THE FORUM FOR COLLABORATIVE HIV RESEARCH

This report is the next stage of the Forum's examination of the dissemination and implementation of clinical practice guidelines for HIV disease. The report provides a summary of the methodology, scope, and principles of patient outcomes research; an area of scientific investigation that has much to offer researchers, clinicians, and policy analysts addressing HIV disease. Through this report, we begin to examine how outcomes research can be better applied to examine the effects of new standards of care for HIV disease. In June, 1998, the Forum held a workshop to examine HIV practice guideline dissemination and implementation. The workshop brought experts in information dissemination together with experts in HIV research, care, and education. The result was a series of recommendations to better disseminate HIV treatment information to providers and patients by government, industry, health care delivery, and community. We now turn our attention to the evaluation of implementation efforts and the effects of that implementation on patient outcomes. This report provides the needed background information to begin our analysis. In September, 1999, the Forum will hold a workshop to discuss how this methodology can be applied to HIV disease. The recommendations from that workshop will be published, along with summaries of the speaker presentations.

The report is written by Helen Schietinger, who has a remarkable ability to take vasts amounts of complex information and summarize it so that it is easy to understand, yet truly substantive. Several people provided their input and assistance in the creation of this report including: Bruce Agins, Sam Bozzette, Vicki Cargill, Ruth Finkelstein, Julia Hidalgo, Jeff Levi, John Ludden, Leona Markson, Jim Neaton, Kathy Wilfert and Robert Zackin. Derek Hodel and William Gist have provided much needed support for the project.

David Barr
Executive Director

The Forum for Collaborative HIV Research,situated within the Center for Health Services Research and Policy at the George Washington University, is an independent public-private partnership composed of representatives from multiple interests in the HIV clinical research arena. These include pharmaceutical companies; public and private third-party payors; health care providers; federal and state government agencies; clinical research centers; and patient advocacy groups. The Forum primarily facilitates ongoing discussion and collaboration between appropriate stakeholders on the development and implementation of new clinical studies in HIV and on the transfer of the results of research into clinical practice. The main purpose of the Forum is to enhance collaboration between interested groups in order to address the critical unanswered questions regarding the optimal medical management of HIV disease. The Forum's Executive Committee, made up of representatives from each of the constituency groups, sets the goals and priorities of the Forum; determining which projects to undertake and how to do so.

For further information about the Forum for Collaborative HIV Research, please call William Gist at 202-530-2334 or visit our website at: www.gwumc.edu/chpr and click on HIV Research.

THE POTENTIAL OF PATIENT OUTCOMES RESEARCH IN HIV DISEASE

A Review of Background Research

Prepared by Helen Schietinger

Table of Contents

OVERVIEW

Definition

Outcomes research is an exciting new field of scientific investigation that has much to offer researchers, clinicians, and policy analysts addressing HIV/AIDS issues. Also called effectiveness research, it draws upon multiple disciplines to examine the long-range effectiveness of specific medical treatments and procedures for patients and for society. This developing discipline focuses on measuring and comparing a broad spectrum of endpoints rather than the immediate clinical impact that is the usual focus of clinical trials. The endpoints include the impacts on the person receiving the treatment, such as quality of life and absence or presence of side effects, as well as the impacts on society, such as the economic and social cost of the treatment. (Kasper, 92)

Outcomes research also addresses the rapidly changing priorities in clinical care and health resource allocation. The field has developed in part because traditional process-oriented health services research does not address the new need for medical decisions to be based on projected health and economic outcomes. Besides this, there is growing awareness that clinical effectiveness involves domains much broader than the biological impact of a treatment on an organism. While clinical trials in HIV/AIDS often focus on laboratory markers as measures of effectiveness, these markers may not encompass certain more subjective aspects of patient outcome. Carefully controlled clinical efficacy studies may have limited application in the real world of clinical care, where social, environmental and behavioral factors have a substantial impact on the prognosis of a given patient being treated with a specific intervention. Moreover, many common medical interventions have never been documented as superior to other interventions or as clinically effective at all. (Atlas, 1996)

The fields of epidemiology, statistics, economics, decision modeling, biomedical research, and behavioral and psychological research all contribute to this new discipline. The stakeholders concerned about the broad clinical effects of health interventions have grown to include not just researchers and clinicians, but also consumers, payors, and other health care providers. (Bayley, 95) The need to know about health outcomes has become important in making both clinical and policy (including resource allocation) decisions.

Scope of outcomes research

Outcomes research can be a complex long-range process that includes the collection and analysis of primary and secondary data from a variety of sources, the synthesis and dissemination of the findings, and finally an evaluation of how effectively the recommended changes are incorporated into clinical practice. Various methodologies are used during this process which, in its most intensive application, involves the following series of steps, applied to the treatment of a specific disease or condition:

1. ***Meta-analysis*** of existing literature;
2. ***Examination of existing data sets***, with analysis of relationships among health care services, patient outcomes, and resource use;
3. ***Development of health status measures***, and implementation of longitudinal studies using these measures;

4. ***Development of clinical guidelines*** based on the findings of all prior outcomes research;
5. ***Dissemination, implementation and evaluation*** of guidelines.

When the steps are examined in more detail, a range of methodological approaches can be identified, all of which have been refined and improved through their application to this process, particularly through focused federally-funded research (see next section).

Meta-analysis. The literature search and meta-analysis of prior research that is conducted regarding a specific disease or treatment identifies clinical or surrogate marker-based efficacy studies comparing various interventions, and attempts to use the compilation of data to enhance the knowledge gleaned from individual studies. The analysis also identifies gaps and weaknesses in the literature and recommends new areas of research.

Examination of existing data sets. Existing data sources, especially large administrative databases (e.g., indicators such as mortality and hospitalizations from the Medicare database), are analyzed to examine large-scale temporal, geographic, and economic trends not only in disease incidence and/or prevalence but also in clinical treatments and procedures. Often the baseline within a given population can be established using large administrative or survey databases. The development of new scientific methods, such as mathematical modeling and complex statistical analysis, has enhanced the capacity for manipulation of these large data sets, and the resulting observations and conclusions that are possible. More work needs to be done to address the limitations and enhance the use of these databases for research purposes.

Development and application of health status measures. Health status measures broaden the scope of outcome research beyond physiological indicators related to disease control or management, to include indicators of patient well-being such as functional capacity and quality-of-life, symptom relief, cost, and cost-benefit analysis that are related, ultimately, to the appropriateness of treatment decisions. (Maklan, 1994; Bayley, 95; Ware, 95) These health status measures are applied in longitudinal studies of health outcomes among patients receiving specific treatments, broadening the scope of clinical outcomes.

Table I lists outcome measures related to back pain. The list includes measures that some investigators would label "process indicators," illustrating again the wide range of elements that draw the attention in outcomes research. When ethical and practical, randomized prospective studies with controls are conducted to examine the effectiveness of clinical interventions in relation to health status measures. Through this prospective research, the extent of understanding about the impact of treatments and procedures is greatly enhanced.

Table I: Examples of Outcomes Measures Related to Back Pain

Table I: Examples of Outcomes Measures Related to Back Pain	
Physiologic	Spine or extremity range of motion Muscle EMG activity Spinal fluid endorphin levels Muscle strength, endurance
Anatomic	Solid fusion mass Disc height Vertebral displacement
Complications	Drug side effects New neurologic deficit Major infection Cardiopulmonary complications Dural tear
Physical examination	Neurological deficits Straight leg raising
Mortality	
Health-related quality of life: Symptoms	Pain duration Pain frequency
Health-related quality of life: Functional status	ADLs Psychological function Recreational activities Social function Health perceptions, general well-being
Health-related quality of life: Role function	Employment status Disability compensation Days of work absenteeism Days of limited activity
Costs, health care use	Medical care costs Compensation Imaging tests Need for first or repeat surgery Assistive devices, physical therapy
Satisfaction	With treatment With results Were expectations met?

(Deyo, Andersson, 1994, p. 2033S)

Development of clinical guidelines. Using the information that has been gathered through investigations of existing literature, large databases, and application of health status measures, the elements of optimal care are improved or identified, resulting in the development or revision of clinical guidelines. Sometimes the research has provided surprising results that have changed practice in clinical care. For example, the findings that back surgery with spinal fusion resulted in a higher complication rate than back surgery without spinal fusion and that the clinical outcomes were

no better resolved the controversy regarding this surgical procedure. (Deyo, 1993)

A new element in the development of clinical guidelines is the use of a consensus process that includes all stakeholders, including consumers and payors, rather than only clinicians or researchers. While there are examples of the development of guidelines through consensus in England and other developed countries (Conroy, 1995; Grimshaw, 1993), we will confine ourselves to examples in the United States.

Dissemination, implementation and evaluation. Dissemination and implementation of clinical guidelines are critical next steps that have often been overlooked in the past. The obvious result is that practitioners continue using out-dated clinical information and skills they received in their pre-service training. To prevent this, it is important to build into the process of guidelines development a plan for their dissemination as well as for the evaluation of whether and how they are being implemented. Only when guidelines implementation is evaluated can the feedback loop be effectively closed to ensure that clinical care actually changes in response to new knowledge. For this purpose, the tools of outcomes research have been applied to quality assurance in clinical settings. In addition, evaluation of guidelines implementation generates new hypotheses and identifies needs for further research, so that outcomes research is an ongoing cyclical process.

The Forum for Collaborative HIV Research convened a meeting earlier this year that identified a need to evaluate previously developed HIV guidelines for care. This has resulted in the current focus of concern with evaluating the implementation of the HIV guidelines for care.

Federal initiatives supporting outcomes research

In 1988, the Health Care Financing Administration (HCFA) began an effectiveness initiative to use the data from the Medicare systems of claims processing and peer review to monitor trends and to assess the effectiveness (meaning both clinical efficacy and appropriateness) of selected medical interventions. (Roper, 88) At the same time, a data resource center was developed to make the Medicare data available for use by private persons and organizations.

In 1989, Congress allocated $15 million for studies to develop and disseminate information on clinical practices that enhance patient outcomes. Several agencies in the Public Health Service assumed roles in this effort, including NIH, HRSA, and the National Center for Health Services Research and Health Care Technology Assessment. The Agency for Health Care Policy and Research initiated the Medical Treatment Effectiveness Program (MEDTEP) to promote patient outcomes research, facilitate development of clinical practice guidelines, improve databases for research purposes, and to disseminate research findings and clinical guidelines. (Clinton, 1991) Because of the substantive and methodological contributions of the treatment effectiveness program, its budget was doubled four years after its initial funding in 1990, under the auspices of the new Agency for Health Care Policy and Research (AHCPR).

In the first two years of the MEDTEP program, 11 Patient Outcomes Research Teams (PORTs) were funded, and additional PORTS were added later. Each PORT examined a specific health condition or procedure affecting large numbers of Americans for which optimal treatment was unclear, costs were

high, and data were available. (Maklan, 1994) These were large five-year projects with average annual budgets of one million dollars each. The areas of investigation of the PORTs included (Greene, 1994):

- Back Pain Outcome Assessment Team
- Consequences of Variation in Treatment for Acute Myocardial Infarction
- Variations in Cataract Management: Patient and Economic Outcomes;
- Assessing Therapies for Benign Prostatic Hypertrophy ad Localized Prostate Cancer
- Variations in the Management and Outcomes of Diabetes
- Outcome Assessment Program in Ischemic Heart Disease
- Analysis of Practices: Hip Fracture Repair and Osteoarthritis
- Outcome Assessment of Patients with Biliary Tract Disease
- Variations in Management of Childbirth and Patient Outcomes
- Assessment of the Variations and Outcomes of Pneumonia
- Secondary and Tertiary Prevention of Stroke
- Low Birthweight in Minority and High Risk Women
- Schizophrenia Patient Outcomes Research Team

Each PORT conducted a series of activities including systematic literature reviews and formal analyses; analyses of variations in practice patterns and patient outcomes; development of guidelines from the findings; and, dissemination of the guidelines and the evaluation of the guidelines implementation. The process was multi-disciplinary, multi-faceted, multi-method, and multi-site. (Maklan, 1994, p. JS14) As the first PORTs completed their work, a second series of PORTs, PORT-II, was initiated to continue investigations in areas of particularly fertile work.

In conjunction with the PORTs, representatives began meeting together in cross-cutting committees to identify and resolve shared methodological and conceptual problems related to the processes being utilized by the MEDTEP investigators. The six Inter-PORT Work Groups included:

- Literature review and meta-analysis;
- Use of claims data;
- Decision modeling;
- Outcomes assessment;
- Cost of care;
- Dissemination.

In analyzing the challenges and lessons learned about each issue across the PORTS, methodological advances have been made that have contributed to the improvement of the field of effectiveness research. (Maklan, 1994) PORTs use existing data sets and research to perform complex analyses of the relationships among health care services, patient outcomes and resource use. After identifying the elements of optimal care for the condition under study, PORTs develop clinical recommendations, disseminate the recommendations, and evaluate the changes in practice patterns and patient outcomes that may result from assimilation of the research findings. (Clinton, 1991)

METHODOLOGIES

Literature review, synthesis, and meta-analysis

Usually the initial step in outcomes research is to conduct a literature review and synthesis, and, when appropriate, meta-analysis, to establish current knowledge and practices regarding the treatment of the disease in question and to identify gaps in the knowledge base. A relatively new methodological discipline, meta-analysis is increasingly valued as an important component of medical research investigations that can achieve the objectives of assessing the evidence that the current assumptions are valid, testing specific hypotheses, and estimating treatment effects.

> *Meta-analysis: the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings*. (Glass, 1998)

Although there is wide variation in the techniques and strategies used to conduct the search and analysis, literature reviews and meta-analyses usually involve certain common elements. There must be a systematic, quantitative, and reproducible method for recording, integrating and summarizing the characteristics or results of independent research studies. (Powe, Turner, et al, 1994) A detailed description of the process employed assures that other researchers can assess the comprehensiveness of the review and the validity of the conclusions.

Meta-analysis usually includes identifying all related articles, selecting studies that meet certain pre-determined inclusion criteria (often only those studies whose subjects are randomized and have controls), aggregating the data from the studies, weighting the data where necessary to allow comparison across studies, generating hypotheses, conducting further statistical manipulation of the data, and developing conclusions.

A number of methodological questions remain unanswered about this new research technique. Some of the persistent problems that must be addressed include: how to adjust for the bias toward findings that exists in the published literature; how to adjust for the differential quality of published studies; how to adjust for heterogeneity of treatment effects; whether to include unpublished studies and studies from non-peer-reviewed journals; whether and how to blind reviewers of independent studies; whether and how to synthesize the results of studies without a control group, and how to synthesize longitudinal data. (DerSimonian, 1986; Powe, Turner, et al, 1994) Tools are being developed to enhance the processes of conducting comprehensive and efficient searches and analyses, but much further work is needed. However, one thing is certain: the process is time- and labor-intensive, and therefore costly. To generate a useful literature synthesis or meta-analysis, researchers must make an adequate investment in both time and labor to assure the findings are as replicable and as valid as possible.
That said, conducting a meta-analysis is definitely more cost-effective and less time-consuming than doing a study from scratch of the magnitude that is possible with a meta-analysis.

The process and some of the drawbacks of the methodology can be illustrated by a review of a meta-analysis of data related to surgery for lumbar spinal stenosis that was conducted in 1990-1. (Turner, Ersek et al, 1992)

> *A comprehensive search of the literature from 1966 to 1991 was conducted. Abstracts were reviewed by two clinicians, and all articles of possible relevance were retrieved. Further*

citations were identified from the bibliographies of retrieved articles. In all, 625 articles were selected. Inclusion and exclusion criteria were developed. Each article was read independently by two clinicians, and information extracted concerning the study methodology, preoperative patient characteristics, surgical methods, and patient outcomes. The data were recorded on standardized coding forms. The clinicians then discussed and reached consensus on all differences between their decisions about each item of data. When agreement could not be reached, a third rater arbitrated. Finally, 74 articles were included in the analysis.

The investigators developed sets of outcome rating criteria, which they applied to each article. An overall rating system of good-to-excellent, fair, and poor outcomes was used, and for articles with sufficient information, ratings were developed for discrete variables: back pain, leg pain, job functioning, and functional disability. Descriptive data analyses were performed to examine the distributions of the variables under study, including patient and procedural characteristics, and surgical outcomes. In addition, the investigators performed other analyses to determine whether overall outcome varied significantly across studies, and to identify any potential predictors of outcome.

The authors were able to make few conclusions; perhaps their words are most expressive: "The most definitive finding of this literature synthesis was the poor scientific quality of the literature. Major deficits in study design, analysis and reporting were the rule, rather than the exception. Because of these flaws, we could not conduct the meta-analysis we had intended. A meta-analysis to address questions of efficacy of various therapies requires the existence of comparative trials (preferably randomized), optimally reporting means and standard deviations of patients in each group on the same outcome measures. A meta-analysis to examine predictors of outcome requires multiple studies reporting statistical associations between predictor measures and the same outcome measures. Neither requirement was present in the existing literature." (Turner, Ersek, 1992, P. 5)

A more successful meta-analysis was conducted regarding vertical HIV transmission and mode of delivery:

To evaluate the relation between elective cesarean section (C-section) and vertical transmission of HIV, a meta-analysis was performed using data in individual patients from 15 prospective cohort studies. North American and European studies of at least 100 mother-child pairs were included. Uniform definitions of modes of delivery were used. Elective C-sections were defined as those performed before onset of labor and rupture of membranes. Multivariate logistic regression analysis was used to adjust for other factors known to be associated with vertical transmission.

The primary analysis included data on 8533 mother-child pairs. After adjustment for receipt of antiretroviral therapy (ART), maternal stage of disease, and infant birth weight, the likelihood of vertical transmission of HIV-1 was decreased by approximately 50 percent with elective C-section, compared with other modes of delivery. The results were similar when the study population was limited to those with rupture of membranes shortly before delivery. The likelihood of transmission was reduced by approximately 87 percent with

both elective C-section and receipt of ART during the prenatal, intrapartum, and neonatal periods, compared with other modes of delivery and the absence of ART. Among mother-child pairs receiving ART during the prenatal, intrapartum, and neonatal periods, rates of vertical transmission were 2 percent among 196 mothers who underwent elective C-section and 7.3 percent among the 1255 mothers with other modes of delivery. The results of this meta-analysis suggest that elective C-section reduces the risk of transmission of HIV from mother to child independently of the effects of treatment with zidovudine. (IPHG, 1999)

Analysis using large administrative and survey databases

Outcomes research often uses large administrative and survey databases in the study of medical treatments, in order to obtain statistical power. The development of the ICD-9-CM has enhanced the utility of insurance claims in the systematic study of diseases in large populations. Analysis of outcomes using these large data sets complements other methods of investigation, such as controlled studies on smaller, randomized groups of people.

Statistical analysis of the data elements in large data sets can identify associations between interventions and outcome measures that can then lead to further research and analysis. In addition, the manipulation of large data sets can be used to establish the baseline incidence or prevalence of a condition or a treatment in the community, to identify geographic variations and time trends in medical utilization, or to identify rare events that would only be seen in such a large sample or population. (Deyo, Taylor, 1994) The large data sets are used to define specific populations to which decision modeling can be applied, or to which smaller experimental samples can be compared. Smaller random samples are obtained from large databases for controlled experiments that can then be generalized to the population represented by the larger database.

Administrative databases can be used in a number of other ways to strengthen clinical research. The data can assist in estimating probabilities of certain outcomes to determine sample size requirements of smaller studies. The databases can sometimes provide a denominator for more explicit research. They can also provide the basis for decisions in designing a study, such as stratification of the data and inclusion and exclusion criteria. Finally, while small randomized controlled studies can be designed with high internal validity, they often are not highly generalizable to larger populations. Use of large databases complements this limitation by providing comparative information on a larger and more representative population.

The development of sophisticated statistical techniques, as well as the development of powerful software and hardware with the capacity to conduct the needed calculations, has facilitated the analysis of larger and larger data sets. Along with this growing technology, databases have emerged that contain very large data sets derived from national surveys and from administrative functions such as processing of insurance claims and recording of archival information (births, deaths, incidence of cancer).

Administrative databases

Administrative databases can be found in the public as well as the private sector. They include data

sets as disparate as statewide hospital discharge registries, cancer registries, death records, insurance claims records, and workers compensation records. The databases developed for administrative purposes are secondary data sources that offer a number of advantages over the primary data collected for investigative purposes. The size of the databases provides the statistical power necessary to study the multiple variables that contribute to a single health event or outcome. Because the data have already been collected, they tend to provide a less expensive and more rapid means of answering some questions than conducting original research to obtain primary data. Moreover, the data often represent all members of specific populations, such as all hospitalized people in a state.

Use of administrative databases also has a number of limitations. The data elements are pre-determined and not under the control of the investigator. It is often unclear whether the absence of certain outcomes is a function of reality or of the database design. The information is usually confined to health care utilization to the exclusion of clinical observations and outcomes. (Deyo, Taylor, 1994) The processes of care, such as the specific services provided during a hospital stay, are usually not available through these records. Longitudinal tracking of data may not be possible as data elements change over time (e.g., between the ICD-9 and ICD-10; as laboratory tests are developed or refined). In addition, the quality of the data is dependent upon the quality of the data collection process (e.g., there may be missed entrees, different criteria being used, etc), which may be somewhat unreliable, particularly in comparison to the high quality that is possible in carefully controlled clinical trials.

Perhaps the largest administrative databases representing an entire population are the Medicare claims records, maintained by the Health Care Financing Administration (HCFA), which include almost all people over the age of 64. They are one of the largest sources of information on health care utilization in the United States. The data are collected within a single health care financing system, under fairly uniform rules. Data include use and costs of all covered services reimbursed by Medicare under fee-for-service. (Lave, 1994)

An advantage of the Medicare databases is that they provide accurate longitudinal data on nearly the entire elderly population, from the date of eligibility until death. This is unlike most other medical claims data, such as private insurance companies, in which beneficiaries change companies frequently, or Medicaid, in which many beneficiaries are eligible for only short periods of time.

A disadvantage of Medicare data is that it does not include the people who do not have a long enough work history to be eligible, such as the extremely poor or non-working disabled. In addition, data are only collected on beneficiaries in the traditional fee-for-service Medicare program, not on those who are enrolled in managed care. The proportion of beneficiaries enrolled in Medicare managed care is increasing each year, leaving fewer beneficiaries in the Medicare databases. Because there may be differences between beneficiaries who select fee-for-service and managed care Medicare, this reduces the representativeness of Medicare claims data.

Another type of administrative database is the hospital discharge registry, which is usually maintained at the state level. Discharge registries often contain more detail about a specific hospital event, sometimes even providing time sequenced diagnoses and treatments, which allows researchers to connect conditions with events (e.g., the fact that an infection is dated as being diagnosed after the date of a surgery may enable a researcher to identify it as nosocomial, or a surgical complication, which would not be possible using Medicare claims data). However, even here, quality of data may be

limited due to coding errors.

National surveys

A number of large national surveys are conducted to collect information regularly on national representative samples of individuals or health care events. The federal government funds many of these surveys. (Deyo, Taylor, 1994) The National Ambulatory Medical Care Survey is an annual survey that obtains information on a representative sample of patient visits to office-based physicians. The National Hospital Discharge Survey obtains information annually on a representative sample of all hospital discharges.

A number of national surveys monitor the health status of Americans. The National Health and Nutrition Examination Survey (NHANES) conducts detailed health histories, physical exams, and laboratory tests on a representative sample of the U.S. population every five years. It is conducted by the National Center for Health Statistics (NCHS), which is part of the Centers for Disease Control (CDC).

Development and use of health status measures

Outcomes research has widened the scope of clinical research beyond immediate physiological and anatomic outcomes of treatment to include functional outcomes such as range of motion and muscle strength, patient-reported outcomes such as pain and emotional well-being, and clinical outcomes such as number of emergency room visits or episodes of pneumonia, and health care costs, including not only costs to the payor and social costs, but also costs to the consumer. Measures have been developed for domains such as functional status, symptoms and symptom relief, role function (work capacity), and satisfaction with treatment. (Deyo, Andersson, 1994; Atlas, 1996) Improvements in quality of life and well being have been integrated into the evaluation of health care services for the first time. The most dramatic change in emphasis has been the growing importance placed on taking into consideration the patient s goals for treatment in measuring the outcomes of treatment. (Ware, 1992)

While some health status measures can be measured in a clinical setting using objective tools (blood analysis, scale, dynamometer), many of them can only be measured by collecting patient reported data and developing scales for comparison to norms in a larger population. Numerous instruments for health status have been developed and, according to R Deyo, these newer questionnaires are valid, reproducible, and sensitive to clinically important changes over time. In fact, the questionnaires are sometimes more accurate in predicting outcomes such as death, disability, and health care expenditures than are high-tech laboratory and imaging tests, and they also capture what is most important for patients. He recommends wider use of modern instruments for measuring health-related quality of life, adding patient-oriented outcome measures to research that would otherwise only focus on physiologic or anatomic variables and using only a few widely accepted instruments for each disease to facilitate comparisons across studies, including meta-analysis. (Deyo, Andersson, 1994, p. 2033S)

A project that has contributed greatly to the field of health status measures is the Medical Outcomes

Study (MOS). The MOS grew out of the 1983 Health Insurance Experiment (HIE) whose goal was to construct scales to measure a broad array of functioning and well-being concepts. (Ware, 1992) The HIE demonstrated that scales constructed from self-administered surveys could be reliable and valid tools for assessing changes in health status. The MOS then developed the first comprehensive array of generic functioning and well-being measures that can be used in diverse populations and health care settings, and also can be used with sick and well populations. The MOS has developed health status measures that not only meet the traditional psychometric standards of reliability, validity and precision, but are also efficient and practical: they have achieved the goal of reducing respondent burden, making them possible to use within health care settings as well as research. For example, in an old survey, 25 items were required to define 7 levels of physical functioning, whereas the MOS Physical Functioning Scale uses 10 items to define 20 levels of functioning. (Fowler, Cleary, 1994) The quality of instruments to measure health status has been improved by the intensive work of the MOS project. (Ware, 1995; Wu, 1997)

Using health status measures in chronically ill populations such as people with HIV/AIDS may pose a problem related to the focus of many of the instruments on functional ability. Whereas increase in function is generally considered a favorable outcome associated with a higher health status or quality of life, for some conditions the opposite may be true. A person may benefit from becoming wheelchair-dependent, thus increasing her/his mobility. A person with chronic pain may experience less functional pain by being more functionally dependent. (Jette, 1980) Thus, it is important to assure that the outcomes identified as favorable for the population being studied are actually being measured as favorable by the health status instrument being used.

The findings that result from use of health status measures may be affected not only by the population or disease to which they are applied, but also by the design of studies. For example, prospective and retrospective assessments of treatment outcomes have been found not to yield the same results (Aseltine, 1995). Retrospective assessment yields a higher estimate of the benefit of treatment than a prospective assessment.

An interesting finding that might be relevant to study of health outcomes in HIV/AIDS treatment is that patients tend to adapt to adverse outcomes. In a study of the effect of radical prostatectomy for prostate cancer on patient quality of life, it was found that patient satisfaction with the treatment and its results were not necessarily equivalent to what had been identified as a successful outcome. There was no association between respondents having adverse outcomes or surgical complications (incontinence and sexual dysfunction) and lower scores on measures of quality of life, how they felt about their surgical treatment, or whether they would choose the surgery again. This reinforces the importance of individualized decision making regarding whether to undergo treatment with possible adverse outcomes or side effects. (Fowler, Barry, 1995)

The use of health status measures may require compromises related to the objectives and methods of the study. Although data collected for pure research purposes may be accessible using interviewers who have unlimited access to respondents, most situations involve constraints of some sort (e.g., lack of privacy or time limits for the interview). Perhaps the most stringent requirements are found when collection of health status measures is to be integrated into a health care system for assessment of health outcomes for quality assessment or health care management purposes. Under these clinical circumstances, the instruments must be accepted by the providers and not pose to great of an added

burden to their clinical responsibilities; the tools must be standardized; there must be comparison data sets available; and, the data collection must be inexpensive (for example, scanners might be used for inputting data from paper forms). (Lansky, 1992)

Experimental designs used in outcomes research

Experimental versus Observational studies

Outcomes research can be conducted using both experimental and observational methodologies. The experimental approach (the randomized controlled trial - RCT), is, of course, the gold standard, because it allows the most control over the variables, and thus achieves the most definitive conclusions. It is designed to answer questions regarding the efficacy of an intervention. Patients are randomized to treatment and control groups, or to one of multiple control groups, and provided the treatment according to a specific protocol. Then their clinical status is measured in a standardized way to compare the clinical or surrogate marker-based outcome of the randomized treatment.

The fact that patients have been randomized to specific treatment (or control) groups lends credence to the assertion that the difference in outcomes is actually a result of the treatment as opposed to other variables. After randomization, non-treatment variables may be assumed to affect all of the patients equally, and individual differences are, with high probability, distributed equally between the groups by randomization. The number of subjects is selected to assure sufficient statistical power to produce a significant and relevant difference between treatment and control groups if one exists.

The criteria for attributing cause (the exposure caused the disease; the treatment caused the favorable clinical outcome) include (Schlesselman, 1982):

- temporal sequence (the possible exposure occurred prior to onset of the disease);
- consistency (the association of the two factors is observed repeatedly);
- strength of association (the relative risk is high);
- biological gradient (dose-response curve demonstrates that a stronger exposure is associated with more virulent disease);
- specificity of effect (with the exposure, the condition appears; without the exposure, the condition does not appear);
- collateral evidence (other findings point in the same direction);
- biological plausibility (there is a biological explanation is plausible).

Rigorously designed RCTs have been tightly controlled in an effort to eliminate all effects except the experimental treatment (i.e., assure internal validity), in order to demonstrate a causal relationship. However, because of the rigid control, they often do not represent actual clinical reality, in which multiple other effects influence whether and how treatments are administered. In addition, unless subjects have also been randomly selected from the larger population, findings are not necessarily generalizable beyond the study participants, reducing the study's external validity. Therefore, observational methodologies often provide the most relevant results in outcomes research. Generally, RCTs are used to determine the efficacy of a treatment under experimental conditions and observational designs are then used to determine the effectiveness and efficiency of the treatment when

applied under usual clinical conditions.

Observational studies include cohort studies, case-control studies, and cross-sectional studies. Cross-sectional studies obtain information about a single point in time. For example, a study might identify associations between subject characteristics and health events or health utilization. There are two kinds of longitudinal observational studies: cohort studies and case control studies.

Alternatives to RCTs are needed in situations in which (Armenian, 1998, p. 136):

- The outcome being studied is rare;
- To conduct a randomized study would be unethical;
- The disease being studied or prevented has a long latency period;
- The intervention being studied is available and being used widely in the community;
- The outcome is the potential side-effects of a treatment;
- An impact assessment of the intervention is needed in the community after its efficacy has already been established.

A number of important questions can be answered by examining information that is already being collected or by collecting information on a pre-existing cohort without applying an intentional intervention. Observations can be obtained through chart reviews, manipulation of administrative or survey databases, collection of data from individuals by surveys or interviews, or collection of data by observing a situation in the field (for example, interactions in an emergency room).

A strength of observational studies is that they can be less artificial than RCTs, in which subjects are randomly assigned to treatment arms. In addition, observational studies can be less costly and time-consuming than randomized controlled studies. While observational studies can be very useful, they also have limitations, the largest of which is that the lack of random assignment of subjects to treatment may reduce the ability of investigators to adequately control for the effects of intervening variables.

The trade-off between traditional clinical research and outcomes research can in some ways be seen as the tension between achieving high internal validity and high external validity. The objective of the former is to achieve the highest possible internal validity, sometimes at the expense of external validity, and the objective of the latter is to achieve the highest possible external validity, which may result in compromise of internal validity to some extent. However, as we shall see, it is possible to build randomization into outcomes research in order to also maximize internal validity.

Randomized controlled trials

> *While outcomes analyses of retrospectively and prospectively assembled cohorts of patients exposed to different intensities of early detection efforts and treatment strategies must continue, these comparisons have inherent biases. Only randomization can provide reasonable assurance that treatment groups are balanced in terms of known and unknown confounders, and that the comparisons of the resulting outcomes are fair and unbiased. (Barry, 1994, p. 1903)*

While RCTs often have limited generalizability because they are conducted in a tightly controlled

manner, the value of their definitive findings has justified some large-scale studies in order to achieve greater generalizability. In an editorial about a component of the Prostate, Lung, Colon, and Ovarian Cancer Screening Project, M J Barry argues for a large-scale RCT:

> *A randomized controlled study will address the question of whether it is beneficial to conduct screening on elderly men for prostate cancer for early detection using digital rectal examination and the prostate specific antigen (PSA) blood test, both methods that have poor specificity: 15% to 40% of those screened have suspicious results (depending on their age), leading to multiple prostatic biopsies, most of which are negative. In order to justify a screening protocol with such a low specificity, it must have a demonstrated mortality benefit. The study designed to answer the question identifies the outcome indicator as mortality from prostate cancer.*
>
> *The Prostate, Lung, Colon, Ovarian Cancer Screening Project trial randomized 74,000 men ages 60 to 74 to an aggressive screening protocol or routine care. When the study is complete, the mortality from prostate cancer in the two groups will be compared. The trial has a 90% power to detect a 20% reduction in prostate cancer mortality between the groups. This large-scale longitudinal study, examining the ultimate clinical outcome of cancer screening, is the only way to answer this question, and Barry argues that the answer will be invaluable for clinical and health policy decisions. (Barry, 1994)*

Obviously, this study is both costly to conduct and is taking a long time to carry out. Sometimes alternative strategies can produce acceptable results.

Cohort studies

A cohort study identifies two groups of people, one that received an intervention or exposure and one that did not, and follows them longitudinally to measure the difference in the proportion of each group that have the hypothesized outcome, usually a disease or condition. For example, had the above prostate screening study been a cohort study, the subjects would not have been randomized into aggressive screening and routine care groups. Rather, groups receiving each would be identified (perhaps by conducting an aggressive screening program in one clinic and using subjects from the community who receive routine care as a comparison group).

The lack of randomization reduces the certainty with which the outcome can be attributed to the intervention or exposure. However, if historical data are used, a cohort study can produce faster results, and can be conducted at less expense. Moreover, if a cohort is selected from a defined population, multiple exposures can be studied, and thus might be more comprehensive in its findings.

Case-control studies

A case-control study enables the investigator to ascertain the association between an outcome and its determinants. Here, a group of people who have the disease or condition (the cases) are identified, and matched with a group of people who do not have the disease or condition (the controls). The study is usually retrospective in nature: their past exposure to the hypothesized causal agent or

treatment is then measured and analyzed for possible differences between the two groups.

The case-control study does not provide an incidence rate, but from such a study it is possible to calculate an odds ratio as a means of quantifying the relative risk (for low-incidence diseases), as well as the relative importance of various suspected determinants of the outcome, as well as any potential interaction between these factors. (Armenian, 1998)

The following is an example of a case-control study:

> *The case-control method is demonstrated by a study that was conducted on women with breast cancer. Ninety-seven women with confirmed incident cancer were selected from a larger cohort study of osteoporotic fractures in which 9,704 white women over 64 years of age were being followed prospectively at four clinical centers across the U.S. The 97 case-subjects included all incident breast cancers in women who were not receiving estrogen-replacement therapy at baseline. Case-controls were 247 women selected at random from the same study who denied a history of breast cancer and did not report use of estrogen at baseline. Their sex-steroid hormone concentrations were measured and the results analyzed.*
>
> *The relative risk for breast cancer in women with the highest concentration of bioavailable estradiol was 3.6 (95% CI, 1.3 to 10.0) compared with women with the lowest concentration. The risk for breast cancer in women with the highest concentration of free testosterone compared with those with the lowest concentration was 3.3 (95% CI, 1.1 to 10.3). The estimated incidence of breast cancer per 1000 person-years was 0.4 in women with the lowest levels of bioavailable estradiol and free testosterone compared with 6.5 in women with the highest concentrations of these hormones. Traditional risk factors for breast cancer, such as family history of breast cancer, smoking, and obesity, were similar in case-patients and controls. (Cauley, 1999)*

The fact that this case-control study was nested in a much larger study provided the opportunity for the controls to be selected randomly from the same population as the cases. Thus, although the population was recruited from four clinical centers, which limits the generalizability of the study, the comparison group can be used without question to determine whether there is an association between the level of sex-steroid hormones and the incidence of breast cancer. In this population, while there may be interactions among the variables to be considered that influence the strength of the associations between each variable and the outcome, the differences between the two groups remains clear.

Cross-sectional studies

A cross-sectional study is one in which the exposure and the outcome are measured simultaneously for each subject. Measurement of the outcome, usually a disease or condition, identifies the number of cases at one point in time in a given population, which is the prevalent cases. A cross sectional approach can also be used to compare two points in time, but at each point, the subjects are different. Thus, rather than subjects being followed longitudinally, as in a cohort study, or studied retrospectively, as is usually done in a case-control study, two cross-sectional snapshots are taken that can then be compared. The difference between this and a case-control study is that rather than

selecting cases with a disease and intentionally identifying matched controls with which to compare them, it is the population that is selected, after which exposure and disease are determined, and only then is the population stratified as to either exposure or disease. Because the exposure and disease are measured simultaneously, it is usually not possible to identify a causal relationship between exposure and disease.

Shortcomings of current clinical research

Much of the clinical research reported in the medical literature, whether it consists of RCTs, cohort studies, case-control studies or cross-sectional studies, is not scientifically sound. This was demonstrated over and over again in literature reviews in specific areas of medical care for meta-analyses. (Deyo, 1993; Powe, Tielsch, 1994; Hoffman, 1991) Reviewers found that much of the reported research lacks rigor, leaving much room for doubt about the certainty of the findings. Common characteristics of studies reported in the medical literature include:

< Lack of a control or comparison group: The subjects often constitute a convenience sample with no inclusion or exclusion criteria other than their availability. Without a control or comparison group, it is not possible to attribute the outcomes to the specific treatments.

< Lack of generalizability: Often the study group is not described in sufficient detail to enable the reader to make an informed guess as to whether the group is similar to other populations. Thus, the study is not generalizable beyond the treatment group or treatment site.

< No standardization of data recording: Studies such as chart reviews depend on existing data that is often not standardized in how it is recorded. This can result in wide variation in data elements and follow-up intervals. The data about individual patients are thus difficult or impossible to compare. Sometimes the demographic or other characteristics that might have contributed to differences in patient outcomes are not available, leaving it unclear as to whether the investigators took these variables into consideration in their analysis.

< Lack of objectivity in measurement: Often, particularly in chart reviews, not only was measurement not standardized, but the clinicians made their own observations. This is hardly an unbiased assessment. Moreover, there may have been as many clinicians as patients, providing a different subjective observation for each subject.

< No accounting for loss to follow-up: Articles commonly provide no explanation for attrition over time or analysis of the difference between subjects who complete the study and those who are lost to follow-up, although there may be real differences between the two groups. (For example, all patients for whom the treatment was unsuccessful may have left the study, giving the impression of a 100% success rate, when the actual rate was much lower.)

< Lack of consumer input: An important weakness of traditional clinical research is the absence of a consumer perspective, which leads not only to inappropriate or incomplete hypotheses being generated, but unrealistic study design, lack of congruence of measures, failure to address key variables, and distorted analysis.

Descriptions of literature reviews used for meta-analyses of back pain research for the back pain

PORT demonstrate many of the common weaknesses of clinical research:

> *In a literature synthesis of 47 articles about patient outcomes after lumbar spinal fusion, outcomes were assessed independently of the operating surgeon in only seven of the studies. There were no randomized trials, although 4 non-randomized studies compared surgery with and without fusion. Most articles did not report potentially important prognostic variables, and lack of comparison groups made it impossible to determine the extent to which patient status at follow-up was due to the fusion versus other factors. (Deyo, 1993)*
>
> *In a literature synthesis of 74 articles regarding surgery for lumbar spinal stenosis, studies primarily involved short-term follow-ups of patients, which may have contributed to misleadingly favorable results. Many of the articles contained inadequate description of the surgery and of patient characteristics, including variables such as complications and reoperations, making the case-series much less useful. (Turner, Ersek, 1992)*
>
> *According to the authors, the literature does not definitively answer the questions we posed, and the statement made over a decade ago that the literature on spinal fusion is totally inadequate remains true today..... Randomized controlled trials are needed to compare the long-term risks and benefits of specific fusion procedures vs other surgical and nonsurgical interventions for patients with specific, rigorously defined lumbar disorders. (Turner, Ersek, August 19, 1992, p. 910)*

What good outcomes research has to offer

As researchers identify the shortcomings of past research they can improve new research so that questions about efficacy and effectiveness are answered with valid, reliable, replicable research. The RCT is the only method available to definitively answer important questions about the efficacy of many medical treatments. Outcomes research can build RCTs into long-term research projects. Moreover, the standardization of what information is obtained and how it is collected can also be made more rigorous. (For example, data collection can be made more systematic by blinding reviewers and establishing strict protocols.) Finally, the outcomes that are selected can include multiple indicators related to health status, long-range clinical status, and cost, in addition to traditionally measured short-term clinical status. Thus, rather than just measuring efficacy of a treatment (treatment under ideal conditions), the researchers also measure effectiveness (treatment in the context of intervening variables such as adherence), while maintaining the same high standards of clinical research.

Decision modeling

Decision modeling, or decision analysis, has become an essential tool of outcomes research. Used to predict outcomes based on an analysis of the complex variables involved in a situation, decision modeling has a range of applications, from directing clinical decisions to projecting the cost-effectiveness of introducing a new procedure or technology into a health care system. The technique involves the development of complex mathematical models using clinical and operational variables such as cost and morbidity and applying values taken from current or projected practice. Some of the

various modeling strategies include Markov models, decision trees, and simulations.

Decision modeling can be applied to clinical guidelines development, policy analysis, program development, and various cost-benefit and cost-effectiveness analyses leading to resource allocation decisions. The PORTS applied decision modeling to clinical care in a number of ways, and at a number of steps. They used decision modeling to help define the optimal path to the desired outcome, through a vast number of patient characteristics and treatment options, and modified by probabilities, patient utilities, risks, and costs. According to Maklan, et al, These complex models, which may incorporate data from literature reviews, claims data analysis, chart abstraction, and surveys, are designed to predict various outcomes and/or costs for patients with the same disease or clinical problem, but who present with different characteristics and who undergo different management strategies. (Maklan, 1994, p. JS18)

Process versus outcomes research

Process research plays an important role in assessing outcomes. Process research assesses how the intervention is actually practiced, while outcomes research assesses what outcomes are achieved. Even if analysis of the outcome demonstrates that the outcome was achieved, it is also important to demonstrate that the intervention was implemented according to the prescribed protocol, in order to validate that the prescribed intervention was actually what achieved the outcome. If the intervention is complex, analysis must also account for variation in implementation of the intervention.

Process evaluation compares what actually happened during the implementation of an intervention with what was planned. A judgement is made according to existing standards of acceptability for each dimension. The evaluation either applies standards and a level of quality based on the literature or is derived by professional consensus. According to Windsor et al, (Windsor, 1984, p 89) process evaluation usually:

< applies nonexperimental designs;
< assesses operating procedures;
< examines structure and process;
< conducts observational analyses;
< performs qualitative observations;
< monitors program effort or activity;
< reviews or audits data systems and records; and,
< employs formative evaluation methods.

Measures of process also include quantifiable operational measures such as:

< number served;
< proportion of target populations served;
< service units;
< program cost;
< cost per identified case;
< proportion of identified cases brought into care (Gordis, 1996).

Measures of outcome, on the other hand, include assessments of effectiveness such as:

- decrease in mortality in the population;
- decrease in case-fatality rate;
- increase in proportion of cases treated earlier in disease process;
- decrease in rate of complications resulting from treatment or disease;
- decrease in recurrences/metastases;
- increase in quality of life of persons served (Gordis, 1996).

Process research is a critical component of outcomes research because it can never be assumed that the prescribed protocol was implemented as planned. If the intervention does not achieve the desired results, it may be that it was not implemented as planned, and only process evaluation will clarify whether this is so.

In health services evaluation, process evaluation can evaluate what a program has done for patients, how frequently and how well it was done, under what circumstances it was successfully done, and what should have been done. It does not, on the other hand, determine whether the patient s condition has improved.

Clinical Guidelines

Development

The development of clinical guidelines involves a series of activities that all must be completed in order for the ultimate objective, improvement of clinical care, to be achieved. The activities are: development, dissemination, implementation, and evaluation. The inputs include primary and secondary data and all the methodologic tools that have been discussed thus far: literature review and meta-analysis, outcomes research using health status measures and other instruments, use of administrative and survey databases, and decision modeling.

The development of guidelines should be a formal group process involving a multidisciplinary team of not just clinicians but experts from clinical research, epidemiology, psychometrics, and mathematics, but also representatives of all the stakeholder groups, including consumers and health care payors (health care industry). (Bozzette and Asche, 1995)

Dissemination

Once the guidelines have been developed, successful dissemination targets a range of audiences including consumers, health care providers, researchers, policy makers, payor representatives from the health care industry, and journalists. (Goldberg, 1994) The media used for dissemination include print media (direct mail, technical journals, health journals, the popular press, and newspapers) as well as electronic media (radio, television, and the internet).

Dissemination must include strategies that foster assimilation, acceptance, adoption and use of the guidelines, ultimately leading to behavior change on the part of the clinicians and consumers.

Historically, guidelines dissemination has not been found to produce behavior change among clinicians, but there is new interest in assuring that this occurs and that a successful outcome is demonstrated. (Horowitz, 1996) A wide range of strategies has been used, including some newly developed ones that have been implemented and evaluated in the processes of the evaluation of guidelines conducted by the PORTs. The following are dissemination strategies that were used in the PORTs (Goldberg, 1994):

Patient-targeted strategies:
< mass information;
< community-based health promotion;
< interaction with practitioner (shared decision-making procedures);

Practitioner-targeted educational strategies:
< curriculum development;
< continuing education;
< opinion leaders;
< academic detailing;
< accreditation procedures;

Practitioner-targeted administrative strategies:
< medical audits;
< feedback;
< peer review;
< reminder systems (computerized and manual).

Academic detailing (AD), or public interest detailing, one of the practitioner-targeted strategies listed above, is a specific QA strategy adopted from pharmaceutical industry strategies for marketing products to clinicians. It is an educational outreach approach to imparting information to practitioners through a one-on-one approach in which the trainer (academic detailer) targets specific categories of practitioners and their opinion leaders. They work directly with individuals in the clinical setting, repeating the essential message, and using positive feedback and reinforcement as follow-up. (Horowitz, 1996)

The need for quality assurance (QA) is frequently the factor that stimulates the dissemination of clinical guidelines within a medical setting. QA involves activities and programs to assure the quality of care in a defined medical setting or program, emphasizing the structure of care, the process of care, and the outcomes of care. Two innovative forms of quality assurance programs are Continuous Quality Improvement (CQI) and Total Quality Management (TQM). (Horowitz, 1996) The essential ingredient is the change of clinical behavior through monitoring and feedback methods that involve the clinicians themselves.

Evaluation

Evaluating the implementation of clinical guidelines should take a variety of factors into consideration to clearly demonstrate that changes in clinical practice are a result of the guidelines, rather than other circumstances. Not only should the outcomes of clinical care be measured, but also the processes of

clinical care. This will enable the investigators to know whether the outcomes that are observed are in fact a result of care that is in accordance with the clinical guidelines. A range of indicators is generally needed to assess that the elements of care are being delivered and link the delivery of care to the targeted outcomes. The outcomes of concern have expanded in the past two decades and now include indicators of patient well-being and functional ability. (Ware, 1992)

In assessing whether standards of care are being applied, inferring a causal relationship is much more challenging than it might seem at first. It may at times be simpler to demonstrate a negative correlation if guidelines are not followed and there is a negative outcome. In either case, multiple factors contribute to outcomes, even when they are limited to easily quantifiable measures rather than measures of health status dependent upon survey instruments.

The challenge of evaluating the dissemination of guidelines and the resultant behavior change by clinicians in a clinical setting is demonstrated in a study of physician compliance with clinical practice guidelines:

> In measuring and attempting to improve physician compliance with a length-of-stay guideline, physician refusal accounts for a small percentage (16%) of noncompliance. Implementation issues (e.g., classification as high-risk when actually low-risk), health care system inefficiency (e.g., delayed discharge caused by wait for tests or nursing home placement), and severity of illness (e.g., change in clinical status because of a complication or unstable co-morbid condition) were the predominant reasons why physicians did not comply with guidelines. Our study further supports the principle that clinical practice guidelines should complement rather than be a substitute for physician judgment. (Ellrodt, 1995)

The Inter-PORT Dissemination Working Group examined the amount of rigor that needs to be applied to evaluating dissemination of guidelines. A critical question was, should potential impact be compromised (e.g., by limiting the number of interventions adopted) to satisfy the rigor that is necessary to prove whether efforts were successful in changing behavior? The work group agreed that such compromises should be limited, and that at a minimum, quasi-experimental designs should be employed wherever possible. (Goldberg, 1994) Thus, it was recognized that the issues of clinical guideline evaluation in the context of the real world of health care services are complex and multidimensional.

APPLICATION TO HIV/AIDS

Methodologies used in HIV/AIDS outcomes research

The research related to HIV/AIDS clinical care resulting in the development of clinical guidelines has been both more dynamic and more chaotic than the processes utilized in the Medical Treatment Effectiveness Program that produced the PORTs. The slower paced evolution of clinical practice related to the chronic and disabling diseases studied by the PORTS provided neater scenarios for organized, step-wise progression of outcomes research: literature search and meta-analysis, outcomes research using all the tools at hand, including large administrative databases, health status measures,

and decision modeling, development and dissemination of clinical guidelines, and finally, evaluation to determine how well the clinical guidelines are being implemented and whether they are achieving the desired outcomes.

HIV/AIDS research has been aggressive since the beginning of the epidemic, and has applied most of the tools of the new discipline of outcomes research. The tool least used has been meta-analysis, because the literature becomes obsolete practically before articles go to press. An example of primary meta-analysis in HIV research was described in the Methodologies section of this paper.

Large administrative and survey databases

While numerous administrative databases have been tapped for use in HIV research, the variety of diagnoses that can be related to HIV disease has made it challenging to utilize hospital records or insurance claims data to study trends in utilization patterns. While there have been states such as California in which Medicaid files have provided rich data for the study of HIV utilization and cost, national data have not been as useful. In the early 1990s, HCFA commissioned the development of an algorithm to identify AIDS cases in the Medicare database, and when the study was complete, the estimate was a wide range of cases, between 10,565 and 22,927 cases. (Thornton, 1997) Instead of this broad estimate, actuarial estimates are often used in federal estimates of the number of people with AIDS or HIV who are Medicare or Medicaid beneficiaries.

In addition to use of pre-existing databases to study HIV/AIDS, HIV-specific databases have been developed. The CDC AIDS surveillance system, the first large database to be used in HIV research, has enabled epidemiologists to track the morbidity (an AIDS diagnosis) and mortality of the disease from the very start. A national survey of the health status and health care utilization of people with AIDS (AIDS Cost and Services Utilization Survey) was conducted from March 1991 to November 1992. Although not a probability sample, it was an attempt to collect demographic and health care cost and utilization information on a wide range of people with AIDS being cared for in a variety of health care settings. (Mohr, 1994) A similar national survey of the health status of people with HIV in care, entitled the HIV Care and Service Utilization Survey (HCSUS), is currently underway. Federal agencies and private foundations fund it, with the Agency for Health Care Policy and Research (AHCPR) as the lead agency. Efforts have been made to obtain as representative a sample as possible in this survey. A wide range of information has been collected by self-report from the respondents, and chart reviews have been conducted on a subsample of those surveyed. A wide range of data is being collected regarding the demographics and health status of people with HIV who are in care and the utilization and cost of care.

There is growing interest in developing or refining large databases to collect more detailed and useful clinical information both for quality assurance and research purposes. (Deyo, Taylor, 1994) One attempt to establish an HIV/AIDS information system demonstrates the challenges and benefits of the effort. The Maryland HIV Information System (HIVIS) was developed by the Maryland Department of Health in the late 1980's to assist in AIDS/HIV surveillance efforts, health services planning, resource and policy development, resource allocation, and program evaluation. The database consists of longitudinal, person-based analysis files that draw upon an HIV registry, state-wide vital statistics records, public and private health insurance claims systems, institution-based data systems, and

information collected from community-based service agencies. (Hidalgo, 1990)

Health status measures

Many health status instruments have been developed to measure the health-related quality of life of people with HIV. (Wu, 1997; Smith, 1997) An excellent review of the leading health-related quality of life measures, with an appendix that includes representative questions from the key instruments to assess various domains, can be found in an article by Wu et al in the journal *Quality of Life Research* (Wu, 1997).

Although these instruments are largely based on the pool of items that were developed for the Medical Outcomes Study (MOS), they have been found to be reliable and valid in people with various stages of HIV disease and demographic characteristics. The HIV-related instruments contain various combinations of the 15 MOS subscales, and vary in reliability, construct validity, internal consistency of scales in a range of HIV-infected patients, as well as floor and ceiling effects. While there was not input from people with HIV in the development of the MOS questions, the items address the concerns identified by consumers in the development of the HCSUS survey. The widespread use of these instruments has contributed to an understanding of quality of life as an outcome of health care for people with HIV.

The MOS measures that have been extensively used in clinical trials include the MOS-HIV, HIV PARSE (Patient Reported Status and Experience Survey), and ACTG SF-21 (AIDS Clinical Trials Group Short-Form 21). The SF-20 (Short Form 20) has been used with large surveys needing to take respondent burden into consideration. The HCSUS survey, a national longitudinal survey, includes 30 items related to health status.

Decision modeling

Decision modeling has been applied to HIV research since the beginning of the epidemic. An area of particular interest has been the projection of the number of people with AIDS and HIV. The study of the natural history of HIV disease, the monitoring of trends in incidence in specific populations, and the passage of time have enabled corrections to be made to the calculations, but estimates continue to be broad. Using a back calculation method of computation, experts estimated that the number of HIV infected people in the U.S. was between 650,000 and 900,000 in 1992. (Holmberg, 1996; Karon, 1996) Modeling has also been applied to health care utilization data to project costs and cost-effectiveness of care, although the extrapolation from current data continues to be problematic because the actual number of people with HIV varies so widely geographically, institutionally, and demographically.

A recent study applied modeling to the findings of an actual clinical trial (Merck protocol 035) to project the long-term outcomes and costs of HIV antiretroviral therapy (ART):

> *A model was developed that used observed HIV RNA levels and CD4 cell counts to estimate the probability that an individual would progress from asymptomatic infection to the first*

AIDS-defining illness and death and to estimate the total net cost of care and long-term cost-effectiveness of ART. The model projected that for an individual without AIDS who received triple therapy (indinavir, zidovudine and lamivudine) the progression to AIDS and death would be delayed more than for a patient who received double therapy (zidovudine and lamivudine) if no other treatment were offered.

Because of this delay in disease progression, the total discounted cost over the initial 5-year period was projected to be $5,100 lower for patients who received triple therapy compared with double therapy if suppression with triple therapy lasts up to 3 years. At 20 years, the incremental costs per life-year gained by adding indinavir to the double regimen was estimated at $13,229, which is well within the range of other widely accepted medical interventions. (Cook, 1999)

Experimental designs

Randomized clinical trials of new drugs have been conducted extensively, through both NIH and pharmaceutical company funding. These studies have had impressive results in demonstrating the efficacy of new treatments both for opportunistic infections and for HIV itself. The consistency of positive findings regarding combinations of drugs used in highly active antiretroviral treatment (HAART) led to the development of a standard of care that was codified into the HIV treatment guidelines in a surprisingly rapid time-frame. Health status measures have been increasingly incorporated into randomized clinical trials, and have been found in some cases to be more sensitive to differences in treatment arms than biological indicators such as CD4 counts. (Bozzette and Kanouse, 1995)

While much attention has been focused on treatment efficacy trials, HIV-related research conducted using other experimental designs has looked at factors related to treatment effectiveness. Cohort studies and cross-sectional studies of convenience samples of people with HIV and AIDS have contributed to the knowledge base regarding the HIV disease process, the needs of people living with HIV, and the processes and outcomes of HIV care. Currently there is a great deal of research related to treatment adherence because of the challenges associated with maintaining the complex regimens of HAART. There are natural history studies of specific groups of people with HIV and AIDS in care. An example is the Multicenter AIDS Cohort Study (MACS), a large-scale cohort study of seronegative and seropositive gay men who have been monitored intensively over a long period of time. Models of care have been developed and studied, with growing interest now being directed toward studies of the efficiency of care within managed care networks.

Process research in HIV

The concern with the quality and cost of HIV care has focused attention on process measures for the evaluation of care within institutions, including the development of systems and indicators for continuous quality improvement programs. Consumers and payors alike are concerned that these be established within systems of managed care in order to monitor the care that is provided. Process measures utilized internally can become data sources for use in research outside these systems of care.

Discussion of the evaluation of HIV-related clinical guidelines

HIV clinical guidelines have been developed using a formal consensus process involving key stakeholders, including clinicians, researchers, consumers, payors, and policy-makers. (CDC, April 24, 1998) The guidelines form a standard of care that is widely accepted by clinicians and consumers. Certain components of the standards are based upon demonstrated efficacy (combination antiretroviral therapy), while other components rely on widespread opinion that they are appropriate (pre-test counseling; case management to link consumers to services and to assist in adherence to medical regimens).

The development of criteria for the evaluation of HIV-related standards of care is challenging for a number of reasons. In an excellent analysis of the issues, Bozzette and Asche explain that using explicit criteria is difficult because of the wide variation in what is considered appropriate care and the frequency with which specific aspects of treatment change. (Bozzette and Asch, 1995) Evaluation of HIV care involves addressing a fairly complex combination of medical and psychical factors. The criteria for meeting medical standards must allow for the different levels of disease and the range of opportunistic infections and complications to be monitored and treated. Moreover, the constant introduction of new HIV-related treatments and frequent advances in what is considered appropriate care require that standards remain open-ended and generic. In this environment of change, there are differences in standards from community to community, among clinicians and their patients. One community may demand the introduction of a new drug into clinical practice when preliminary research findings point to its efficacy, while another community may expect to wait until further investigation demonstrates more long-term benefit.

The persistence of ambiguity and doubt in the rapidly advancing field of HIV treatment lead to the need for individualized treatment regimens that are based on careful thought on the part of both the clinician and the consumer. Thus, medical standards, while possible, must be flexible. As Bozzette and Asche point out, what may appear as over-treatment or under-treatment in a clinical record review using standardized treatment criteria may be a treatment regimen decided upon by an informed clinician/patient team.

Moreover, because surrogate-marker-based outcome measures such as viral load have been developed and refined over time, longitudinal comparisons using the same measures are often not possible: the standard of measurement today quickly becomes inadequate.

Psychosocial aspects of clinical care are increasingly considered important components of standard practice, and certainly have become embedded in HIV clinical guidelines. The most obvious of these are pre- and post-test counseling associated with HIV testing and case management, both of which are considered important in assuring access to care and adherence to the complex antiretroviral regimens. These components of care are not only important in terms of achieving psychosocial outcomes such as well-being and quality of life, they are considered critical to delivering the medical processes of care. Thus, psychosocial care elements must be evaluated despite the lack of evidence of links with measurable outcomes.

The fact that two key outcome indicators, morbidity (measured as diagnosis of CDC-defined AIDS)

and mortality, have seen significant drops at least temporally associated with the advent of the use of antiretroviral combination therapy (CDC, 1999) creates a paradox in terms of evaluating HIV-related clinical guidelines: the outcomes have already been achieved. The challenge of actually evaluating the clinical standards of care that are now being implemented involves documenting that the processes of care are occurring and that shorter-term process and outcome measures are being achieved. Establishing a causal relationship between the process of care and the health and well-being of patients might be possible by measures such as the degree of implementation (this is somewhat analogous to the documentation of a dose-response relationship).

Overall, internal evaluation of the guidelines will be accomplished within individual institutions or health care systems, where the continuum of indicators will also include morbidity and mortality. However, these cannot be compared across sites because they have been shown to differ widely not only demographically but geographically. (CDC, 1999) Thus, one of the challenges is to quantify the criteria for the standard of care, including indicators of process and outcome, and establish the monitoring and evaluation infrastructure for ongoing or intermittent mechanisms of evaluation. A number of criteria have been considered such as immunization against common infections (e.g., hepatitis B), routine screening for latent infections (e.g., MTB and syphilis), regular monitoring of immune status (e.g., frequency of viral load monitoring). However, the optimum level of care (e.g., treatment regimen, frequency of laboratory monitoring, or duration of an adherence intervention) can differ widely according to expert opinion, clinical setting, patient preference, and numerous other factors that make quantifying the standard extremely problematic. (Bozzette, 1995) What constitutes too little care (below the minimum of necessary care) and too much care (above the maximum level of appropriate care) often leaves a wide range in HIV care.

> *Clinical practices become outdated very quickly as randomized trial results outpace the slow diffusion and validation of formally untested practices so common in other areas of medicine and as fads in therapy rapidly rise and fall based on shifting preliminary data. In short, one might expect that the consensus of expert opinion may be less reliable in HIV care. (Bozzette, 1995, S47)*

REFERENCES

Armenian HK. Case-control methods. In *Epidemiology and Health Services*. Eds: Armenian HK and Shapiro S. New York: Oxford University Press. 1998.

Aseltine RH, Carlson KJ, Fowler FJ, and Barry MJ. 1995. Comparing prospective and retrospective measures of treatment outcomes. *Medical Care* 33(4):AS67-AS76.

Atlas S, Singer DE, Keller RB, Patrick DL, and Deyo RA. 1996. Application of outcomes research in occupational low back pain: The Maine lumbar spine study. *American Journal of Industrial Medicine* 29:584-589.

Barry MJ. November 1994. Proving early detection and treatment of prostate cancer does more good than harm: the need to support randomized trials actively. *The Journal of Urology* 152:1903-1904.

Bayley RB, London MR, Grunkemeier GL, and Lansky DJ. 1995. Measuring the success of treatment in patient terms. *Medical Care* 33(4, suppl): pp. AS226-AS235.

Bozzette SA and Asch S. 1995. Developing quality review criteria from standards of care for HIV disease: A framework. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 8(Suppl 1):S45-S52.

Bozzette SA, Berry SH, Duan N, Frankel MR, Leibowitz AA, Lefkowitz D, et al. Dec 1998. The care of HIV-infected adults in the U.S. *New England Journal of Medicine* 339(26):1897-1904.

Bozzette, SA, Finkelstein DM, Spector SA, Frame P, Powderly WG, Weili H, Phillips L, Craven D, van der Horst C, and Feinberg J. March 16, 1995. A randomized trial of three antipneumocystis agents in patients with advanced HIV infection. *The New England Journal of Medicine* 332(11):693-699.

Bozzette, SA, Kanouse DE, Berry S, Duan N. January 25, 1995. Health status and function with zidovudine or zalcitabine as initial therapy for AIDS: A randomized controlled trial. JAMA 273(4):295-301.

Cauley JA, Lucas FL, Lewis HK, Stone K, Browner W, and Cummings SR. 16 February 1999. Elevated serum estradiol and testosterone concentrations are associated with a high risk for breast cancer. *Annals of Internal Medicine* 130(4 Pt 1):271-277.

CDC. 1999. *Trends in the HIV& AIDS Epidemic*. Atlanta GA: DHHS.

CDC. April 24, 1998. Report to the NIH Panel to define principles of therapy of HIV infection. *MMWR* 47(RR-5).

CDC. April 24, 1998. Guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents. *MMWR* 47(RR-5).

Clinton JJ. October 16, 1991. Outcomes Research: A way to improve medical practice. *JAMA* 266(15):2057.

Conroy M. July 1995. Clinical guidelines: their implementation in general practice. *British Journal of General Practice* 45:371-375.

Cook J, Dasbach E, Coplan P, Markson L, et al. 1999. Modeling the long-term outcomes and costs of HIV antiretroviral therapy using HIV RNA levels: Application to a clinical trial. *AIDS Research and Human Retroviruses* 15(6):499-508.

DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. *Controlled Clinical Trials* 7:177-88.

Deyo RA, Andersson G, Bombardier C, Cherkin DC, Keller RB, Lee CK, et al. 1994. Outcome measures for studying patients with low back pain. *Spine* 19(18S):2032S-2036S.

Deyo RA, Ciol MA, Cherkin DC, Loeser JD, and Bigos SJ. 1993. Lumbar spinal fusion: A cohort study of complications, reoperations, and resource use in the Medicare population. *Spine* 18(11):1463-1470

Deyo RA, Taylor VM, Diehr P, Conrad D, Cherkin DC, Ciol M, and Kreuter W. 1994. Analysis of automated administrative and survey databases to study patterns and outcomes of care. *Spine* 19(18S):2083S-2091S.

Ellrodt AG, Conner L, Reidinger M, and Weingarten S. 1995. Measuring and improving physician compliance with clinical practice guidelines. *Annals of Internal Medicine* 122:277-282.

Feinglass, J, Yarnold PR, McCarthy WJ, and Martin GJ. 1998. A Classification Tree Analysis of selection for discretionary treatment. *Medical Care* 36(5):740-747.

Fowler FJ, Barry MJ, Lu-Yao G, Wasson J, Roman A, and Wennberg J. June 1995. Effect of radical prostatectomy for prostate cancer on patient quality of life: Results from a Medicare study. *Urology* 45(6):1007-1015.

Fowler FJ, Cleary PD, Magaziner J, Patrick DL, and Benjamin KL. 1994. Methodological issues in measuring patient-reported outcomes: The agenda of the Work Group on Outcomes Assessment. *Medical Care* 32(7):JS65-JS76.

Glass G. 1976. Primary, secondary and meta-analysis of research. Educational Research 5:3-9. Cited in Goodman, SN. Meta-analysis in health services research. In *Epidemiology and Health Services*. Eds: Armenian HK and Shapiro S. New York: Oxford University Press. 1998. Gordis, Leon. 1996. *Epidemiology*. Philadelphia: W.B. Saunders.

Goldberg HI, Cummings MA, Steinberg EP, Ricci EM, Shannon T, Soumerai SB, et al. 1994. Deliberations on the dissemination of PORT products: Translating research findings into improved

patient outcomes. *Medical Care* 32(7):JS90-JS110, Supplement.

Greene R, Bondy PK, and Maklan CW. June 1994. The National Medical Effectiveness Research Initiative: The search for what really works in treating common clinical conditions. *Diabetes Care* 17(Supplement 1):45-49.

Grimshaw JM and Russell IT. Nov 27, 1993. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *The Lancet* 342:1317-1322.

Hidalgo J. 1990. Development and application of statewide AIDS information systems in health services planning and evaluation. *Evaluation and Program Planning* 13:39-46.

Hoffman RM, Kent DL and Deyo RA. 1991. Diagnostic accuracy and clinical utility of thermography for lumbar radiculopathy: a meta-analysis. *Spine* 16(6):623-628.

Holmberg SD. May 1996. The estimated prevalence and incidence of HIV in 96 large US Metropolitan Areas. *American Journal of Public Health* 86(5):642-654.
Horowitz CR, Goldberg HI, Martin DP, Wagner EH, Fihn SD, Christensen DB, and Cheadle AD. November 1996. Conducting a randomized controlled trial of CQI and Academic detailing to implement clinical guidelines. *Journal of Quality Improvement* 22(11):734-750.

International Perinatal HIV Group. April 1,1999. The mode of delivery and the risk of vertical transmission of human immunodeficiency virus type 1: A meta-analysis of 15 prospective cohort studies. *NEJM* 340(13)977-87.

Jette AM. 1980. Health status indicators: Their utility in chronic-disease evaluation research. *Journal of Chronic Disease* 33:567-579.

Karon JM, Rosenberg PS, McQuillan G, Khare M, Gwinn M, and Petersen LR. July 10, 1996. Prevalence of HIV infection in the US, 1984-1992. *JAMA* 276(2):126-131.

Kasper JF, Mulley AG, and Wennberg JE. June 1992. Developing shared decision-making programs to improve the quality of health care. QRB pp. 183-190.

Kind P and Dolan P. 1995. The effect of past and present illness experience on the valuations of health states. *Medical Care* 33(4):AS255-AS263.

Lansky D, Butler JBV, and Waller FT. May 1992. Using health status measures in the hospital setting: From acute care to Outcomes Management. *Medical Care* 30(5):MS57-MS73.

Lave JR, Pashos CL, Anderson GF et al. 1994. Costing medical care: Using Medicare administrative data. *Medical Care* 32(7):JS77-89.

Leape LL, Hilborne LH, Bell R, Kamberg C, and Brook RH. 1999. Underuse of cardiac procedures: Do women, ethnic minorities, and the uninsured fail to receive needed revascularization? *Annals of Internal Medicine* 130:183-192.

Maklan CW, Greene R, and Cummings MA. 1994 Methodological challenges and innovations in patient outcomes research. *Medical Care* 32(7, suppl): pp. JS13-JS21.

McDowell I and Newell C. 1996. *Measuring Health A Guide to Rating Scales and Questionnaires*. New York: Oxford University Press.

Mohr PE. September 1994. Patterns of Health Care Use Among HIV-Infected Adults: Preliminary Results. *ACSUS Report* 3:1-17.

Powe NR, Tielsch JM, Schein OD, Luthra R, Steinberg EP, and the Cataract Patient Outcome Research Team. Feb 1994. Rigor of research methods in studies of the effectiveness and safety of cataract extraction with intraocular lens implantation. *Archives of Ophthalmology* 112:228-238.

Powe NR, Turner JA, Maklan CW and Ersek M. Alternative methods for formal literature review and meta-analysis at AHCPR Patient Outcomes Research Teams. 1994. *Medical Care* 32(7):JS22-JS37.

Roper, WL, Winkenwerder W, Hackbarth GM, and Krakauer H. 1988. Effectiveness in health care: An initiative to evaluate and improve medical practice. *NEJM* 319(18)1197-1202.

Sackett DL, Chambers LW, MacPherso AS, Goldsmith CH, and Mcauley RG. 1977. The development and application of indices of health: General methods and a summary of results. *American Journal of Public Health* 67(5):423-428.

Schlesselman JJ. 1982. *Case-Control Studies: Design, Conduct, Analysis*. New York: Oxford University Press.

Smith KW, Avis NE, Mayer KH, and Swislow L. 1997. Use of the MqoL-HIV with asymptomatic HIV-positive patients. *Quality of Life Research* 6:555-560.

Stewart AL and Ware JE, eds. 1992. *Measuring Functioning and Well-Being: the Medical Outcomes Study Approach*. Durham NC: Duke University Press.

Thornton C, Fasciano N, Turner BJ, Cherlow A, Bencio DS. March 1997. *Methods for Identifying AIDS Cases in Medicare and Medicaid Claims Data*. Baltimore MD: Mathematica Policy Research, Inc.

Turner JA, Ersek M, Herron L, Haselkorn J, Kent D, Ciol MA, and Deyo R. August 19, 1992. Patient Outcomes after lumbar spinal fusions. *JAMA* 268(7):907-911.

Turner JA, Ersek M, Herron L, and Deyo R. 1992. Surgery for lumbar spinal stenosis: Attempted meta-analysis of the literature. *Spine* 17(1):1-8.

Ware JE, Measures for a new era of health assessment. 1992. In *Measuring Functioning and Well-Being*. Durham NC: Duke University Press.

Ware JE, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, and Raczek A. 1995. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: Summary of results from the Medical Outcomes Study. *Medical Care* 33(4, suppl):AS264-AS279.

Windsor RA, Baranowski T, Clark N, and Cutter G. 1984. Evaluation of Health Promotion and Education Programs. Mountainview CA: Mayfield Publishing Company.

Wu AW, Hays RD, Kelly S, Malitz F, and Bozzette SA. 1997. Applications of the Medical Outcomes Study health-related quality of life measures in HIV/AIDS. *Quality of Life Research* 6:531-554.

SUGGESTED BIBLIOGRAPHY

Agency for Health Care Policy and Research. December 1996. Publications of the Patient Outcomes Research Teams (PORTS): PORT and PORT II Projects. Rockville MD: DHHS.

Bozzette SA and Asch S. 1995. Developing quality review criteria from standards of care for HIV disease: A framework. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 8(Suppl 1):S45-S52.

Bozzette SA, Berry SH, Duan N, Frankel MR, Leibowitz AA, Lefkowitz D, et al. Dec 1998. The care of HIV-infected adults in the U.S. *New England Journal of Medicine* 339(26):1897-1904.

Deyo RA. Practice variations, treatment fads, rising disability: Do we need a new clinical research paradigm? *Spine* 18(15):2153-2162.

Deyo RA and Carter WB. May 1992. Strategies for improving and expanding the application of health status measures in clinical settings: A research-developer viewpoint. *Medical Care* 30(5 Suppl):MS176-MS186.

Deyo RA, Taylor VM, Diehr P, Conrad D, Cherkin DC, Ciol M, and Kreuter W. 1994. Analysis of automated administrative and survey databases to study patterns and outcomes of care. *Spine* 19(18S):2083S-2091S.

Fowler FJ, Cleary PD, Magaziner J, Patrick DL, and Benjamin KL. 1994. Methodological issues in measuring patient-reported outcomes: The agenda of the Work Group on Outcomes Assessment. *Medical Care* 32(7):JS65-JS76.

Goldberg HI, Cummings MA, Steinberg EP, Ricci EM, Shannon T, Soumerai SB, et al. 1994. Deliberations on the dissemination of PORT products: Translating research findings into improved patient outcomes. *Medical Care* 32(7):JS90-JS110, Supplement.

Jette AM. 1980. Health status indicators: Their utility in chronic-disease evaluation research. *Journal of Chronic Disease* 33:567-579.

Lave JR, Pashos CL, Anderson GF, Brailer D, Bubolz T, Conrad D, et al. 1994. Costing medical care: Using Medicare administrative data. *Medical Care* 32(7, Suppl):JS77-JS89.

Maklan CW, Greene R, and Cummings MA. 1994. Methodological challenges and innovations in patient outcomes research. *Medical Care* 32(7, suppl): pp. JS13-JS21.

McDowell I and Newell C. 1996. *Measuring Health A Guide to Rating Scales and Questionnaires*. New York: Oxford University Press.

Powe NR, Turner JA, Maklan CW and Ersek M. Alternative methods for formal literature review and meta-analysis at AHCPR Patient Outcomes Research Teams. 1994. *Medical Care* 32(7):JS22-JS37.

Wu AW, Hays RD, Kelly S, Malitz F, and Bozzette SA. 1997. Applications of the Medical Outcomes Study health-related quality of life measures in HIV/AIDS. *Quality of Life Research* 6:531-554.