modification/critique, producing a final group response (sometimes statistical).

Direct costs: the fixed and variable costs of all resources (goods, services, etc.) consumed in the provision of an intervention as well as any consequences of the intervention such as adverse effects or goods or services induced by the intervention. Includes direct medical costs and direct nonmedical costs such as transportation or child care.

Disability-adjusted life years (DALYs): a unit of health care status that adjusts age-specific life expectancy by the loss of health and years of life due to disability from disease or injury. DALYs are often used to measure the global burden of disease.

Discounting: the process used in cost analyses to reduce mathematically future costs and/or benefits/outcomes to their present value, e.g., at an annual rate of five or ten percent. These adjustments reflect that given levels of costs and benefits occurring in the future usually have less value in the present than the same levels of costs and benefits realized in the present.

Disease management: a systematic process of managing care of patients with specific diseases or conditions (particularly chronic conditions) across the spectrum of outpatient, inpatient, and ancillary services. The purposes of disease management may include: reduce acute episodes, reduce hospitalizations, reduce variations in care, improve health outcomes, and reduce costs. Disease management may involve continuous quality improvement or other management paradigms. It may involve a cyclical process of following practice protocols, measuring the resulting outcomes, feeding those results back to clinicians, and revising protocols as appropriate.

Dissemination: any process by which information is transmitted (made available or accessible) to intended audiences or target groups.

Effect size: same as **treatment effect**. Also, a dimensionless measure of treatment effect that is typically used for continuous variables and is usually defined as the difference in mean outcomes of the treatment and control group divided by the standard deviation of the outcomes of the control group. One type of meta-analysis involves averaging the effect sizes from multiple studies.

Effectiveness: the benefit (e.g., to health outcomes) of using a technology for a particular problem under general or routine conditions, for example, by a physician in a community hospital or by a patient at home.

Effectiveness research: see **outcomes research.**

Efficacy: the benefit of using a technology for a particular problem under ideal conditions, for example, in a laboratory setting, within the protocol of a carefully managed randomized controlled trial, or at a "center of excellence."

Endpoint: a measure or indicator chosen for determining an effect of an intervention.

Evidence-based medicine: the use of current best evidence from scientific and medical research to make decisions about the care of individual patients. It involves formulating questions relevant to the care of particular patients, searching the scientific and medical literature, identifying and evaluating relevant research results, and applying the findings to patients.

Evidence table: a summary display of selected characteristics (e.g., of methodological design, patients, outcomes) of studies of a particular intervention or health problem.

External validity: the extent to which the findings obtained from an investigation conducted under particular circumstances can be generalized to other circumstances. To the extent that the circumstances of a particular investigation (e.g., patient characteristics or the manner of delivering a treatment) differ from the circumstances of interest, the external validity of the findings of that investigation may be questioned.

Factual database: an indexed computer or printed source that provides information in the form of guidelines for diagnosis and treatment, patient indications, or other authoritative information. Examples are *PDQ,* a computer database on cancer management, and *DRUGLINE,* a computer database on drug indications, contraindications, and interactions.

False negative error: occurs when the statistical analysis of a trial detects no difference in outcomes between a treatment group and a control group when in fact a true difference exists. This is also known as a **Type II error**. The probability of making a Type II error is known as b.

False positive error: occurs when the statistical analysis of a trial detects a difference in outcomes between a treatment group and a control group when in fact there is no difference. This is also known as a **Type I error**. The probability of a Type I error is known as a.

Follow-up: the ability of investigators to observe and collect data on all patients who were enrolled in a trial for its full duration. To the extent that data on patient events relevant to the trial are lost, e.g., among patients who move away or otherwise withdraw from the trial, the results may be affected, especially if there are systematic reasons why certain types of patients withdraw. Investigators should report on the number and type of patients who could not be evaluated, so that the possibility of bias may be considered.

Gray literature: research reports that are not found in traditional peer-reviewed publications, for example: government agency monographs, symposium proceedings, and unpublished company reports.

Health care technology assessment (HCTA): the systematic evaluation of properties, effects, and/or impacts of health care technology. It may address the direct, intended consequences of technologies as

well as their indirect, unintended consequences. Its main purpose is to inform technology-related policymaking in health care. HCTA is conducted by interdisciplinary groups using explicit analytical frameworks drawing from a variety of methods.

Health-related quality of life (HRQL) measures: patient outcome measures that extend beyond traditional measures of mortality and morbidity, to include such dimensions as physiology, function, social activity, cognition, emotion, sleep and rest, energy and vitality, health perception, and general life satisfaction. (Some of these are also known as health status, functional status, or quality of life measures.)

Health services research: a field of inquiry that examines the impact of the organization, financing and management of health care services on the delivery, quality, cost, access to and outcomes of such services.

Healthy-years equivalents (HYEs): the number of years of perfect health that are considered equivalent to (i.e., have the same utility as) the remaining years of life in their respective health states.

Historical control: a control group that is chosen from a group of patients who were observed at some previous time. The use of historical controls raises concerns about valid comparisons because they are likely to differ from the current treatment group in their composition, diagnosis, disease severity, determination of outcomes, and/or other important ways that would confound the treatment effect. It may be feasible to use historical controls in special instances where the outcomes of a standard treatment (or no treatment) are well known and vary little for a given patient population.

Hypothesis testing: a means of interpreting the results of a clinical trial that involves determining the probability that an observed treatment effect could have occurred due to chance alone if a specified hypothesis were true. The specified hypothesis is normally a **null hypothesis,** made prior to the trial, that the intervention of interest has no true effect. Hypothesis testing is used to determine if the null hypothesis can or cannot be rejected.

Incidence: the rate of occurrence of new cases of a disease or condition in a population at risk during a given period of time, usually one year.

Indication: a clinical symptom or circumstance indicating that the use of a particular intervention would be appropriate.

Indirect costs: the cost of time lost from work and decreased productivity due to disease, disability, or death. (In cost accounting, it refers to the the overhead or fixed costs of producing goods or services.)

Intangible costs: the cost of pain and suffering resulting from a disease, condition, or intervention.

Internal validity: the extent to which the findings of a study accurately represent the causal relationship between an intervention and an outcome in the particular circumstances of that study. The internal validity of a trial can be suspect when certain types of biases in the design or conduct of a trial could have affected outcomes, thereby obscuring the true direction, magnitude, or certainty of the treatment effect.

Investigational Device Exemption (IDE): a regulatory category and process in which the U.S. Food

and Drug Administration (FDA) allows specified use of an unapproved health device in controlled settings for purposes of collecting data on safety and efficacy/effectiveness; this information may be used subsequently in a premarketing approval application.

Investigational New Drug Application (IND): an application submitted by a sponsor to the U.S. FDA prior to human testing of an unapproved drug or of a previously approved drug for an unapproved use.

Large, simple trials: prospective, randomized controlled trials that use large numbers of patients, broad patient inclusion criteria, multiple study sites, minimal data requirements, and electronic registries; their purposes include detecting small and moderate treatment effects, gaining effectiveness data, and improving external validity.

Literature review: a summary and interpretation of research findings reported in the literature. May include unstructured qualitative reviews by single authors as well as various systematic and quantitative procedures such as meta-analysis. (Also known as overview.)

Marginal benefit: the additional benefit (e.g., in units of health outcome) produced by an additional resource use (e.g., another health care intervention).

Marginal cost: the additional cost required to produce an additional unit of benefit (e.g., unit of health outcome).

Markov model: A type of quantitative modeling that involves a specified set of mutually exclusive and exhaustive states (e.g., of a given health status), and for which there are transition probabilities of moving from one state to another (including of remaining in the same state). Typically, states have a uniform time period, and transition probabilities remain constant over time.

MEDLARS: the *Med*ical *L*iterature *A*nalysis and *R*etrieval *S*ystem of about 40 computer databases managed by the U.S. NLM.

MEDLINE: a bibliographic database that is the most used of about 40 *MEDLARS* databases managed by the U.S. NLM. It is the computer version of the printed *Index Medicus.* Citations for 7.5 million articles published since 1966 from about 3,700 health and biomedical journals are compiled in *MEDLINE,* which is updated at a rate of 6,600 articles every week. About 75% of citations are for English-language articles.

MeSH: *Me*dical *S*ubject *H*eadings, the controlled vocabulary of about 16,000 terms used for *MEDLINE* and certain other *MEDLARS* databases.

Meta-analysis: systematic methods that use statistical techniques for combining results from different studies to obtain a quantitative estimate of the overall effect of a particular intervention or variable on a defined outcome. This combination may produce a stronger conclusion than can be provided by any individual study. (Also known as data synthesis or quantitative overview.)

Moving target problem: changes in health care that can render the findings of HCTAs out of date, sometimes before their results can be implemented. Included are changes in the focal technology, changes in the alternative or complementary technologies i.e., that are used for managing a given health problem, emergence of new competing technologies, and changes in the application of the

technology (e.g., to different patient populations or to different health problems).

N of 1 trial: a clinical trial in which a single patient is the total population for the trial, including a single case study. An N of 1 trial in which random allocation is used to determine the order in which an experimental and a control intervention are given to a patient is an N of 1 RCT.

New Drug Application (NDA): an application submitted by a sponsor to the FDA for approval to market a new drug (a new, nonbiological molecular entity) for human use in U.S. interstate commerce.

Nominal group technique: a face-to-face group judgment technique in which participants generate silently, in writing, responses to a given question/problem; responses are collected and posted, but not identified by author, for all to see; responses are openly clarified, often in a round-robin format; further iterations may follow; and a final set of responses is established by voting/ranking.

Null hypothesis: in hypothesis testing, the hypothesis that an intervention has no effect, i.e., that there is no true difference in outcomes between a treatment group and a control group. Typically, if statistical tests indicate that the *P* value is at or above the specified a-level (e.g., 0.01 or 0.05), then any observed treatment effect is not statistically significant, and the null hypothesis cannot be rejected. If the P value is less than the specified a-level, then the treatment effect is statistically significant, and the null hypothesis is rejected. If a confidence interval (e.g., of 95% or 99%) includes zero treatment effect, then the null hypothesis cannot be rejected.

Number needed to treat: a measure of treatment effect that provides the number of patients who need to be treated to prevent one outcome event. It is the inverse of absolute risk reduction (1 ¸ absolute risk reduction); i.e., $1.0 \div (P_c - P_t)$. For instance, if the results of a trial were that the probability of death in a control group was 25% and the probability of death in a treatment group was 10%, the number needed to treat would be $1.0 \div (0.25 - 0.10) = 6.7$ patients. (See also **absolute risk reduction, relative risk reduction,** and **odds ratio.**)

Observational study: a study in which the investigators do not manipulate the use of an intervention (e.g., do not randomize patients to treatment and control groups), but only observe patients who are (and sometimes patients who are not) exposed to the intervention, and interpret the outcomes.

Odds ratio: a measure of treatment effect that compares the probability of a type of outcome in the treatment group with the outcome of a control group, i.e., $[P_t \div (1 - P_t)] \div [P_c \div (1 - P_c)]$. For instance, if the results of a trial were that the probability of death in a control group was 25% and the probability of death in a treatment group was 10%, the odds ratio of survival would be $[0.10 \div (1.0 - 0.10)] \div [(0.25 \div (1.0 - 0.25)] = 0.33$. (See also **absolute risk reduction, number needed to treat,** and **relative risk.**)

Outcomes research: evaluates the impact of health care on the health outcomes of patients and populations. It may also include evaluation of economic impacts linked to health outcomes, such as cost effectiveness and cost utility. Outcomes research emphasizes health problem- (or disease-) oriented evaluations of care delivered in general, real-world settings; multidisciplinary teams; and a wide range of outcomes, including mortality, morbidity, functional status, mental well-being, and other aspects of health-related quality of life. It may entail any in a range of primary data collection methods and synthesis methods that combine data from primary studies.

***P* value:** in hypothesis testing, the probability that an observed difference between the intervention and control groups is due to chance alone if the null hypothesis is true. If *P* is less than the a-level (typically 0.01 or 0.05) chosen prior to the study, then the null hypothesis is rejected.

Parallel (or independent groups) design: a trial that compares two groups of patients, one of which receives the treatment of interest and one of which is a control group (e.g., a randomized controlled trial). (Some parallel trials have more than one treatment group; others compare two treatment groups, each acting as a control for the other.)

Patient selection bias: a bias that occurs when patients assigned to the treatment group differ from patients assigned to the control group in ways that can affect outcomes, e.g., age or disease severity. If the two groups are constituted differently, it is difficult to attribute observed differences in their outcomes to the intervention alone. Random assignment of patients to the treatment and control groups minimizes opportunities for this bias.

Peer review: the process by which manuscripts submitted to health, biomedical, and other scientifically oriented journals and other publications are evaluated by experts in appropriate fields (usually anonymous to the authors) to determine if the manuscripts are of adequate quality for publication.

Phase I, II, III, and IV studies: phases of clinical trials of new technologies (usually drugs) in the development and approval process required by the FDA (or other regulatory agencies). Phase I trials usually involve approximately 20-80 healthy volunteers to determine a drug's safety, safe dosage range, absorption, metabolic activity, excretion, and the duration of activity. Phase II trials are controlled trials in approximately 100-300 volunteer patients (with disease) to determine the drug's efficacy and adverse reactions (sometimes divided into Phase IIa pilot trials and Phase IIb well-controlled trials). Phase III trials are larger controlled trials in approximately 1,000-3,000 patients to verify efficacy and monitor adverse reactions during longer-term use (sometimes divided into Phase IIIa trials conducted before regulatory submission and Phase IIIb trials conducted after regulatory submission but before approval). Phase IV trials are postmarketing studies to monitor long-term effects and provide additional information on safety and efficacy, including for different regimens patient groups.

Placebo: an inactive substance or treatment given to satisfy a patient's expectation for treatment. In some controlled trials (particularly investigations of drug treatments) placebos that are made to be indistinguishable by patients (and providers when possible) from the true intervention are given to the control group to be used as a comparative basis for determining the effect of the investigational treatment.

Placebo effect: the effect on patient outcomes (improved or worsened) that may occur due to the expectation by a patient (or provider) that a particular intervention will have an effect. The placebo effect is independent of the true effect (pharmacological, surgical, etc.) of a particular intervention. To control for this, the control group in a trial may receive a placebo.

Power: the probability of detecting a treatment effect of a given magnitude when a treatment effect of at least that magnitude truly exists. For a true treatment effect of a given magnitude, power is the probability of avoiding Type II error, and is generally defined as (1 - b).

Precision: the degree to which a measurement (e.g., the mean estimate of a treatment effect) is derived from a set of observations having small variation (i.e., close in magnitude to each other). A precise estimate is not necessarily an accurate one. (Contrast with **accuracy.**)

Predictive value negative: an operating characteristic of a diagnostic test; predictive value negative is the proportion of persons with a negative test who truly do not have the disease, determined as: [true negatives _ (true negatives + false negatives)]. It varies with the prevalence of the disease in the population of interest. (Contrast with **predictive value negative.**)

Predictive value positive: an operating characteristic of a diagnostic test; predictive value positive is the proportion of persons with a positive test who truly have the disease, determined as: [true positives ¸ (true positives + false positives)]. It varies with the prevalence of the disease in the population of interest. (Contrast with **predictive value negative.**)

Premarketing Approval (PMA) Application: an application made by the sponsor of a health device to the FDA for approval to market the device in U.S. interstate commerce. The application includes information documenting the safety and efficacy/effectiveness of the device.

Prevalence: the number of people in a population with a specific disease or condition at a given time, usually expressed as a ratio of the number of affected people to the total population.

Primary study: an investigation that collects original (primary) data from patients, e.g., randomized controlled trials, observational studies, series of cases, etc. (Contrast with **synthetic/integrative study**).

Probability distribution: portrays the relative likelihood that a range of values is the true value of a treatment effect. This distribution is typically shown in the form of a bell-shaped curve. An estimate of the most likely true value of the treatment effect is the value at the highest point of the distribution. The area under the curve between any two points along the range gives the probability that the true value of the treatment effect lies between those two points. Thus, a probability distribution can be used to determine an interval that has a designated probability (e.g., 95%) of including the true value of the treatment effect.

Prospective study: a study in which the investigators plan and manage the intervention of interest in selected groups of patients. As such, investigators do not know what the outcomes will be when they undertake the study. (Contrast with **retrospective study**.)

Publication bias: unrepresentative publication of research reports that is not due to the quality of the research but to other characteristics, e.g., tendencies of investigators to submit, and publishers to accept, positive research reports (i.e., ones with results showing a beneficial treatment effect of a new intervention).

Quality-adjusted life year (QALY): a unit of health care outcomes that adjusts gains (or losses) in years of life subsequent to a health care intervention by the quality of life during those years. QALYs can provide a common unit for comparing cost-utility across different interventions and health problems. Analogous units include disability-adjusted life years (DALYs) and healthy-years equivalents (HYEs).

Quality assessment: a measurement and monitoring function of quality assurance for determining how well health care is delivered in comparison with applicable standards or acceptable bounds of care.

Quality assurance: activities intended to ensure that the best available knowledge concerning the use of health care to improve health outcomes is properly implemented. This involves the implementation of health care standards, including quality assessment and activities to correct, reduce variations in, or otherwise improve health care practices relative to these standards.

Quality of care: the degree to which health care is expected to increase the likelihood of desired health outcomes and is consistent with standards of health care. (See also **quality assessment** and **quality assurance**.)

Random variation (or random error): the tendency for the estimated magnitude of a parameter (e.g., based upon the average of a sample of observations of a treatment effect) to deviate randomly from the true magnitude of that parameter. Random variation is independent of the effects of systematic biases. In general, the larger the sample size is, the lower the random variation is of the estimate of a parameter. As random variation decreases, precision increases.

Randomization: a technique of assigning patients to treatment and control groups that is based only on chance distribution. It is used to diminish patient selection bias in clinical trials. Proper randomization of patients is an indifferent yet objective technique that tends to neutralize patient prognostic factors by spreading them evenly among treatment and control groups. Randomized assignment is often based on computer-generated tables of random numbers.

Randomized controlled trial (RCT): a true prospective experiment in which investigators randomly assign an eligible sample of patients to one or more treatment groups and a control group and follow patients' outcomes. (Also known as **randomized clinical trial**.)

Receiver operating characteristic (ROC) curve: a graphical depiction of the relationship between the true positive ratio (sensitivity) and false positive ratio (1 - specificity) as a function of the cutoff level of a disease (or condition) marker. ROC curves help to demonstrate how raising or lowering the cutoff point for defining a positive test result affects tradeoffs between correctly identifying people with a disease (true positives) and incorrectly labeling a person as positive who does not have the condition (false positives).

Register: see **database**.

Reliability: the extent to which an observation that is repeated in the same, stable population yields the same result (i.e., test-retest reliability). Also, the ability of a single observation to distinguish consistently among individuals in a population.

Relative risk reduction: a type of measure of treatment effect that compares the probability of a type of outcome in the treatment group with that of a control group, i.e.: $(P_c - P_t) \div P_c$. For instance, if the results of a trial show that the probability of death in a control group was 25% and the probability of death in a control group was 10%, the relative risk reduction would be: $(0.25 - 0.10) \div 0.25 = 0.6$. (See also **absolute risk reduction, number needed to treat,** and **odds ratio**.)

Retrospective study: a study in which investigators select groups of patients that have already been treated and analyze data from the events experienced by these patients. These studies are subject to bias because investigators can select patient groups with known outcomes. (Contrast with **prospective study**.)

Safety: a judgment of the acceptability of risk (a measure of the probability of an adverse outcome and its severity) associated with using a technology in a given situation, e.g., for a patient with a particular health problem, by a clinician with certain training, or in a specified treatment setting.

Sample size: the number of patients studied in a trial, including the treatment and control groups, where applicable. In general, a larger sample size decreases the probability of making a false-positive error (a) and increases the power of a trial, i.e., decreases the probability of making a false-negative error (b). Large sample sizes decrease the effect of random variation on the estimate of a treatment effect.

Sensitivity: an operating characteristic of a diagnostic test that measures the ability of a test to detect a disease (or condition) when it is truly present. Sensitivity is the proportion of all diseased patients for whom there is a positive test, determined as: [true positives ¸ (true positives + false negatives)]. (Contrast with **specificity.**)

Sensitivity analysis: a means to determine the robustness of a mathematical model or analysis (such as a cost-effectiveness analysis or decision analysis) that tests a plausible range of estimates of key independent variables (e.g., costs, outcomes, probabilities of events) to determine if such variations make meaningful changes the results of the analysis. Sensitivity analysis also can be performed for other types of study; e.g., clinical trials analysis (to see if inclusion/exclusion of certain data changes results) and meta-analysis (to see if inclusion/exclusion of certain studies changes results).

Series: an uncontrolled study (prospective or retrospective) of a series (succession) of consecutive patients who receive a particular intervention and are followed to observe their outcomes. (Also known as **clinical series** or **series of consecutive cases.**)

Specificity: an operating characteristic of a diagnostic test that measures the ability of a test to exclude the presence of a disease (or condition) when it is truly not present. Specificity is the proportion of nondiseased patients for whom there is a negative test, expressed as: [true negatives ¸ (true negatives + false positives)]. (Contrast with **sensitivity.**)

Statistical significance: a conclusion that an intervention has a true effect, based upon observed differences in outcomes between the treatment and control groups that are sufficiently large so that these differences are unlikely to have occurred due to chance, as determined by a statistical test. Statistical significance indicates the probability that the observed difference was due to chance if the null hypothesis is true; it does not provide information about the magnitude of a treatment effect. (Statistical significance is necessary but not sufficient for **clinical significance**.)

Statistical test: a mathematical formula (or function) that is used to determine if the difference in outcomes of a treatment and control group are great enough to conclude that the difference is statistically significant. Statistical tests generate a value that is associated with a particular *P* value. Among the variety of common statistical tests are: *F, t, Z,* and *chi-square.* The choice of a test depends upon the conditions of a study, e.g., what type of outcome variable used, whether or not the

patients were randomly selected from a larger population, and whether it can be assumed that the outcome values of the population have a normal distribution or other type of distribution.

Surrogate endpoint: an outcome measure that is used in place of a primary endpoint (outcome). Examples are decrease in blood pressure as a predictor of decrease in strokes and heart attacks in hypertensive patients, and increase in T-cell (a type of white blood cell) counts as an indicator of improved survival of AIDS patients. Use of a surrogate endpoint assumes that it is a reliable predictor of the primary endpoint(s) of interest.

Synthetic (or integrative) study: a study that does not generate primary data but that involves the qualitative or quantitative consolidation of findings from multiple primary studies. Examples are literature review, meta-analysis, decision analysis, and consensus development. (Contrast with primary study.)

Technological imperative: the inclination to use a technology that has potential for some benefit, however marginal or unsubstantiated, based on an abiding fascination with technology, the expectation that new is better, and financial and other professional incentives.

Technology: the application of scientific or other organized knowledge--including any tool, technique, product, process, method, organization or system--to practical tasks. In health care, technology includes drugs; diagnostics, indicators and reagents; devices, equipment and supplies; medical and surgical procedures; support systems; and organizational and managerial systems used in prevention, screening, diagnosis, treatment and rehabilitation.

Treatment effect: the effect of a treatment (intervention) on outcomes, i.e., attributable only to the effect of the intervention. Investigators seek to estimate the true treatment effect using the difference between the observed outcomes of a treatment group and a control group. (See **effect size**.)

Type I error: same as **false-positive error**.

Type II error: same as **false-negative error.**

Utility: the relative desirability or preference (usually from the perspective of a patient) for a specific health outcome or level of health status.

Validity: The extent to which a measure accurately reflects the concept that it is intended to measure.

REFERENCES

Altman, L. K. "Officials Apologize for Disputed Cancer Study." The New York Times; April 14, 1994: A8.

Anderson, G. F., M. A. Hall, and E. P. Steinberg. "Medical Technology Assessment and Practice Guidelines: Their Day in Court." American Journal of Public Health 83, no. 3 (1993): 1635-9.

Andersson, I., K. Aspegren, L. Janzon, T. Landberg, K. Lindholm, F. Linell, O. Ljungberg, J. Ronstam, and B. Sigfusson. "Mammographic Screening and Mortality From Breast Cancer: The Malm Mammographic Screening Trial." BMJ 297, no. 6654 (1988): 943-8.

Angell, M., and J. P. Kassirer. "Clinical Research -- What Should the Public Believe." New England Journal of Medicine 331, (1984): 189-90.

Anonymous. "Panel Questions Rules on Trauma Research," The New York Times; April 24, 1994:10.

Antczak-Bouckoms, A., E. Burdick, S. Klawansky, and F. Mosteller. "Using Medical Registries and Data Sets for Technology Assessment." International Journal of Technology Assessment in Health Care 7, no. 2 (1991): 123-8.

Applebaum, P. S., L. H. Roth, C. W. Lidz, P. Benson, and W. Winslade. "False Hopes and Best Data: Consent to Research and the Therapeutic Misconception." Hastings Center Report 17, April (1987): 20-24.

Aristotle. "Posterior Analytics." In Basic Works of Aristotle, edited by R. McKeown, 72a.25-72b.4, p. 113. New York: Random House, 1941.

Aspirin Myocardial Infarction Study Research Group. "A Randomized, Controlled Trial of Aspirin in Persons Recovered From Myocardial Infarction." Journal of the American Medical Association 243, no. 7 (1980): 661-9.

Auston, I., M. A. Cahn, and C. R. Selden. "Literature Search Methods For the Development of Clinical Practice Guidelines." In Clinical Practice Guideline Development: Methodology Perspectives edited by K. A. McCormick, S. R. Moore, and R. A. Siegel, 123-7. Rockville, MD: Agency for Health Care Policy Research, pub. no. 95-0009, 1994.

Banta, H. D., and B. R. Luce. Health Care Technology and its Assessment: An International Perspective. New York: Oxford University Press, 1993.

Barber, B. "Resistance by Scientists to Scientific Discovery." Science 134, (1961): 596-602.

Barnes, R.V., L. Lawton, and D. Briggs. "Clinical Benchmarking Improves Clinical Paths: Experience With Coronary Artery Bypass Grafting." Joint Commission on Quality Improvement 20, no. 5 (1994): 267-76.

Beauchamp, T. L., and J. F. Childress. Principles of Biomedical Ethics, 4th ed. New York: Oxford University Press, 1993.

Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. Journal of the American Medical Association 1996;276:637-9.

Bennett, J. C., for the Board on Health Sciences Policy of the Institute of Medicine. "Inclusion of Women in Clinical Trials -- Policies for Population Subgroups." New England Journal of Medicine 329, (1993): 288-92.

Bergner, M., R. A. Bobbitt, W. B. Carter, and B. S. Gilson. "The Sickness Impact Profile: Development and Final Revision of a Health Status Measure." Medical Care 19, no. 8 (1981): 787-805.

Berkey, T. "Benchmarking in Health Care: Turning Challenges into Success." Joint Commission on Quality Improvement 20, no. 5 (1994): 277-84.

Bero, L. and Rennie, D. The Cochrane Collaboration. Preparing, maintaining, and disseminating sytematic reviews of the effects of health care. Journal of the American Medical Assocation 274, no. 24 (1995): 1935-8.

Blendon, R. J., and K. Donelan. "Discrimination Against People with AIDS: The Public's Perspective." New England Journal of Medicine 319, (1988): 1022-26.

Breddin, K., D. Loew, K. Lechner, K. Uberla, and E. Walter. "Secondary Prevention of Myocardial Infarction: A Comparison of Acetylsalicylic Acid, Placebo and Phenprocoumon." Haemostasis 9, no. 6 (1980): 325-44.

Bronzino, J. D., V. H. Smith, and M. L. Wade. Medical Technology and Society: An Interdisciplinary Perspective. New York: McGraw Hill Publishing Co., 1991.

Brooks, H., and R. Bowers. "The Assessment of Technology." Science 222, no. 2 (1970): 13-20.

Bunch, W. H., and V. M. Dvonch. "The Value of Functional Independence Measure Scores." American Journal of Physical Medicine and Rehabilitation 73, no. 1 (1994): 40-3.

Buring, J. E., M. A. Jonas, and C. H. Hennekens. "Large and Simple Randomized Trials." In Tools for Evaluating Health Technologies: Five Background Papers, U.S. Congress, Office of Technology Assessment, 67-91. BP-H-142. Washington, DC: U.S. Government Printing Office, 1995.

Buxton, M. J. "Problems in the Economic Appraisal of New Health Technology: The Evaluation of Heart Transplants in the UK." In Economic Appraisal of Health Technology in the European Community, edited by M. F. Drummond. Oxford Medical Publications, 1987.

Canadian Coordinating Office for Health Technology Assessment. Guidelines for Economic Evaluation of Pharmaceuticals: Canada. 1st Edition. Ottawa, 1994. Canadian Coordinating Office for Health Technology Assessment. (brochure). Ottawa, 1995.

Cebul, R. D. "Randomized, Controlled Trials Using the Metro Firm System." Medical Care 29, suppl. 7 (1991): JS9-18.

Chalmers, I. "The Cochrane Collaboration: Preparing, Maintaining, and Disseminating Systematic Reviews of the Effects of Health Care." Annals of the New York Academy of Sciences 703, (1993): 156-63.

Chalmers, T. C. "Implications of Meta-analysis: Need for a New Generation of Randomized Control Trials." In Clinical Practice Guideline Development: Methodology Perspectives, edited by K. A. McCormick, S. R. Moore, and R. A. Siegel, 1-3. Rockville, MD: Agency for Health Care Policy Research, 1994.

Chalmers, T.C. "The Impact of Controlled Trials on the Practice of Medicine." Mount Sinai Journal of Medicine 41, no. 6 (1974): 753-9.

Chalmers, T. C., C.S. Frank, and D. R. Reitman. "Minimizing the Three Stages of Publication

Bias."Journal of the American Medical Association 263, no. 10 (1990): 1392-5.
Chalmers, T. C., P. Hewett, D. Reitman, and H. S. Sacks. "Selection and Evaluation of Empirical Research in Technology Assessment."International Journal of Technology Assessment in Health Care 5, no. 4 (1989): 521-36.
Chalmers, T. C., H. Smith, Jr., B. Blackburn, B. Silverman, B. Schroeder, D. Reitman, and A. Ambroz. "A Method for Assessing the Quality of a Randomized Controlled Trial." Controlled Clinical Trials 2, no. 1 (1981): 31-49.

Coates & Jarratt, Inc.Course Workbook: Technology Assessment. Anticipating the Consequences of Technological Choices. August 17, 1992. Washington, DC

Cochrane, A. L.Effectiveness and Efficiency: Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust, 1972.

Coleman, R. E., W. H. Briner, and B. A. Siegel,"Clinical PET Scanning. A Short-Lived Orphan." International Journal of Technology Assessment in Health Care 8, no. 4 (1992): 610-22.

Comroe, J. J., Jr., and R. D. Dripps."Scientific Basis for the Support of Biomedical Science." Science 192, 1976: 105-11.

Dickersin, K., and Y. I. Min."Publication Bias: The Problem That Won't Go Away." Annals of the New York Academy of Sciences 703, (1993): 135-46.

Dickersin, K., R. Scherer, and C. Lefebvre."Identifying Relevant Studies for Systematic Reviews." BMJ 309, no. 6964 (1994): 1286-91.

Doll, H. A., N. A. Black, A. B. Flood, and K. McPherson."Criterion Validation of the Nottingham Health Profile: Patient Views of Surgery for Benign Prostatic Hypertrophy." Social Science and Medicine 37, no. 1 (1993): 115-22.

Donabedian, A."Quality Assessment and Assurance: Unity of Purpose, Diversity of Means." Inquiry 25, no. 1 (1988): 173-92.

Donaldson, M. S., and H. C. Sox, Jr., eds.Setting Priorities for Health Technology Assessment: A Model Process. Washington, DC: National Academy Press, 1992.

Doughtery, C. J."Ethical Perspectives on Prospective Payment." Hastings Center Report 19, January/February (1989): 5-11.
Dougherty, C.J. "Quality-adjusted Life Years and Ethical Values of Health Care." American Journal of Physical Medicine and Rehabilitation 73, no. 1 (1994): 61-5.

Drummond, M., A. Brandt, B. Luce, and J. Rovira."Standardizing Methodologies for Economic Evaluation in Health Care. Practice, Problems, and Potential." International Journal of Technology Assessment in Health Care 9, no. 1 (1993): 26-36.

Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ 1996; 313:275-83.

DuBose, E. R., R. Hamel, and L. J. O'Connell, eds.A Matter of Principles? Ferment in U.S. Bioethics.

Valley Forge, PA: Trinity Press International, 1994.

Eckman, M. H., H. J. Levine, S. G. Pauker."Decision Analytic and Cost-effectiveness Issues Concerning Anticoagulant Prophylaxis in Heart Disease." Chest 102, suppl. 4 (1992): 538-49S.

Eddy, D. M."Selecting Technologies for Assessment." International Journal of Technology Assessment in Health Care 5, no. 4 (1989): 485-501.
Eddy, D. M.A Manual for Assessing Health Practices & Designing Practice Policies: The Explicit Approach. Philadelphia: American College of Physicians, 1992.

Eddy, D. M., V. Hasselblad, and R. Shachter."A Bayesian Method for Synthesizing Evidence: The Confidence Profile Method." International Journal of Technology Assessment in Health Care 6, no. 1 (1990): 31-55.

Elixhauser, A., B. R. Luce, W. R. Taylor, and J. Reblando."Health Care CBA/CEA: An Update on the Growth and Composition of the Literature." Medical Care 31, suppl. 7 (1993): JS1-11, JS18-149.

Ellenberg, S. S."Do Large, Simple Trials Have a Place in the Evaluation of AIDS Therapies?" Oncology 6, (1992): 55-9, 63.

Elwood, P. C."British Studies of Aspirin and Myocardial Infarction." American Journal of Medicine 74, no. 6A (1983): 50-4.

Elwood, P. C., A. L. Cochrane, M. L. Burr, P. M. Sweetnam, G. Williams, E. Welsby, and S. J. Hughes."A Randomized Controlled Trial of Acetyl Salicylic Acid in the Secondary Prevention of Mortality from Myocardial Infarction." BMJ 1, no. 905 (1974): 436-40.

Elwood, P. C., and P. M. Sweetnam."Aspirin and Secondary Mortality After Myocardial Infarction." Lancet 2, no. 8156-8157 (1979): 1313-5.

Engelhardt, H. T., Jr."The Disease of Masturbation: Values and the Concept of Disease." Bulletin of the History of Medicine 48, (1974): 234-48.
Engelhardt, H. T., Jr. "Ethics in Pluralist Societies." Perspectives in Biology and Medicine 26, (1982): 64-78.

Enkin, M., M. J. N. C. Keirse, M. Renfrew, and J. Neilson.A Guide to Effective Care in Pregnancy and Childbirth. Second Edition. New York: Oxford University Press, 1995.

Essink-Bot, M. L., M. E. Stouthard, and G. J. Bonsel."Generalizability of Valuations on Health States Collected With the EuroQol© Questionnaire." Health Economics 2, no. 3 (1993): 237-46.

EuroQol© Group."EuroQol© -- A New Facility for the Measurement of Health Related Quality of Life." Health Policy 16, no. 3 (1990): 199-208.

Evidence-based Medicine Working Group."Evidence-based Medicine: A New Approach to Teaching the Practice of Medicine." Journal of the American Medical Association 268, no. 17 (1992): 2420-5.

Feeny, D., G. ." In Contemporary Moral Controversies in Technology, edited by A. P. Iannone, 17-26. New York: Oxford University Press, 1987.

Frawley, S. J. "Building a Database of Data Sets for Health Services Research." Proceedings of the Annual Symposium on Computer Applications in Medical Care (1994): 377-81.

Freedman, B. "Equipoise and the Ethics of Clinical Research." New England Journal of Medicine 317, (1987): 141-45.

Freiman, J. A., T. C. Chalmers, H. Smith, Jr., and R. R. Kuebler. "The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Controlled Trial. Survey of 71 "Negative" Trials." New England Journal of Medicine 299, no. 13 (1978): 690-4.

Friede, A., J. A. Reid, and H. W. Ory. "CDC WONDER: A Comprehensive On-line Public Health Information System of the Centers for Disease Control and Prevention." American Journal of Public Health 83, no. 9 (1993): 1289-94.

Gafni, A., and S. Birch. "Searching for a Common Currency: Critical Appraisal of the Scientific Basis Underlying European Harmonization of the Measurement of Health Related Quality of Life (EuroQol©)." Health Policy 23, no. 3 (1993): 219-28.

Gallagher M., T. Hares, J. Spencer, C. Bradshaw, and I. Webb. "The Nominal Group Technique: A Research Tool for General Practice?" Family Practice 10, no. 1 (1993): 76-81.

Gann, M. J., and J. D. Restuccia. "Total Quality Management in Health Care: A View of Current and Potential Research." Medical Care Review 51, no. 4 (1994): 467-500.

Gelijns, A., and N. Rosenberg. "The Dynamics of Technological Change in Medicine." Health Affairs 13, no. 3 (1994): 28-46.

Gerard, K., and G. Mooney. "QALY League Tables: Handle With Care." Health Economics 2, no. 1 (1993): 59-64.

Gold, J. A., M. J. Zaremski, E. R. Lev, and D. H. Shefrin. "Daubert v. Merrell Dow. The Supreme Court Tackles Scientific Evidence in the Courtroom." Journal of the American Medical Association 270, no. 24 (1993): 2964-7.

Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.

Goldberg, H. I., M. A. Cummings, E. P. Steinberg, E. M. Ricci, T. Shannon, S.B. Soumerai, B. S. Mittman, J. Eisenberg, and D. A. Heck. "Deliberations on the Dissemination of PORT Products: Translating Research Findings into Improved Patient Outcomes." Medical Care 32, suppl. 7 (1994): JS90-110.

Goodman, C., ed. Medical Technology Assessment Directory: A Pilot Reference to Organizations, Assessments, and Information Resources. Washington, DC: Institute of Medicine, 1988.

Goodman, C. "It's Time to Rethink Health Care Technology Assessment." International Journal of Technology Assessment in Health Care 8, no. 2 (1992): 335-58.
Goodman, C. Literature Searching and Evidence Interpretation for Assessing Health Care Practices.

Stockholm: Swedish Council on Technology Assessment in Health Care, 1993.
Goodman, C. The moving target problem and other lessons from percutaneous transluminal coronary angioplasty. In A. Szczepura, J. Kankaanpää, eds. Assessment of Health Care Technologies: Case Studies, Key Concepts and Strategic Issues. New York: John Wiley & Sons, 1996:29-65.

Goodman, C., and S. R. Baratz, eds. Improving Consensus Development for Health Technology Assessment: An International Perspective. Washington, DC: National Academy Press, 1990.

Goodman, C., Snider, G., and Flynn, K. Health Care Technology Assessment in VA. Boston, Mass.: Management Decision and Research Center. Washington, DC: Health Services Research and Development Service, 1996.

Goodman, C. S. "Technology Assessment in Healthcare: A Means for Pursuing the Goals of Biomedical Engineering." Medical and Biological Engineering and Computing 31, no. 1 (1993): HTA3-10.

Goodwin, J. S., and J. M. Goodwin. "The Tomato Effect: Rejection of Highly Efficacious Therapies." Journal of the American Medical Association 251, (1984): 2387-90.

Grimes, D. A. "Technology Follies: The Uncritical Acceptance of Medical Innovation." Journal of the American Medical Association 269, no. 23 (1993): 3030-3.

Gruber, M. "The Development of a Position Statement Using the Delphi Technique." Gastroenterology Nursing 16, no. 2 (1993): 68-71.

Hasselblad, V., and L. V. Hedges. "Meta-Analysis of Screening and Diagnostic Tests." Psychological Bulletin 117 no. 1 (1995): 167-78.

Haynes, L. M., A. A. Patterson, and S. U. Wade. "Drug Information Service for Drug Procurement in the Veterans Affairs Health-care System: Preliminary Experience." American Journal of Hospital Pharmacy 49, no. 3 (1992): 59-58.

Hayward R. S. A., M. C. Wilson, S. R. Tunis, E. B. Bass, G. Guyatt. "Users' Guides to the Medical Literature. VIII. How to Use Clinical Practice Guidelines. A. Are the Recommendations Valid?" Journal of the American Medical Association 274, no. 7 (1995): 570-4.

Health Care Technology Institute. Decomposition of the Health Care Spending Residual. Alexandria, VA, 1994.

Heitman, E. "Cultural Diversity and the Clinical Encounter: Intercultural Dialogue in Multi-ethnic Patient Care." In Theological Analyses of the Clinical Encounter, edited by G. P. McKenny and J. Sande, 202-22. Amsterdam: Klewer Academic Publishers, 1994.

Hilborne, L. H., L. L. Leape, J. P. Kahan, R. E. Park, C. J. Kamberg, and R. H. Brook. Percutaneous Transluminal Coronary Angioplasty: A Literature Review of Ratings of Appropriateness and Necessity. Santa Monica, CA: RAND, JRA-01,1991.

Hogness, J. R., M. Van Antwerp. The Artificial Heart: Prototypes, Policies, and Patients. Washington, DC: National Academy Press, 1991.

In re. Helga Wanglie, Fourth Judicial District (Dist. Ct. Probate Ct. Div.) PX-91-283. Minnesota, Hennepin County.

Institute of Medicine. Assessing Medical Technologies. Washington, DC: National Academy Press, 1985.
Institute of Medicine. Assessment of Diagnostic Technology in Health Care. Rationale, Methods, Problems, and Directions. Washington, DC: National Academy Press, 1989.
Institute of Medicine. Consensus Development at the NIH: Improving the Program. Washington, DC: National Academy Press, 1990.

International Committee of Medical Journal Writers. "Conflict of Interest." Lancet 341, no. 8847 (1993): 742-3.
International Committee of Medical Journal Writers. Uniform requirements for manuscripts submitted to biomedical journals. New England Journal of Medicine 1997336:309-15.
In the Matter of Baby K, 832 F.Supp 1022 (E.D. Va. 1993) and 16 F.3d 590 (4th Cir. 1994).

Irwig, L., A. N. Tosteson, C. Gatsonis, J. Lau, G. Colditz, T. C. Chalmers, and F. Mosteller. "Guidelines for Meta-analyses Evaluating Diagnostic Tests." Annals of Internal Medicine 120, no. 8 (1994): 667-76.

Jairath N., and J. Weinstein. "The Delphi Methodology (Part Two): A Useful Administrative Approach." Canadian Journal of Nursing Administration 7, no. 4 (1994): 7-20.

Jenkinson, C., R. Fitzpatrick, and M. Argyle. "The Nottingham Health Profile: An Analysis of its Sensitivity to Differentiating Illness Groups." Social Science and Medicine 27, no. 12 (1988): 1411-4.

Johannesson M., Pliskin J. S., Weinstein, M. C. Are healthy-years equivalents an improvement over quality-adjusted life years? Medical Decision Making 13, no. 4 (1993):281-6.

Joint Strategy Group of the Government and the Association of the British Pharmaceutical Industry. Guidelines on Good Practice in the Conduct of Economic Evaluation of Medicines. London: Association of the British Pharmaceutical Industry, 1994.

Kahan, J. P., S. J. Bernstein, L. L. Leape, L. H. Hilborne, R. E. Park, L. Parker, C. J. Kamberg, and R. H. Brook. "Measuring the Necessity of Medical Procedures." Medical Care 32, no. 4 (1994): 357-65.

Kahan, J. P., Kanouse, D. E., and J. D. Winkler. "Stylistic Variations in National Institutes of Health Consensus Statements, 1979-1983." International Journal of Technology Assessment in Health Care 4, no. 2 (1988): 289-304.

Kant, I. Groundwork of the Metaphysic of Morals. Translated by H.J. Paton, New York: Harper Torch Books, 1964.

Kaplan, R. M., and Anderson, J. P. "A General Health Policy Model: Update and Applications." Health Services Research 23, (1988): 203-35.

Kaplan, R. M., J. P. Anderson, A. W. Wu, W. C. Mathews, F. Kozin, and D. Orenstein. "The Quality

of Well-Being Scale: Applications in AIDS, Cystic Fibrosis, and Arthritis." Medical Care 27, suppl. 3 (1989): S27-43.

Kassirer, J. P., and M. Angell. "Financial Conflicts of Interest in Biomedical Research." New England Journal of Medicine 329, no. 8 (1993): 570-1.

Kassirer, J. P. "Our Stubborn Quest for Diagnostic Certainty: A Cause of Excessive Testing." New England Journal of Medicine 320, (1989): 1489-91.

Katz, S., T. D. Downs, H. R. Cash, and R. C. Grotz. "Progress in the Development of the Index of ADL." Gerontologist 10, part I (1970): 20-30.

Kazis, L. E., J. J. Anderson, and R. F. Meenan. "Effect Sizes for Interpreting Changes in Health Status." Medical Care 27, suppl. 3 (1989): S178-189.

Kleinman L. C., J. Kosecoff, R. W. Dubois, and R. H. Brook. "The Medical Appropriateness of Tympanostomy Tubes Proposed for Children Younger Than 16 Years in the United States." Journal of the American Medical Association 271, no. 16 (1994): 1250-5.

Kolata, G. "Patients Lawyers Lead Insurers to Pay for Unproven Treatments: Coverage Even When Policy Excludes a Therapy." The New York Times; March 28, 1994: A1, A7.

Kritchevsky, S. B., and B. P. Simmons. "Continuous Quality Improvement: Concepts and Applications for Physician Care." Journal of the American Medical Association 266, no. 13 (1991): 1817-23.

L'Abbé, K. A., A. S. Detsky, and K. O'Rourke. "Meta-analysis in Clinical Research." Annals of Internal Medicine 107, no. 2 (1987): 224-33.

Laird, N. M., and F. Mosteller. "Some Statistical Methods for Combining Experimental Results." International Journal of Technology Assessment in Health Care 6, no. 1 (1990): 5-30.

Langley PC. The November 1995 revised Australian guidelines for the economic evaluation of pharmaceuticals. PharmacoEconomics 1996;9:341-52.

La Puma, J., and E. F. Lawlor. "Quality-adjusted Life-years: Ethical Implications for Physicians and Policymakers." Journal of the American Medical Association 263, no. 21 (1990): 2917-2921.

Lara, M. E., and C. Goodman, eds. National Priorities for the Assessment of Clinical Conditions and Medical Technologies. Washington, DC: National Academy Press, 1990.

Lau, J., E. M. Antman, J. Jiminez-Silva, B. Kupelnick, F. Mosteller, and T. C. Chalmers. "Cumulative Meta-analysis of Therapeutic Trials for Myocardial Infarction." New England Journal of Medicine 327, no. 4 (1992): 248-54.

Lazaridis, E. N., M. A. Rudberg, M. A. Furner, and C. K. Cassel. "Do Activities of Daily Living Have a Hierarchical Structure? An Analysis Using the Longitudinal Study of Aging." Journal of Gerontology 49, no. 2 (1994): M47-51.

Leape, L. L., L. H. Hilborne, J. P. Kahan, W. B. Stason, R. E. Park, C. J. Kamberg, and R. H. Brook. Coronary Artery Bypass Graft: A Literature Review and Ratings of Appropriateness and Necessity. Santa Monica, CA: RAND, 1991.

Lenz, S., S. Myers, S. Nordlund, D. Sullivan, and V. Vasista. "Benchmarking: Finding Ways to Improve." Joint Commission on Quality Improvement 20, no. 5 (1994): 250-9.

Levi, J. "Unproven AIDS Therapies: The Food and Drug Administration and ddI." In Biomedical Politics, by Institute of Medicine, 9-37. Washington, DC: National Academy Press, 1991.

The Lewin Group. Outcomes and Effectiveness Research in the Private Sector: Final Report to the Agency for Health Care Policy and Research. Fairfax, Va., 1997.

Linacre, J. M., A. W. Heinemann, B. D. Wright, C. V. Granger, and B. B. Hamilton. "The Structure and Stability of the Functional Independence Measure." Archives of Physical Medicine and Rehabilitation 75, no. 2 (1994): 127-32.

Littenberg, B., and L. E. Moses. "Estimating Diagnostic Accuracy from Multiple Conflicting Reports: A New Meta-analytic Method." Medical Decision Making 13, no. 4 (1993): 313-21.

Lohr, K. N., ed. Institute of Medicine. Medicare: A Strategy for Quality Assurance. Volume I. Washington, DC: National Academy Press, 1990.

Lohr, K. N., and R. A. Rettig, eds. Quality of Care and Technology Assessment. Report of a Forum of the Council on Health Care Technology. Washington, DC: National Academy Press, 1988.

MacIntyre, A. "Utilitarianism and Cost-benefit Analysis: An Essay on the Relevance of Moral Philosophy to Bureaucratic Theory." In Ethical Theory and Business, edited by T. L. Beauchamp and N. E. Bowie, 266-87. Englewood Cliffs, NJ: Prentice-Hall, 1983.

Mant D., and G. Fowler. "Mass Screening: Theory and Ethics." British Medical Journal 300, (1990): 916-18.

Mason JM. Cost-per QALY league tables: their roles in pharmacoeconomic analysis. Pharmacoeconomics 1994;5:472-81.

Mastroianni, A. C., R. Faden, and D. Federman. Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies. Washington, DC: National Academy Press, 1994.

McGlynn, E. A., J. Kosecoff, and R. H. Brook. "Format and Conduct of Consensus Development Conferences. Multi-nation Comparison." International Journal of Technology Assessment in Health Care 6, no. 3 (1990): 450-69.

McHorney, C. A., J. E. Ware, Jr., J. F. Lu, and C. D. Sherbourne. "The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of Data Quality, Scaling Assumptions, and Reliability Across Diverse Patient Groups." Medical Care 32, no. 1 (1994): 40-66.

McKinlay, J. B. "From "Promising Report" to "Standard Procedure": Seven Stages in the Career of a

Medical Innovation." Milbank Memorial Fund Quarterly 59, no. 3 (1981): 374-411.

Meenan, R. F., J. H. Mason, J. J. Anderson, A. A. Guccione, and L. E. Kazis. "AIMS2. The Content and Properties of R revised and Expanded Arthritis Impact Measurement Scales Health Status Questionnaire." Arthritis and Rheumatism 35, no. 1 (1992): 1-10.

Mehrez, A., and A. Gafni. "Healthy-years Equivalents Versus Quality-adjusted Life Years: In Pursuit of Progress." Medical Decision Making 13, no. 4 (1993): 287-92.

Merigan, T. C. "You Can Teach an Old Dog New Tricks: How AIDS Trials are Pioneering New Strategies." New England Journal of Medicine 323, no. 19 (1990): 1341-3.

Mills, R. M., Jr., and J. M. Kalan. "Developing a Rational Management Strategy for Angina Pectoris After Coronary Bypass Surgery: A Clinical Decision Analysis." Clinical Cardiology 14, no. 3 (1991): 191-7.

Mittman, B. S., and A. L. Siu. "Changing Provider Behavior: Applying Research on Outcomes and Effectiveness in Health Care." In Improving Health Policy and Management: Nine Critical Research Issues for the 1990s, edited by S. M. Shortell and U. E. Reinhardt, 195-226. Ann Arbor, MI: Health Administration Press, 1992.

Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1996;272:122-4.

National Academy of Engineering, Committee on Public Engineering Policy. A Study of Technology Assessment. Washington, DC: U.S. Government Printing Office, 1969.

National Research Council, Committee on the Life Sciences and Social Policy. Assessing Biomedical Technologies: An Inquiry into the Nature of the Process. Washington, DC: National Academy of Sciences, 1975.

Nelkin, D. "The Social Power of Genetic Information." In The Code of Codes: Scientific and Social Issues in the Human Genome Project, edited by D. J. Kelves and L. Hood. Cambridge: Harvard University Press, 1992.

Neuhauser, D., and A. M. Lewicki. "What Do We Gain From the Sixth Stool Guaiac?" New England Journal of Medicine 293, no. 5 (1975): 226-8.

Newcomer, L. N. "Defining Experimental Therapy -- A Third Party Payer's Dilemma." New England Journal of Medicine 323, no. 24 (1990): 1702-4.

Newhouse, J. P. "Medical Care Costs: How Much Welfare Loss?" Journal of Economic Perspectives 6, no. 3 (1992): 3-21.

Nord, E. "The QALY -- A Measure of Social Value Rather Than Individual Utility?" Health Economics 3, no. 2 (1994): 89-93.

O'Brien, B. J., M. J. Buxton, and D. L. Patterson. "Relationship Between Functional Status and Health-related Quality-of-life After Myocardial Infarction." Medical Care 31, no. 10 (1993): 950-5.

Olson, C. M. "Consensus Statements: Applying Structure." Journal of the American Medical Association 273, no. 1 (1995): 72-3.

Ontario Ministry of Health. Ontario Guidelines for Economic Analysis of Pharmaceutical Products. Toronto, 1993.

Oxman, A. D., and F. H. Guyatt. "Guidelines for Reading Literature Reviews." Canadian Medical Association Journal 138, no. 8 (1988): 697-703.

Passamani, E. "Clinical Trials: Are They Ethical?" New England Journal of Medicine 324, no. 22 (1991): 1589-92.

Patrick, D. L., and R. A. Deyo. "Generic and Disease-specific Measures in Assessing Health Status and Quality of Life." Medical Care 27, suppl. 3 (1989): S217-32.

Patrick, D. L., H. E. Starks, K. C. Cain, R. F. Uhlmann, and R. A. Pearlman. "Measuring Preferences for Health States Worse Than Death." Medical Decision Making 14, no. 1 (1994): 9-18.

Pauker, S. G., and J. P. Kassirer. "Decision Analysis." New England Journal of Medicine 316, no. 5 (1987): 250-8.

Peters, W. P., and M. C. Rogers. "Variation in Approval By Insurance Companies of Coverage for Autologous Bone Marrow Transplantation for Breast Cancer." New England Journal of Medicine 330, no. 7 (1994): 473-7.

Peto, R., R. Collins, and R. Gray. "Large-scale Randomized Evidence: Large, Simple Trials and Overviews of Trials." Annals of the New York Academy of Sciences 703, (1993): 314-40.

Phelps, C. E., and C. Mooney. "Correction and Update on ´Priority Setting in Medical Technology Assessment.'" Medical Care 30, no. 8 (1992): 744-51.

Phelps, C. E., and A. I. Mushlin. "On the (Near) Equivalence of Cost-Effectiveness and Cost-Benefit Analyses." International Journal of Technology Assessment in Health Care 7, (1991): 12-21.

Poe, N. R., and R. I. Griffiths. "Clinical-economic Trials." In Tools for Evaluating Health Technologies: Five Background Papers. U.S. Congress, Office of Technology Assessment, BP-H-142, 125-49. Washington, DC: U.S. Government Printing Office, 1995.

Reiser, S. J. "Criteria for Standard Versus Experimental Therapy." Health Affairs 13, no. 3 (1994): 127-36.

Relman, A. S. "Financial Incentives in Clinical Investigation." New England Journal of Medicine 320, (1989): 933-34.

Rettig, R. A. Health Care in Transition: Technology Assessment in the Private Sector. Santa Monica, Ca.: RAND, 1997.

Rettig, R. A., and N. G. Levinsky, eds. Kidney Failure and the Federal Government. Washington, DC:

National Academy Press, 1991.

Richardson, J. "Cost Utility Analysis: What Should Be Measured?" Social Science and Medicine 39, no. 1 (1994): 7-21.

Rothman KJ, Michels KB. The continuing unethical use of placebo controls. N Engl J Med 1994;331:394-7.

Rovira, J. "Standardizing Economic Appraisal of Health Technology in the European Community." Social Science and Medicine 38, no. 12 (1994): 1675-8.

Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine. New York: Churchill Livingstone, 1997.

Sacks, H. S., J. Berrier, D. Reitman, V. A. Ancona-Berk, and T. C. Chalmers. "Meta-analysis of Randomized Controlled Trials." New England Journal of Medicine 316, no. 8 (1987): 450-5.

Sheldon T., Chalmers, I. The UK Cochrane Centre and the NHS Centre for reviews and dissemination: respective roles within the informaiton systems strategy of the NHD R&D programme, coordination and principles underlying collaboration. Health Economics 3, no. 3 (1994): 201-3.

Sibbald, W. J., and K. J. Inman. "Problems in Assessing the Technology of Critical Care Medicine." International Journal of Technology Assessment in Health Care 8, (1992): 419-43.

Simpson, K. N., and L. B. Snyder. "Informing the Mammography Coverage Debate: Results of Meta-analysis, Computer Modeling, and Issue Analysis." International Journal of Technology Assessment in Health Care 7, (1991): 616-31.

Skrabanek, P. "Mass Mammography: The Time for Reappraisal." International Journal of Technology Assessment in Health Care 5, (1989): 423-30.

Small, N. C., C. B. Sledge, and J. N. Katz. "A Conceptual Framework for Outcomes Research in Arthroscopic Meniscectomy: Results of a Nominal Group Process." Arthroscopy 10, no. 5 (1994): 486-92.

Smith, R., and M. Dobson. "Measuring Utility Values for QALYs: Two Methodological Issues." Health Economics 2, no. 4 (1993): 349-55.

Sontag, S. Illness as Metaphor. New York: Strauss and Giroux, 1978.

Soumerai, S. B., and H. L. Lipton. "Computer-Based Drug-Utilization Review -- Risk, Benefit, or Boondoggle?" New England Journal of Medicine 332, (1995): 1641-5.

Stason, W. B., and M. C. Weinstein. "Allocation of Resources to Manage Hypertension." New England Journal of Medicine 296, (1977): 732-9.

Steffen, G. E. "Quality Medical Care: A Definition." Journal of the American Medical Association 260, (1988): 56-61.

Steinberg, E. P., J. M. Tielsch, O. D. Schein, J. C. Javitt, P. Sharkey, S. D. Cassard, and L. M. Legro. "The VF-14. An Index of Functional Impairment in Patients With Cataract." Archives of Ophthalmology 112, no. 5 (1994): 630-8.

Tanouye, E. "Bristol-Myers to Market Antidepressant by Claiming Less Sexual Dysfunction." Wall Street Journal; January 6, 1995: B2.

Taylor, K. S. "Hospital Tech Assessment Teams Field Requests -- and Politics." Hospitals and Health Networks 68, no. 16 (1994): 58.

Thacker, S. B. "Meta-analysis: A Quantitative Approach to Research Integration." Journal of the American Medical Association 259, no. 11 (1988): 1685-9.

The Coronary Drug Project Research Group. "Aspirin in Coronary Heart Disease." Journal of Chronic Diseases 29, no. 10 (1976): 625-42.

The Ischemic Optic Neuropathy Decompression Trial Research Group. "Optic Nerve Decompression Surgery for Nonarteritic Anterior Ischemic Optic Neuropathy (NAION) Is Not Effective and May Be Harmful." Journal of the American Medical Association 273, no. 8 (1995): 625-32.

The Nuremberg Code, from Trials of War Criminals Before the Nuremberg Military Tribunals under Control Council Law No. 10, vol. 2, 181-82. Washington, DC: U.S. Government Printing Office, 1949.

The Persantine-Aspirin Reinfarction Study Research Group. Persantine and Aspirin in Coronary Heart Disease. Circulation 62, no. 3 (1980): 449-61.

Thornbury, J. R., and D. G. Fryback. "Technology Assessment -- an American View." European Journal of Radiology 14, no. 2 (1992): 147-56.

Thornquist, M. D., N. Urban, A. Tseng, C. Edelstein, B. Lund, and G. S. Omenn. "Research Cost Analyses to Aid in Decision Making in the Conduct of a Large Prevention Trial, CARET. Carotene and Retinol Efficacy Trial." Controlled Clinical Trials 14, no. 4 (1993): 325-39.

Thornton, J. G., R. J. Lilford, and N. Johnston. "Decision Analysis in Medicine." BMJ 304, no. 6834 (1992): 1099-103.

Torrance, G. W., D. Feeny. "Utilities and Quality-adjusted Life Years." International Journal of Technology Assessment in Health Care 5, no. 4 (1989): 559-75.

Tymstra, T. "The Imperative Character of Medical Technology and the Meaning of 'Anticipated Decision Regret.'" International Journal of Technology Assessment in Health Care 5, (1989): 207-13.

Tyson, J., and K. A. Kennedy. "Are Small Randomized Controlled Trials Unethical? Research Ethics in an Age of Meta-analysis." Pediatric Research 6, (1994): 47a.

University HealthSystem Consortium. Clinical Practice Advancement Center. Technology Report: Implementation of Hospital-Based Technology Assessment. Oak Brook Ill.: UHC Services

Corporation, 1996.

U.S. Congress, House of Representatives. Committee on Science and Astronautics. Technology Assessment. Statement of Emilio Q. Daddario, Chairman, Subcommittee on Science Research and Development. 90th Cong., 1st sess., Washington, DC, 1967.

U.S. Congress, Office of Technology Assessment. Technology Assessment in Business and Government. Summary and Analysis. Washington, DC: U.S. Government Printing Office, 1977.
U.S. Congress, Office of Technology Assessment. The Impact of Randomized Clinical Trials on Health Policy and Medical Practice: Background Paper. Washington, DC: U.S. Government Printing Office, 1983.
U.S. Congress, Office of Technology Assessment. Strategies for Medical Technology Assessment. Washington, DC: U.S. Government Printing Office, 1983.
U.S. Congress, Office of Technology Assessment. Identifying Health Technologies That Work: Searching for Evidence. OTA-H-608. Washington, DC: U.S. Government Printing Office, 1983.
U.S. Congress, Office of Technology Assessment. Protecting Privacy in Computerized Medical Information. Washington, DC: U.S. Government Printing Office, 1994.

U.S. Department of Health and Human Services, Agency for Health Care Policy and Research. Extracranial-Intracranial Bypass to Reduce the Risk of Ischemic Stroke. Health Technology Assessment Reports. No. 6. Rockville, MD, 1990.
U.S. Department of Health and Human Services, Agency for Health Care Policy and Research. Intermittent Positive Pressure Breathing: Old Technologies Rarely Die. Rockville, MD, 1993.
U.S. Department of Health and Human Services, Agency for Health Care Policy and Research. Clinical Practice Guideline Number 9: Management of Cancer Pain. Rockville, MD, 1994: 223-4.

U.S. Preventive Services Task Force. Guide to Clinical Services. Second Edition. Baltimore: Williams & Wilkins, 1996.

U.S. Preventive Services Task Force. Guide to Clinical Preventive Services. An Assessment of the Effectiveness of 169 Interventions. Baltimore: Williams & Wilkins, 1989.

van den Broek, S. A., D. J. van Veldhuisen, P. A. de Graeff, M. L. Landsman, J. Hillege, and K. I. Lie. "Comparison Between New York Heart Association Classification and Peak Oxygen Consumption in the Assessment of Functional Status and Prognosis in Patients With Mild to Moderate Chronic Congestive Heart Failure." American Journal of Cardiology 70, no. 3 (1992): 359-63.

Varmus, H., and Satcher, D. Ethical complexities of conducting research in developing countries. New England Journal of Medicine 337, no. 14 (1997): 1003-5.

Wakefield, D. S., and B. J. Wakefield. "Overcoming Barriers to Implementation of TQM/CQI in Hospitals: Myths and Realities." QRB. Quality Review Bulletin 19, no. 3 (1993): 83-8.

Ware, J. E. Jr., and C. D. Sherbourne. "The MOS 36-item Short Form Health Survey (SF-36): I. Conceptual Framework and Item Selection." Medical Care 30, no. 6 (1992): 473-83.

Weiss, R., and S. O. Thier. "HIV Testing is the Answer -- What's the Question?" New England Journal of Medicine 328, (1988): 1010-12.

Williams, B., and J. J. C. Smart. Utilitarianism: For and Against. Cambridge, England: Cambridge University Press, 1973.

World Development Report 1993: Investing in Health. The World Bank. New York: Oxford University Press, 1993.

Yusuf, S., P. Held, K. K. Teo, and E. R. Toretsky. "Selection of Patients for Randomized Controlled Trials: Implications of Wide or Narrow Eligibility Criteria." Statistics in Medicine 9, no. 1-2 (1990): 73-83.

Zeman, R. K., F. Al-Kawas, and S. B. Benjamin. "Gallstone Lithotripsy: Is There Still Cause for Optimism?" Radiologia 178, no. 1 (1991): 33-5.

Zola, I. K. "Medicine as an Institution of Social Control." Sociological Review 20, (1972): 487-504.
Zola, I. K. "Pathways to the Doctor -- From Person to Patient." Social Science and Medicine 7, (1973): 677-89.