American Thoracic Society Workshop
MEDICAL SECTION OF THE AMERICAN LUNG ASSOCIATION

Outcomes Research -in Critical Care

Results of the American Thoracic Society Critical Care Assembly Workshop on Outcomes Research

GORDON D. RUBENFELD, DEREK C. ANGUS, MICHAEL R. PINSKY, J. RANDALL CURTIS, ALFRED F. CONNORS, jr., GORDON R. BERNARD, and The Members of the Outcomes Research Workshop

INTRODUCTION

The terms "outcomes research," "patient-centered care," "treatment effectiveness," and "cost-effectiveness" have found their way into the daily discourse of clinicians, researchers, health policymakers, and managed care providers. Several factors have contributed to the interest in outcomes research in general, and specifically critical care outcomes research. First, there is growing realization that the evidence base is lacking for much of daily clinical practice and that this may be particularly the case in critical care (1). Second, the rapid growth in health care costs, attributed in large part to increasing use of expensive technology, has forced clinicians, payers, and society to consider the value of medical care (2). By some estimates critical care accounts for up to 1% of the entire gross domestic product of the United States (3). Finally, recent studies using nonexperimental methods have raised important questions about the validity of these methods in critical care compared with traditional randomized trials (4). Outcomes research is being used today to formulate clinical practice guidelines, to assess the quality of medical care, and to inform health policy decisions. In an effort to provide a readily accessible document elaborating the focus, methods, goals, and limitations of outcomes research for critical care practitioners, investigators, and other outcomes research consumers, the American Thoracic Society convened a Workshop on Outcomes Research in Critical Care, the summary of which is presented below.

METHODS

The workshop was conducted April 4–6, 1997. Panelists from many disciplines were invited including: clinical trial design, evidence-based medicine, severity of illness measurement, observational outcomes research, health economics, medical database management, and statistical modeling. Three breakout groups were established: special considerations in critical care outcomes research, outcome measures, and data and mathematical modeling. Plenary sessions for the entire workshop were moderated by a different cochair to reduce the bias of any individual chair. At the completion of the workshop, a writing committee was chosen and the resulting manuscript was then distributed to all panelists for comment.

The document is designed as both a resource to understand current methods as well as an agenda for further study. After presenting a working definition of outcomes research, there is a discussion of the unique features of critical care that require special consideration in designing an outcomes research question. Then we review the outcome measures, including quality of life, available to investigators in this area and discuss their merits. The data section outlines databases suitable for critical care outcomes research, their strengths and weaknesses, and current efforts to improve their quality. Finally, the modeling section addresses the problematic issue of using analytic techniques to draw valid conclusions from observational data.

WHAT IS OUTCOMES RESEARCH?

Outcomes research focuses on the effects of medical care on individuals and society. To address its questions, outcomes research relies on a variety of disciplines including clinical epidemiology, informatics, anthropology, economics, health services research, health policy, and biostatistics (5). Formulating an exact definition for outcomes research is probably less useful than gaining insight into its emphases relative to traditional clinical research (Table 1). One essential feature of outcomes research is the central role of patient-centered and policy-relevant outcomes. Whereas traditional clinical research explores the mechanisms of disease through their biologic manifestations, outcomes research studies the effect of treatments on endpoints important to patients and society. Traditional clinical research variables might include blood pressure, tumor size, and peak expiratory flow whereas outcomes research variables would include survival, quality of life, satisfaction with care, and cost.

It is the research focus, and not any specific methodology, that distinguishes outcomes research from traditional clinical research. Outcomes research uses a variety of study designs and techniques (Figure 1). Qualitative research methods generate hypotheses, identify the relevance of findings to specific groups, and describe complex phenomena that do not necessarily lend themselves to formal hypothesis testing. Quantitative methods include both experimental designs where the investigator controls assignment and nonexperimental or observational methods. Randomized controlled trials, when designed as effectiveness trials can yield outcomes information (6). Meta-analysis incorporates information from several randomized clinical trials (RCT) or observational studies done in different settings and therefore may approximate the infor-

(Received in original form /u/y 24, 7 998 ond in revised form November 3, 7 998)

This workshop was supported by the American Thoracic Society.

Correspondence and requests for reprints should be addressed to Gordon D. Rubenfeld, M.D., M.Sc., Box 359762, Harborview Medical Center, 325 9th Ave., Seattle, WA 98104. E-mail: nodrog@u.washington.edu

Am J Respir Crit Care Med Vol 160. pp 358-367, 1999
Internet address: www.atsjournals.org

TABLE 1

A COMPARISON OF FEATURES BETWEEN TRADITIONAL CLINICAL RESEARCH AND OUTCOMES RESEARCH

Traditional Clinical Research	Outcomes Research
Efficacy	Effectiveness
Mechanisms of disease	Impact of disease on the patient
Experimental	Observational
Feasibility	Cost-effectiveness
The effect of biochemical and physiologic factors on biophysiologic outcomes	The effect of socioeconomic factors on patient-centered outcomes
Disease-centered	Patient- and community-centered
Provider-oriented	Consumer-oriented
Inventing technology	Assessing technology
Drugs and devices	Processes and delivery of care
Methods from the "hard" sciences (physics, biochemistry, physiology)	Methods from the "social" sciences (economics, social and behavioral sciences, epidemiology)

mation provided by larger studies on more heterogeneous populations. By synthesizing data from RCTs and observational studies, decision analysts produce a model of a specific clinical problem. Once built, a valid decision model simulates a clinical situation and can provide outcome information about populations not necessarily included in the original trials (7). Economic models add cost data to decision models and generate cost-effectiveness ratios which provide information about the relative value of medical care options (8).

The Workshop focused on observational outcomes research which includes prospective and retrospective cohort studies, case-control studies, and surveys. Observational outcomes studies on severity of illness, mortality, and cost have a long history in critical care (9-11). Observational outcomes research can be used to identify associations between various exposures and outcomes. Exposure is defined broadly in outcomes research and can include medications, medical procedures, sociodemographic status, organization of health care delivery, access to health care, region of country, or provider. Associations between exposures and outcomes are particularly important when there is evidence to believe a causal relationship exists. Some observed associations are due to a third variable that can obscure or create the appearance of a causal association when none exists. For example, uninsured patients may have higher intensive care unit (ICU) mortality because they present with more severe disease. Identifying and minimizing the effect of these confounding variables constitutes a large part of the analysis in observational outcomes research.

Randomized controlled trials offer an effective solution to the problem of confounding variables. Randomization breaks the association between an exposure and confounding variables by allocating the exposure by chance. However, RCTs are expensive and, particularly for rare diseases, can take years to complete. Because of their cost and the regulatory demands on drug and device manufacturers, RCTs are frequently designed as efficacy studies in highly defined patient populations with experienced providers and, therefore, provide little evidence about effectiveness (12). Some exposures, for example, socioeconomic status, access to specialty care, or type of medical insurance may be difficult or impossible to randomize (13, 14). Refinements to treatments known to be effective, for example, identifying the optimal timing for intervention, would require RCTs with multiple treatment arms. Diagnostic and monitoring technology are rarely subjected to evaluation with RCTs to judge their effect on clinical outcome. RCTs, precisely because they assign the intervention, are poorly suited to describing variation in current practice (15-17). Observational studies can generate hypotheses about the effectiveness of treatments that can be further tested using other research methods. Finally, although randomization is ethically justified by equipoise in the medical community, it may be difficult for individual clinicians and patients to subject treatment decisions to chance (18). Therefore, at least in principle, observational outcomes research can provide complementary information to RCTs.

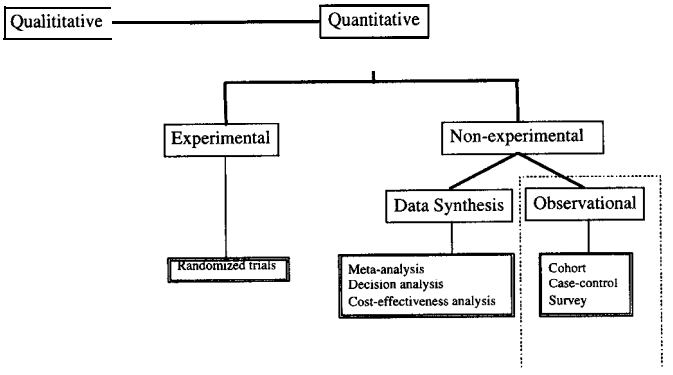

Figure 1. Study designs in outcomes research.

WHAT IS UNIQUE ABOUT CRITICAL CARE MEDICINE?

Critical care presents several unique challenges for outcomes research. To formulate a research question an investigator must be able to define a disease, treatment, patient population, or provider to study. Operationalizing these variables for critical care outcomes research is complex. Compare the investigator studying outcomes after myocardial infarction to one studying the same research question in sepsis. Myocardial infarction is a common disease that is readily diagnosed with several laboratory tests. Patients with myocardial infarction receive treatments that are relatively specific for the diagnosis and are generally cared for by cardiologists, internists, or family physicians. Contrast sepsis which is defined by clinical criteria that continue to evolve and does not have an accepted diagnostic test or unique treatment (19-21). Critically ill septic patients may receive their care from a variety of physicians including surgeons, anesthesiologists, family physicians, general internists, pulmonologists, and pediatricians. Critical care is a challenge to the outcomes researcher precisely because the key variables of disease, patient population, therapy, and provider are difficult to define.

A pragmatic solution is to adopt a geographic definition of critical care (22, 23). By this definition, the critically ill are pa-

tients cared for in an ICU, critical care is treatment provided in an ICU, and intensivists are physicians who attend in an ICU. However, flexible nurse staffing, mobile technology, intermediate care units, and the growth of subacute care for chronically critically ill patients make the assumption that critical care begins and ends at the ICU door problematic. In addition, some patients, such as those recovering from operative procedures, may receive care in an ICU but not be critically ill (24-27).

In the absence of simple solutions to defining critical care or the critically ill for outcomes research, one must carefully understand the research focus and how it was operationalized in any given study. Every observational outcomes research project will have to define a critically ill population and or critical care to identify the appropriate patients, exposures, and outcomes to study. The current literature relies heavily on geography to define the critically ill; however, this is not the only, or necessarily the optimal solution. In addition to location, a critically ill population could be defined in terms of the treatment or monitoring they receive (for example, all patients who receive mechanical ventilation or electrocardiographic telemetry regardless of location in the hospital) or diagnosis (all patients with septic shock regardless of location in the hospital). Some studies will choose to define critical care instead of a critically ill population. Critical care can also be defined by location, by specific treatments or monitors (for example, cardiopulmonary resuscitation or pulmonary artery catheterization), by the type of ICU (for example, open versus closed ICU), or by the provider (28, 29). Because there is no single solution to this problem, readers and investigators will need to approach each study critically to see how the critically ill patients and critical care were defined, and to identify the potential biases introduced by these choices.

Unlike patients with myocardial infarction, most of whom are admitted to the hospital for this problem, critically ill patients use the ICU as one of many resources during an episode of hospital care frequently directed at some other underlying disease. Therefore, separating the outcomes of the intensive care treatment from the outcomes of the primary disease and its treatment can be difficult. For example, death in the ICU from multiple organ failure complicating a bone marrow transplant is both an outcome of the transplant and of the ICU care (30). The research question will direct the investigator to choose a perspective that encompasses the disease process, an individual hospitalization or to focus on the critical care component.

One area of outcomes research where critical care uniquely excels is the availability of severity of illness measures. In the absence of randomization, observational outcomes research relies heavily on the ability to mathematically control for differences in patients' characteristics with severity of illness measures. Critical care has developed a rich selection of disease-specific and generic severity of illness measures (9, 31).

In summary, outcomes research in critical care is complicated because key variables are more difficult to operationally define than other diseases and medical specialties. Because critical care is a component of treatments for complications of other illnesses, the line between critical care outcomes and the outcomes and process of care for the primary disease is blurred.

OUTCOME MEASURES

There is a growing recognition that clinical research needs to define and focus on the outcomes of medical care that are important to patients, their families, and to society. These outcomes have been coined "patient-centered outcomes" (32). Intensive care clinical research has often focused on survival and physiologic impairment. Although this is the appropriate focus in some settings, our society is raising questions about the effect of intensive, aggressive medical care on patient-centered outcomes (33). Mortality is, in most situations, the primary outcome for critical care research because critical care therapy is directed at overcoming an acute life-threatening illness. But other outcomes are important to patients and their families including quality of life, functional status, freedom from pain and other symptoms, and satisfaction with medical care (34).

Mortality

Although death is an unequivocal endpoint, there are several timepoints at which to measure it: ICU or hospital survival, time until death, death at a fixed timepoint (30 d, 6 mo, 1 yr). The "correct" mortality endpoint depends on the specific research question, the mechanisms and timing of the disease and/or treatment under study, and the study design. There are some important general features to consider with each type of mortality assessment.

Practice patterns can change the interpretation of mortality endpoints. Hospital mortality may be reduced by the transfer of ventilator-dependent patients directly from the ICU to a long-term care facility. Survival to a fixed time avoids this problem; however, the selection of the time point requires careful consideration. Thirty-day survival may be less useful when ICU length of stay approaches 30 d, for example, in the fibroproliferative phase of the acute respiratory distress syndrome (ARDS). Whereas l-yr survival might avoid many of these problems, long-term survival may reflect the patient's underlying disease prognosis more than the effectiveness of their ICU care. For example, 1-yr survival might not be the best outcome for a study of the effect of an ICU-based intervention on patients with hematological malignancy. Finally, assumptions about the disease process should be tested. Early sepsis studies assumed that major mortality due to sepsis would occur within 14 or 30 d from the onset of illness (35). However, longer term studies suggest that there is ongoing mortality for substantially longer than 30 d related to sepsis and its complications (36). Thirty-day mortality may not reflect the delayed effects of sepsis or the therapy under study.

Survival analysis techniques use the actual survival time as the endpoint. Because survival time is a continuous variable, it is a more sensitive outcome measure than mortality. However, time until death in the modern ICU reflects time until withdrawal of life support (37). Thus, time to death is influenced by institutional, provider, and geographic factors that affect the decision to withdraw life support (38, 39). Furthermore, survival analysis may find a statistical difference between patients with a median 7-d survival and those with a median S-d survival (40). Although potentially useful to investigators in identifying promising treatments for further study, it is unlikely that these represent meaningful outcomes to patients. In fact, a treatment that prolongs time until death without affecting mortality could do more harm than good.

Patient-assessed Health Outcomes

Patient-assessed outcomes include quality of life, functional status, symptoms, and satisfaction with medical care (41). "Health-related quality of life" (HRQL) is used to refer to the subjective experience of the effect of health and treatment on one's satisfaction with life. The term "functional status" is used to describe an individual's ability to perform tasks in everyday life. Operationally, many of the instruments designed

to measure HRQL or functional status encompass features of both, which can make it difficult to distinguish these concepts.

The term "health status" is used to encompass both functional status and HRQL (42). Health status instruments include disease-specific measures such as those developed for asthma and COPD as well as generic measures intended for use in a broad range of patients (43, 44). Critical illness is not a disease with specific symptoms and functional impairments, making it unlikely that a critical care disease-specific instrument will be helpful. Nonetheless, the unique features of critical illness make it important that generic health status instruments be validated in this setting before they are used in clinical studies. Additional instruments allow researchers to transform patients' values (utilities) for different health states on a single linear scale that ranges from zero to one. Patient utilities can be used to calculate quality-adjusted life years (QALYs) which incorporate the quality and the quantity of life into a single measure for use in economic analyses (45). For example, a patient is discharged to a chronic ventilator weaning facility for 6 mo and dies on the ventilator. If the utility for life on a ventilator in a weaning facility is 0.2, the patient has experienced 0.1 QALYs. If the methodology is correct, this means that 6 mo on a ventilator in a weaning facility is approximately equal in value to patients as 1 mo in perfect health. Patient and family satisfaction with care represent another important domain of patient-centered outcomes.

Investigators have had to address the unique features of critical care and critical illness when measuring patient-assessed outcomes in this patient population (Table 2). First, many critically ill patients are too ill to fill out questionnaires or be interviewed. Important information about the burdens of illness and satisfaction with care in those who die cannot be measured directly (46). Family member interviews can be a reliable and valid measure of a patient's health status, however, studies suggest that this proxy information is not as useful as that obtained directly from the competent patient (47).

In some settings mortality is the most relevant outcome and patient-assessed outcomes are less important, for example, an illness after which patients return to their premorbid health status. There exists a continuum between mortality and quality of life as the primary endpoints of critical care outcomes research such that higher fatality rates from acute illnesses with minimal residual disabilities favor mortality as an outcome and chronic or terminal diseases with profound disability in survivors should emphasize quality-of-life measures. Available research suggests that the tradeoff between quantity of life and quality of life is a highly individualized decision and generalizations are problematic (48, 49).

Physiologic Measures

As a general rule, physiologic measures are not patient-centered outcomes because patients are not concerned about a physiologic abnormality if it does not affect their quality or quantity of life. However, physiologic measures are important in outcomes research because: (I) physiologic variables are essential in the assessment of severity of illness; (2) physiologic variables may be useful as surrogate markers of mortality, functional status, and quality of life; and (3) physiologic variables help researchers understand the mechanisms of disease and its treatment.

Process of Care as an Outcome

In certain situations the process of medical care is the outcome of interest. When there is evidence to support the link between a treatment and an important outcome, process measures may be a more sensitive indicator of the quality of care than the outcome variable (50, 51). For example, heparin prophylaxis for deep venous thrombosis will be a more sensitive marker for quality of care than the outcome of deep venous thrombosis because even in untreated patients this is a relatively unusual and frequently undetected complication. Other processes of care may not have evidence to support their use, but instead have ***prima facie*** validity. For example, communication with patients and families about treatment preferences is an important outcome even if the link has not been established between this process of care and patient satisfaction or quality of life. Finally, process of care is an important outcome variable when studying the factors that influence clinicians' treatment decisions (16).

Life-support-free Days

A major problem in developing nonmortality critical care outcomes is the competing effect of high death rates (52). Outcomes like duration of mechanical ventilation or incidence of organ failure may be lower in the group with higher mortality simply because of the competing effect of mortality. Patients who die cannot develop further organ failures or prolonged ventilator dependence. One proposed solution to this problem is to calculate organ-failure-free days or life-support-free days (52). For example, to calculate ventilator-free days a predefined measurement period, typically 28 d, is specified. Patients who die or are mechanically ventilated longer than this period are assigned zero ventilator-free days. All survivors accrue one ventilator-free day for each day after entry into the study that they are both alive and free of mechanical ventilation. By combining a mortality and morbidity measure the free day outcome may avoid the problem of competing mortality; however, further studies of the statistical performance and clinical significance of this variable are needed before it can be recommended as a primary endpoint.

Quality of Care at the End of Life and Quality of Death

The mortality among patients in the ICU is relatively high compared with other areas of medicine. The goals of medicine and the measures of quality of care may change dramatically

TABLE 2

USE OF PATIENT-ASSESSED OUTCOMES IN CRITICAL CARE CLINICAL RESEARCH

Type of Outcome	Specific Instrument
Quality of life	Perceived Quality of Life (PQOL) (46, 87, 88)
Health-related quality of life	Sickness Impact Profile (SIP) (46, 89, 90) Medical Outcome Study Short-Form 36 (SF-36) (91)
Depression or anxiety	Center for Epidemiologic Studies-Depression (CES-D) (92) Hospital Anxiety and Depression Scale (93)
Utilities	Time tradeoff (48)

in patients who are dying compared with patients for whom death is not imminent (53). Assessing the quality of care at the end of life, the concordance between patients' preferences for treatment and the treatments they receive, and the quality of communication with patients and families are important outcome measures of critical care.

There are a number of issues that complicate research in death and dying compared with other areas of outcomes research. Primarily the "outcome" of interest is less clear. Instruments to assess the quality of death and dying through surveys of patients, their families, and their caregivers are only now being developed; and their applicability to the ICU setting will need to be determined. Optimal processes of care have not been determined and considerable variations in care exist. It is essential for critical care physicians to appreciate the significance of this outcome even though techniques to assess it are in their infancy.

Cost

A detailed discussion of economic analysis is beyond the scope of this review and the interested reader should refer to reviews of this topic (45, 54). However, economic analyses, including cost-effectiveness analysis, are an important component of outcomes research. Recent consensus reports have made methodologic recommendations for performing and reporting economic analyses that should improve the quality of research in this area (55, 56). Additional work will need to be done to generate standards specific for economic analyses in critical care.

These outcome measures reflect a spectrum from which the investigator may choose depending on the research question. Table 3 shows how these measures could be incorporated into a registry of patients with the ARDS to answer a range of questions about the outcomes of this disease.

DATA SOURCES FOR OBSERVATIONAL OUTCOMES RESEARCH IN CRITICAL CARE

Observational outcomes research frequently relies on large datasets which contain information on patient characteristics, exposures, and outcomes. Datasets vary in their feasibility of use (size, cost, and access for the investigator), accuracy (ability of coded variables to convey the actual clinical situation), and appropriateness (whether the dataset contains all of the elements necessary to answer the research question) (57). The dataset is as important a part of the researcher's methods as the study design and data analysis. While readers can look to standard references to evaluate the study design and analysis, there is no readily available guide to the limitations of the various data sources that can be used in this research. Complicating the evaluation of the datasets is the inconsistent terminology used to identify the sources and original use for the data. For example, the term "administrative data" is used to indicate data that were originally collected for reasons other than research. This is an inappropriately broad term in that it groups datasets that vary widely in terms of feasibility of use, accuracy, and appropriateness. To address these inconsistencies and illustrate the types of data sources available to critical care outcomes researchers we present the examples that follow.

TABLE 3

OUTCOME MEASURES IN A HYPOTHETICAL REGISTRY OF PATIENTS WITH ACUTE RESPIRATORY DISTRESS SYNDROME

Mortality	28-d all cause mortality
Survival	Days until death
Quality of life	SF-36, St. George's Respiratory Questionnaire
Physiologic measures	Pulmonary function testing at 6 mo
	Exercise testing
Process of care	Use of pulmonary artery catheter
	Ventilator settings
	Type of stress ulcer prophylaxis
cost	Cost of episode of care
	Cost of care after hospital discharge
	Therapeutic Intervention Severity Score (TISS) points

Encounter Data

Encounter datasets represent databases whose purpose is to maintain a record of health care encounters, typically maintained by payers, to track reimbursement. Data are typically grouped by ambulatory care visit or hospital admission. However, these data can be summed to generate data on a region, and, if unique patient identifiers are available, can also be linked to create a longitudinal history for individual patients. Encounter datasets can be extremely large. For example, the Medicare Provider Analysis and Review (MEDPAR) file contains records for 100% of Medicare beneficiaries who use hospital inpatient services. The national file consists of approximately 11 million records.

There is considerable variation in the quality and detail of these datasets. For example, datasets may allow different number of diagnoses or procedure codes, and have varying detail regarding ICU utilization (58). When data are collected for billing purposes, financial incentives driving the data collection can impair its value (59, 60). Datasets coded by medical records departments suffer from the vagueness of the International Classification of Diseases system and the errors inherent in the abstraction process (61). Analyses based on variables available in encounter data may miss important relationships present in detailed clinical data (62). Authors presenting results from such studies should therefore specify clearly the characteristics of the dataset and its limitations.

Enrollment Data

In addition to information about patients' encounters with the medical system, one might also wish to know about the denominator population from which the encounter numerator is drawn. This allows the investigator to draw inferences about population incidence and utilization for various diseases. Census data can provide age-, race-, and gender-specific data as well as some information on socioeconomic variables at the national and regional level. Other denominator sources, including Health Maintenance Organization enrollment databases, are available for selective analyses. In the United States, the Health Care Financing Administration has data on Medicare enrollees organized by state, county, and zip code. Investigators who are limiting their analysis to a specific insurer or managed care organization can often obtain enrollment data on the insured population. These data are limited by transfer between insurance plans, which makes estimating a stable denominator for any given year a challenge.

Electronic Clinical Data

The electronic medical record promises to be a boon to outcomes research should a usable format gain widespread acceptance. There are several electronic medical records designed specifically for use in the ICU (63). Whether or not a formal electronic medical record exists, virtually all clinical information in the modern hospital, including laboratory data, pharmacy, and many diagnostic test results are converted to electronic form. The extent to which data from these separate systems can be collated into a usable resource and subse-

quently analyzed for research purposes has not been explored. Barriers to the effective use of hospital databases include their sheer volume, the lack of standardized data collection and computer formats, and protection of confidentiality and proprietary interests.

Data Registries

Data registries are databases focused on a particular disease or intervention, for example, ARDS or cardiovascular diseases (64, 65). Data registries have the advantage of collecting a set of predefined data on a large number of patients with homogenous characteristics. Frequently, data registries rely on collection from a single or several academic institutions and therefore may not generalize to other clinical settings. Because they only include patients with the disease or treatment, many important questions about incidence, variability in management, and comparative outcome cannot be answered.

Performance Data

In the current era of health care reform, several organizations and agencies have begun to collect data explicitly for the purpose of assessing institutional and provider performance. Some of these performance data sources are constructed ***de novo,*** such as the Society of Cardiovascular Surgery or Greater Cleveland Health Quality Coalition, while others, such as the Pennsylvania Health Care Cost Containment Council coronary bypass surgery and acute myocardial infarction reports involve the augmentation of existing data with clinical-based severity information (66-68). There has been considerable debate over whether the available data are sufficient to allow for the discrimination between "better" and "worse" providers (69). These datasets are of great potential value in that they are collected across a number of institutions and usually contain data significantly richer in clinical detail than other encounter datasets.

Survey Data

Traditionally, assessments of critical care and critical care therapies have been based on clinical measures and objective outcomes such as mortality. Increasingly, however, attention is being focused on the importance of patient-centered outcomes. Information on such outcomes is available through surveys of patient preferences and satisfaction with care (70). While most such surveys have not collected ICU-specific information to date, it is likely that this information may appear in the future and be a valuable data source for outcomes research in critical care.

Clinical Trial Data

Randomized controlled trials contain a considerable amount of detailed information on the demographics, severity of illness, and outcomes of critically ill patients that has been collected prospectively under rigorous conditions. In addition to meta-analysis, which combines results from several RCTs to estimate the treatment effect, it may be possible to perform a secondary cohort analysis on the patients in clinical trials. This approach has been used in clinical trials of acute myocardial infarction and pulmonary embolism to explore the effect of different risk factors on outcome (71, 72). While these datasets seem promising, the same factors that limit the assessment of effectiveness by RCTs may make extrapolation from their patient data to broader populations problematic. Consequently, readers and researchers should consider carefully the applicability of these data sources for the potential questions asked.

Prospective Cohort Data

There are several datasets that were collected with the specific intention to be representative of ICU populations (73, 74). These databases are valuable sources of rich clinical detail in broad, generally representative critically ill populations. The principal limitations governing their use include limited accessibility to investigators and the data elements collected. For example, the Acute Physiology and Chronic Health Evaluation III (APACHE III) database collected daily information only through Day 7 and thus is not a good source for understanding the characteristics of resource use in patients with long ICU length of stay. Nevertheless, the level of clinical detail, attention to ICU-specific diagnoses and variables, and size make these datasets the current standard against which other data sources ought to be compared for many critical care observational outcomes research questions.

National Minimal ICU Data Sets

In part as a response to the limitations of prospective cohort data, several countries and national societies have developed or proposed the widespread collection of a core set of data on all ICUs and ICU patients (75). These projects are ambitious and some have been plagued by a lack of funding and slow initiation. Nevertheless, since their purpose is to understand national ICU delivery, including an understanding of the demographics and outcomes of critically ill patients, the cost and processes of care, and the extent to which practice variation occurs, such datasets, once generated, will be very useful resources.

TECHNIQUES TO ADJUST FOR CONFOUNDING

In observational outcomes research, an investigator compares a group of subjects with an intervention to a control group. If the groups are different with regard to a variable that is also associated with the outcome (e.g., if one group is older than the other and age is associated with mortality), then an estimate of the treatment effect on outcome will be biased. In this example, age biases the investigator's estimate and is called a confounder. Any variable that is associated with the exposure and with the outcome in the dataset is a potential confounder. Conversely, variables that are only associated with outcome and not with exposure, or only with exposure and not with outcome, do not bias the association between exposure and outcome. Common confounding variables include age, diagnosis, severity of illness, and indication for treatment. Variables can act both as confounders and as exposures depending on the research question and causal model. For example, region of country is an exposure variable for questions that involve geographic variation in outcomes or utilization. However, in an analysis of the effect of pulmonary artery catheters on mortality, geographic region may be an important confounder because technology utilization and mortality vary by region.

Essentially, the goal of identifying and controlling for confounding variables is to justify the statement:

> All things being equal, patients with exposure "x" (e.g., a new antisepsis therapy, specialist care in the ICU, or pulmonary artery catheters) are y% more likely to have the "outcome" (e.g., survival, prolonged hospital stay, improved quality of life at 6 months).

To the extent that all confounding variables are identified, measured accurately, and adjusted for with an appropriate technique, this goal can be achieved. A variety of techniques is used to try to avoid or control for confounding in observa-

tional outcomes research to determine an unbiased estimate of the effect of an exposure.

Matching

Exposed subjects are matched with unexposed patients on the basis of the variables known to be related both to the exposure and to the outcome. The major advantages of matching are its simplicity and the fact that many readers find it a persuasive technique. However, matching has important limitations. Once the subjects have been selected by matching, the matching variables cannot be explored as exposure variables and the investigator is committed to a matched analysis. Overmatching, matching for a variable that is associated with the exposure but not the outcome, will reduce the statistical power of the study (76). Finally, matching on multiple confounding variables is often not feasible because an exact match may not be present among the controls.

Stratification

Subclasses or strata based on the confounding variables are constructed and comparisons are made between exposed and control groups within each stratum (for example, age less than 6.5 and age greater than 65). This technique is also a conceptually simple method of adjusting for confounding in which both the intervention and control groups are divided by subclasses of confounding variables. An important limitation of stratification techniques is that, as the number of confounding variables increases, the number of strata increase exponentially, resulting in strata that do not provide information because they do not contain both intervention and control patients. This method is best when the number of observations is large or the number of confounding variables is small.

Multivariable Adjustment

Multivariable adjustment can be achieved by modeling the relationship between the presence or absence of the exposure (the independent variable) and the outcome (the dependent variable) adjusted for all confounding variables. A variety of mathematical models is available depending on the outcome variable and the research question. The logistic regression model is used when the outcome variable is binary (for example, mortality). The exposure effect is presented as an odds ratio with confidence interval adjusted for the confounding variables in the model. The Cox proportional hazards model is used when the outcome variable is survival time or time until an event other than death. Here the exposure effect is an adjusted hazard ratio and reflects the increased rate of outcome in the exposed. Finally, linear regression, or generalizations of this model, can be used when the outcome is continuous, such as cost, duration of hospital stay, or other continuous variables. The effect size can be presented as the adjusted difference in outcome for the exposed compared with control patients. Using these techniques, the association of a treatment with the outcome can be estimated while mathematically holding all the confounding variables constant.

The validity of multivariate techniques depends on the extent to which the model fits the data and the mathematical assumptions of the model are not violated. Model fit and suitability should be tested by appropriate means including plots of the data and measures of fit (77-79). Failure to consider these possibilities can lead to a flawed analysis and erroneous conclusions.

Propensity Scores

Propensity scores are developed in much the same way as ICU severity of illness scores like APACHE or the Simplified Acute Physiology Score (SAPS). However, instead of providing the estimated probability of hospital mortality, propensity scores combine multiple variables to generate an individual's probability of exposure to the variable of interest. The propensity score can be used for matching, stratification, or multivariable adjustment. Because propensity scores correlate with the use of a treatment, they can be used to control for indication bias which results when patients who receive a treatment are different from those who do not (80). There are important differences between severity of illness and propensity scores. Some variables may have a small effect on mortality and a relatively large effect on the decision to use a therapy. For example, the progression of disease, response to initial therapy, diagnostic uncertainty, and even the patients' insurance status may all affect the treatment selection process.

Instrumental Variables

Instrumental variables, like propensity scores, are associated with the exposure, but are selected because they are independent of patient outcomes. When outcomes are analyzed by the instrumental variable, a random distribution of patient characteristics is expected between the exposed and control groups. Theoretically, this simulates random assignment of patients to the exposure. An example is the use of patient's distance from a hospital that uses a high-intensity approach to managing myocardial infarction (81). Patients living closer to high-intensity hospitals are more likely to receive aggressive care, but, hopefully, for the validity of this analysis, are not medically different from patients living further from a high-intensity hospital. By comparing patients by distance and adjusting for differences in measured covariables, one hopes to create an unbiased estimate of the effect of treatment at a high-intensity hospital.

CAN WE BELIEVE THE RESULTS FROM OBSERVATIONAL OUTCOMES STUDIES?

Ultimately, after reading an observational outcomes study, readers must decide to what extent they believe the estimate of the exposure's effect. Observational outcomes research, regardless of the study design, is subject to a variety of biases in addition to the issues of confounding raised earlier (82, 83). Biases can be introduced in the way the patient sample is selected, in the way the outcome, exposure, and confounders are defined or measured, and in the analysis (84). Readers are cautioned not to combine all such concerns into an assessment of whether the data collection was "prospective" or "retrospective"; neither approach is necessarily superior. Instead the reader should focus attention on the specific nature of the bias, its direction, and magnitude. Biases do not always invalidate a study. For example, an author studying the effect of antibiotics on ICU mortality recognizes that patients with more severe sepsis receive broader spectrum empiric antibiotics. In the study dataset, early broad-spectrum antibiotics are associated with reduced mortality. Here the indication bias is in the opposite direction of the study findings and actually strengthens the conclusion that early broad-spectrum antibiotics are associated with reduced mortality.

One of the most vexing problems in observational outcomes research is the question of missed or poorly measured confounders. After all of the measures above have been taken, the reader must still address the possibility that the observed effect (or lack of effect) is due to failure to account completely for a confounder. This effect, known as "residual confounding," is always a possibility in observational research. We recommend the reader consider the question in this form:

"Were there any omitted variables whose effect would have been strong enough and in the right direction to eliminate the observed effect?" Although these assessments are generally qualitative, they can be formally addressed. Sensitivity analysis can estimate how large an effect an unmeasured confounder would have to exert to eliminate or generate the observed study effect (85). Thus, if a study that appears to have included most clinically important variables and adjusted appropriately for known confounders reports a tenfold increase in effect with the intervention, the reader might safely assume that the exposure has some independent effect on the outcome.

On the other hand, when important clinical variables are omitted or when sensitivity analysis shows that a confounder with a small effect could have generated the observed results, and the study reports only a small effect size, one might have less confidence in the results. Ultimately, the criteria for judging all scientific knowledge apply: reproducibility of the findings from multiple studies with different methods, plausibility of the results, and the ability to use the results to make novel verifiable predictions.

IMPLICATIONS FOR THE PRACTICING CLINICIAN

In addition to seeing more observational outcomes research studies used to assess the effectiveness of treatments, clinicians will increasingly encounter the methods described previously in their practice. The database and analysis methods of observational outcomes research are used to "profile" individual clinicians for their cost and outcomes. In these reports the unit of analysis is the individual clinician; the outcome is mortality, hospital length of stay, or some other measure of resource use; and the confounders are disease severity and diagnosis. The question is whether differences in patient outcome can be reliably detected between clinicians using these techniques. Another local use of outcomes research is in quality improvement programs which should use quantitative methods to evaluate the effect of interventions. A common strategy is the "before-after" design. For example, baseline data on the incidence of ventilator-associated pneumonia and duration of mechanical ventilation are collected prior to the institution of a new antibiotic-prescribing strategy. After this strategy has been instituted, the quality improvement group reevaluates the outcomes taking care to use the techniques described previously to account for potential confounding differences in the patient population. Finally, outcomes research techniques are used at the policy level to assess clinician workforce needs and to set reimbursement levels for different procedures (86).

CONCLUSION

Multiple demands of fiscal constraint, rapidly advancing technology, reorganization of health care delivery in the United States, evidence-based medicine, and an informed and empowered consumer base have all contributed to the prominent role of outcomes research in current medical decisionmaking. In the future, detailed data on medical practice and outcomes from a broad range of providers will shed new light on current practice and suggest optimal care. The high financial and emotional burdens, limited evidence base, and changing technology make continued outcomes research essential in critical care.

Some critical care practitioners will feel more confident in experimental research methods than the observational research designs discussed in this report. It is important that clinicians and investigators appreciate that it is not a question of choosing between experimental and observational outcomes research; rather, it is a matter of understanding the strengths and weaknesses of each. The methods are complementary precisely because many important research questions are not amenable to an experimental design and many experimental designs have limited generalizability. Answers to important clinical and policy questions will require an evolving body of outcomes research derived from quantitative and qualitative methods as well as experimental and observational studies. For pulmonary and critical care clinicians to play an active role in determining policy and practice through the application of outcomes research, it is essential that they understand the strengths and limitations of observational outcomes research.

Members of the American Thoracic Society Outcomes Research in Critical Care Workshop, Toronto, Canada, April 4–6, 1997

Cochairs: Derek C. Angus, M.D., M.P.H., Gordon Bernard, M.D., Michael R. Pinsky, M.D., Gordon Rubenfeld, M.D., M.Sc. ***Critical Cure Subgroup:*** Gordon Rubenfeld, M.D., M.Sc. (Chair), Jean Carlet, M.D., Paul N. Lanken, M.D., Judith Lave, Ph.D., John Marshall, M.D., Arthur Slutsky, M.D., Ph.D. ***Outcomes Definitions Subgroup:*** Gordon Bernard, M.D. (Chair), Carolyn Clancy, M.D., Deborah Cook, M.D., J. Randall Curtis, M.D., M.P.H., Phillip Jacobs, Ph.D., Richard Straube, M.D. ***Data Model Assessment Subgroup:*** Derek C. Angus, M.D., M.P.H. (Chair), Alfred Connors, M.D., Lisa Iezzoni, M.D., James Jollis, M.D., Janelle Klar, Ph.D., Mark S. Roberts, M.D., M.P.P.. Walter T. Linde-Zwirble.

References

1. Cook, D. J., W. J. Sibbald, J. L. Vincent, and F. B. Cerra. 1996. Evidence based critical care medicine: what is it and what can it do for us? Evidence Based Medicine in Critical Care Group. *Crit.* Care ***Med.*** 24:334–***331.***
2. Schwartz, W. B. 1987. The inevitable failure of current cost containment strategies: why they can provide only temporary relief. ***J.A.M.A. 257:*** 220–224.
3. Jacobs, P., and T. W. Noseworthy. 1990. National estimates of intensive care utilization and costs: Canada and the United States. *Crit.* Care ***Med.*** 18:1282–1286.
4. Connors, A. F., Jr., T. Speroff, N. V. Dawson, C. Thomas, F. E. Harrell, Jr., D. Wagner, N. Desbiens, L. Goldman, A. W. Wu, R. M. Califf, W. J. Fulkerson, Jr., H. Vidaillet, S. Broste, P. Bellamy, J. Lynn, and W. A. Knaus. 1996. The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. ***J.A.M.A.*** 276:889–897.
5. Curtis, J., G. Rubenfeld, and L. Hudson. 1998. Training pulmonary and critical care physicians in outcomes research: should we take the challenge? ***Am. J. Respir. Crit. Care*** Med. 157:1012–1015.
6. Yusuf, S., R. Collins, and R. Peto. 1984. Why do we need some large, simple randomized trials? ***Stat. Med.*** 3:409–422
7. Matchar, D. B., G. P. Samsa, J. R. Matthews, M. Ancukiewicz, G. Parmigiani, V. Hasselblad, P. A. Wolf, R. B. D'Agostino, and J. Lipscomb. 1997. The Stroke Prevention Policy Model: linking evidence and clinical decisions. ***Ann. Intern. Med.*** 127(8, Pt. 2):704–711.
8. Petitti, D. B. 1994. Meta-analysis, Decision Analysis, and Cost-effectiveness Analysis: Methods. Oxford University Press, New York.
9. Lemeshow, S., and J. R. Le Gall. 1994. Modeling the severity of illness of ICU patients: a systems update. ***J.A.M.A.*** 272:1049–1055.
10. Knaus, W. A., J. E. Zimmerman, D. P. Wagner, E. A. Draper, and D. E. Lawrence. 1981. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit.* ***Care Med.*** 9:591–597.
11. Berenson, R. A. 1984. Intensive care units (ICUs): clinical outcomes, costs, and decisionmaking. Congress of the U.S.. Office of Technology Assessment, Washington, DC.
12. Wennberg, D., F. Lucas, J. Birkmeyer, C. Bredenberg, and E. Fisher. 1998. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. *J.A.M.A.* 27912781281.
13. Jollis, J. G., E. R. DeLong, E. D. Peterson, L. H. Muhlbaier, D. F. Fortin, R. M. Califf, and D. B. Mark. 1996. Outcome of acute myocardial infarction according to the specialty of the admitting physician. N. *Engl. J.* Med. 335:1880–1887.

14. Angus, D. C., W. T. Linde Zwirble, C. A. Sirio, A. J. Rotondi, L. Chelluri, R. C. Newbold, J. R. Lave, and M. R. Pinsky. 1996. The effect of managed care on ICU length of stay: implications for Medicare. *J.A.M.A.* 276:1075–1082.

15. Davidoff, F. 1997. Databases in the next millennium. ***Ann. Intern.*** Med. 127(8, Pt. 2):770–774.

16. Wennberg, J. E., J. L. Freeman, R. M. Shelton, and T. A. Bubolz. 1989. Hospital use and mortality among Medicare beneficiaries in Boston and New Haven. N. *Engl. J.* Med. 321:1168–1173.

17. Blumberg, M. S., and G. S. Binns. 1994. Swan-Ganz catheter use and mortality of myocardial infarction patients. Health Care *Financ.* ***Rev.*** 15:91–103.

18. Guyatt, G., Group OICS. 1991. A randomized controlled trial of right-heart catheterization in critically ill patients. *J. Intensive* Care Med. 6: 91-95.

19. Bone, R. C. 1992. Toward an epidemiology and natural history of SIRS (systemic inflammatory response syndrome). ***J.A.M.A.*** 268:3452–3455.

20. Bone, R. C. 1991. Sepsis, the sepsis syndrome, multi-organ failure: a plea for comparable definitions. ***Ann. Intern. Med.*** 114:332–333.

21. Vincent, J. L. 1997. Dear SIRS, I'm sorry to say that I don't like you. *Crit.* Care ***Med.*** 25:372–374.

22. Vincent, J. L., D. J. Bihari, P. M. Suter, H. A. Bruining, J. White, M. H. Nicolas-Chanoin, M. Wolff, R. C. Spencer, and M. Hemmer. 1995. The prevalence of nosocomial infection in intensive care units in Europe: results of the European Prevalence of Infection in Intensive Care (EPIC) Study. EPIC International Advisory Committee. *J.A.M.A.* 274:639–644.

23. Legall, J. R., S. Lemeshow, G. Leleu, J. Klar, J. Huillard, M. Rue, D. Teres, and A. Artigas. 1995. Customized probability models for early severe sepsis in adult intensive care patients. ***J.A.M.A.*** 273:644–650.

24. Wagner, D. P., W. A. Knaus, and E. A. Draper. 1987. Identification of low-risk monitor admissions to medical-surgical ICUs. ***Chest*** 92:423–428.

25. Daly, B. J., E. B. Rudy, K. S. Thompson, and M. B. Happ. 1991. Development of a special care unit for chronically critically ill patients. ***Heart*** Lung 20:45–51.

26. Krieger, B. P., P. Ershowsky, and D. Spivack. 1990. One year's experience with a noninvasively monitored intermediate care unit for pulmonary patients. ***J.A.M.A.*** 264:1143–1146.

27. Latriano, B., P. McCauley, M. E. Astiz, D. Greenbaum, and E. C. Rackow. 1996. Non-ICU care of hemodynamically stable mechanically ventilated patients. ***Chest*** 109:1591–1596.

28. Carson, S. S., C. Stocking, T. Podsadecki, J. Christenson, A. Pohlman, S. MacRae, J. Jordan, H. Humphrey, M. Siegler, and J. Hall. 1996. Effects of organizational change in the medical intensive care unit of a teaching hospital: a comparison of 'open' and 'closed' formats. *J.A.M.A.* 276:322–328.

29. Li, T. C., M. C. Phillips, L. Shaw, E. F. Cook, C. Natanson, and L. Goldman. 1984. On-site physician staffing in a community hospital intensive care unit: impact on test and procedure use and on patient outcome. ***J.A.M.A.*** 252:2023–2027.

30. Rubenfeld, G. D., and S. W. Crawford. 1996. Withdrawing life support from mechanically ventilated recipients of bone marrow transplants: a case for evidence-based guidelines. ***Arm. Intern. Med.*** 125:625–633.

31. Hadom, D., E. Keeler, W. Rogers, and R. Brook. 1993. Assessing the performance of mortality prediction models. Rand, Santa Monica, CA.

32. Laine, C., and F. Davidoff. 1996. Patient-centered medicine: a professional evolution. ***J.A.M.A.*** 275:152–156.

33. Curtis, J. R., and G. D. Rubenfeld. 1997. Aggressive medical care at the end of life: does capitated reimbursement encourage the right care for the wrong reason? [Editorial]. ***J.A.M.A.*** 278:1025–1026.

34. Deyo, R. A. ***1991.*** The quality of life, research, and care. ***J.A.M.A. 114: 695697.***

35. Bone, R. C., C. J. Fisher, T. P. Clemmer, G. J. Slotman, C. A. Metz, and R. A. Balk. 1987. A controlled clinical trial of high dose methylprednisolone in the treatment of severe sepsis and septic shock. *N. Engl. J.* ***Med.*** 317:653–658.

36. Quartin, A. A., R. M. Schein, D. H. Kett, P. N. Peduzzi, and Group ftDoVASSCS. 1997. Magnitude and duration of the effect of sepsis on survival. ***J.A.M.A.*** 277:1058–1063.

37. Prendergast. T. J., and J. M. Luce. 1997. Increasing incidence of withholding and withdrawal of life support from the critically ill. ***Am. J.*** *Respir. Crit.* ***Care*** Med. 155:15–20.

38. Christakis, N. A., and D. A. Asch. 1993. Biases in how physicians choose to withdraw life support. *Lancet* 342:642–646.

39. Jayes, R. L., J. E. Zimmerman, D. P. Wagner, E. A. Draper, and W. A. Knaus. 1992. Do-not-resuscitate orders in intensive care units: current practices and recent changes. ***J.A.M.A.*** 270:2213–2217.

40. Knaus, W. A., F. E. Harrell, Jr., J. F. LaBrecque, D. P. Wagner, J. P. Pribble, E. A. Draper, C. J. Fisher, Jr., and L. Soll. 1996. Use of predicted risk of mortality to evaluate the efficacy of anticytokine therapy in sepsis. The rhIL-1ra Phase III Sepsis Syndrome Study Group. *Crit.* ***Care Med.*** 24:46–56.

41. Curtis, J. R., D. P. Martin, and T. M. Martin. 1997. Patient-assessed health outcomes in chronic lung disease: what are they, how do they help us, and where do we go from here? ***Am. J. Respir. Crit. Care*** Med. 156:1032–1039.

42. Fowler, F. J., P. D. Cleary, J. Magaziner, D. L. Patrick, and K. L. Benjamin. 1994. Methodological issues in measuring patient-reported outcomes: the agenda of the work group on outcomes assessment. ***Med. Care*** 32:JS65–JS76.

43. Juniper, E. F., G. H. Guyatt, R. S. Epstein, P. J. Ferrie, R. Jaeschke, and T. K. Hiller. 1992. Evaluation of impairment of health related quality of life in asthma: development of a questionnaire for use in clinical trials. ***Thorax*** 47:76–83.

44. Guyatt, G. H., M. Townsend, L. B. Berman, and S. 0. Pugsley. 1987. Quality of life in patients with chronic airflow limitation. *Br. J. Dis.* ***Chest*** 81:45–54.

45. Gold, M. R., L. B. Russell, J. E. Seigel, and M. C. Weinstein. 1996. Cost-Effectiveness in Health and Medicine. Oxford University Press, New York.

46. Danis, M., D. L. Patrick, L. I. Southerland, and M. L. Green. 1988. Patients' and families' preferences for medical intensive care. ***J.A.M.A.*** 260:797–802.

47. Sprangers, M. A., and N. K. Aaronson. 1992. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *J. Clin. Epidemiol.* 45:743–760.

48. Tsevat, J., N. V. Dawson, A. W. Wu, J. Lynn, J. R. Soukup, E. F. Cook, H. Vidaillet, and R. S. Phillips. 1998. Health values of hospitalized patients 80 years or older. HELP Investigators. Hospitalized Elderly Longitudinal Project. ***J.A.M.A.*** 279:371–375.

49. Nease, R. F., Jr., T. Kneeland, G. T. O'Connor, W. Sumner, C. Lumpkins, L. Shaw, D. Pryor, and H. C. Sox. 1995. Variation in patient utilities for outcomes of the management of chronic stable angina: implications for clinical practice guidelines. Ischemic Heart Disease Patient Outcomes Research Team. ***J.A.M.A.*** 273:1185–1190.

50. Curtis, J. R., M. Ullman, A. C. Collier, M. R. Krone, B. R. Edlin, and C. L. Bennett. 1997. Variation in medical care of HIV-related *Pneumocystis carinii* pneumonia: a comparison of process and outcome at two hospitals. ***Chest*** 112:398–405.

51. Mant, J., and N. Hicks. 1995. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. ***B.M.J.*** 311:793–796.

52. Mangialardi, R. J., and G. B. Bernard. 1997. Designing the perfect trial in ARDS and ALI. *In* J. J. Marini and T. W. Evans, editors. Update in Intensive Care and Emergency Medicine: Acute Lung Injury. Springer-Verlag, Berlin. 421-429.

53. Foley, K. 1997. Competent care for the dying instead of physician-assisted suicide. N. ***Engl. J. Med.*** 336:54–58.

54. Rubenfeld, G. D. 1998. Cost-effectiveness considerations in critical care. New ***Horizons*** 6:33–40.

55. Weinstein, M. C., J. E. Siegel, M. R. Gold, M. S. Kamlet, and L. B. Russell. 1996. Recommendations of the Panel On Cost-Effectiveness in Health and Medicine. ***J.A.M.A.*** 276:1253–1258.

56. Siegel, J. E., M. C. Weinstein, L. B. Russell, and M. R. Gold. 1996. Recommendations for reporting cost-effectiveness analyses. ***J.A.M.A. 276: 1339-1341.***

57. Weinberger, M., S. Hui, and C. Laine. 1997. Measuring quality, outcomes, and cost of care using large databases. ***Ann. Intern. Med.*** 127(8 Pt. 2):665–774.

58. Iezzoni, L. I. 1990. Using administrative diagnostic data to assess the quality of hospital care: pitfalls and potential of ICD-9-CM. ***Int. J. Technol. Assess. Health Care*** 6:272–281.

59. Jollis, J. G., M. Ancukiewicz, E. R. DeLong, D. B. Pryor, L. H. Muhlbaier, and D. B. Mark. 1993. Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. ***Ann. Intern. Med.*** 119:844–850.

60. Iezzoni, L. I., S. M. Foley, J. Daley, J. Hughes, E. S. Fisher, and T. Heeren. 1992. Comorbidities, complications, and coding bias: does the number of diagnosis codes matter in predicting in hospital mortality? ***J.A.M.A.*** 267:2197–2203.

61. Romano, P. S., L. L. Roos, H. S. Luft, J. G. Jollis, and K. Doliszny. 1994. A comparison of administrative versus clinical data: coronary artery bypass surgery as an example. Ischemic Heart Disease Patient Outcomes Research Team. *J. Clin.* ***Epidemiol.*** 47:249–260.

62. Poses, R. M., W. R. Smith, D. K. McClish, and M. Anthony. 1995. Controlling for confounding by indication for treatment: are administrative data equivalent to clinical data? Med. Care 33(Suppl. 4):AS36–46.
63. Clemmer, T. P., and R. M. Gardner. 1991. Medical informatics in the intensive care unit: state of the art 1991. *Int. J. Clin. Monit. Comput.* **8:** 237–250.
64. Milberg, J. A., D. R. Davis, K. P. Steinberg, and L. D. Hudson. 1995. Improved survival of patients with acute respiratory distress syndrome (ARDS): 1983-1993. ***J.A.M.A.*** 273:306–309.
65. Califf, R. M., and D. B. Mark. 1994. Percutaneous intervention, surgery, and medical therapy: a perspective from the Duke Databank for Cardiovascular Diseases. ***Semin. Thorac. Cardiovasc. Surg.*** 6:120–128.
66. Pennsylvania Health Care Cost Containment Council (PHC4). 1995. A consumer guide to coronary artery bypass graft surgery, 1993 data. Volume IV. Pennsylvania Health Care Cost Containment Council, Harrisburg, PA.
67. Pennsylvania Health Care Cost Containment Council (PHC4). 1995. Focus on heart attack in Western Pennsylvania: a 1993 summary report for health benefits purchasers, health care providers, policymakers and consumers. Pennsylvania Health Care Cost Containment Council, Harrisburg, PA.
68. Sirio, C. A., D. C. Angus, and G. E. Rosenthal. 1994. Cleveland Health Quality Choice (CHQC)-an ongoing collaborative, community-based outcomes assessment program. New ***Horizons*** 2:321–325.
69. Iezzoni, L. I. 1997. The risks of risk adjustment. *J.A.M.A.* 278:1600–1607.
70. Spoeri, R. K., and R. Ullman. 1997. Measuring and reporting managed care performance: lessons learned and new initiatives. ***Ann. Intern.*** Med. 127(8, Pt. 2):726–732.
71. Mak, K. H., D. J. Moliterno, C. B. Granger, D. P. Miller, H. D. White, R. G. Wilcox, R. M. Califf, and E. J. Topol. 1997. Influence of diabetes mellitus on clinical outcome in the thrombolytic era of acute myocardial infarction. GUSTO-I Investigators. Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries. *J.* Am. *Coll. Cardiol.* 30:171–179.
72. Carson, J. L., M. A. Kelley, A. Duff, J. G. Weg, W. J. Fulkerson, H. I. Palevsky, J. S. Schwartz, B. T. Thompson, J. Popovich, Jr., T. E. Hobbins, et ***al.*** 1992. The clinical course of pulmonary embolism. N. *Engl. J. Med.* 326:1240–1245.
73. Knaus, W. A., D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, A. Damiano, and F. E. Harrell, Jr. 1991. The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. ***Chest*** 100:1619–1636.
74. Lemeshow, S., D. Teres, J. Klar, J. S. Avrunin, S. H. Gehlbach, and J. Rapoport. 1993. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. *J.A.* ***M.A.*** 270:2478–2486.
75. Nelson, L. D. 1997. Data, data everywhere. *Crit. Care Med.* 25:1265.
76. Day, N. E., D. P. Byar, and S. B. Green. 1980. Overadjustment in case-control studies. ***Am. J. Epidemiol.*** 112:696–706.
77. Hadorn, D. C., D. Draper, W. H. Rogers, E. B. Keeler, and R. H. Brook. 1992. Cross-validation performance of mortality prediction models. ***Stat. Med.*** 11:475–489.
78. Concato, J., A. R. Feinstein, and T. R. Holford. 1993. The risk of determining risk with multivariable models. ***Ann. Intern.*** Med. 118:201–210.
79. Angus, D. C., and M. R. Pinsky. 1997. Risk prediction: judging the judges. ***Intensive Care Med.*** 23:363–365.
80. Rosenbaum, P., and D. Rubin D. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55.
81. McClellan, M., B. J. McNeil, and J. P. Newhouse. 1994. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. ***J.A.M.A.*** 272:859–***866.***
82. Moses, L. E. 1995. Measuring effects without randomized trials? Options, problems, challenges. ***Med. Care*** 33(Suppl. 4):AS8–AS14.
83. Byar, D. P. 1991. Problems with using observational databases to compare treatments. ***Stat. Med.*** 10:663–666.
84. Sackett, D. L. 1979. Bias in analytic research. ***J. Chronic Dis.*** 32:51–63.
85. Rosenbaum, P. R., and D. B. Rubin. 1983. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J.* ***R. Stat.*** *Soc.* 45:212–218.
86. Goodman, D. C., E. S. Fisher, T. A. Bubolz, J. E. Mohr, J. F. Poage, and J. E. Wennberg. 1996. Benchmarking the US physician workforce: an alternative to needs based or demand based planning. ***J.A.M.A. 276: 1811-1817.***
87. Patrick, D. L., M. Danis, L. I. Southerland, and G. Hong. 1988. Quality of life following intensive care. ***J. Gen. Intern. Med.*** 3:218–223.
88. Thiagarajan, J., P. Taylor, E. Hogbin, and S. Ridley. 1994. Quality of life after multiple trauma requiring intensive care. *Anaesthesia* 49:211–218.
89. Miranda, D. R. 1994. Quality of life after cardiopulmonary resuscitation. ***Chest*** 106:524–530.
90. Sage, W. M., M. H. Rosenthal, and J. F. Silverman. 1986. Is intensive care worth it? An assessment of input and outcome for the critically ill. ***Crit. Care Med.*** 14:777–782.
91. Perl, T. M., L. Dvorak, T. Hwang, and R. P. Wenzel. 1995. Long-term survival and function after suspected gram-negative sepsis. ***J.A.M.A.*** 274:338–345.
92. Chelluri, L., M. R. Pinsky, M. P. Donahoe, and A. Grenvik. 1993. Long-term outcome of critically ill elderly patients requiring intensive care. ***J.A.M.A.*** 269:3119–3123.
93. Yau, E., A. Z. Rohatiner, T. A. Lister, and C. J. Hinds. 1991. Long-term prognosis and quality of life following intensive care for life-threatening complications of haematological malignancy. *Br. J. Cancer* 64:938–42.