Health Outcomes

The Outcomes of Outcomes and Effectiveness Research: Impacts and Lessons from the First Decade

Daniel Stryer, Sean Tunis, Heddy Hubbard, and Carolyn Clancy

Objective. To assess the outcomes of the Agency for Healthcare Research and Quality's (AHRQ; formerly the Agency for Health Care Policy and Research, AHCPR) first decade of focus on outcomes and effectiveness research (OER) and to identify needs and opportunities for the study of OER in the coming years.

Data Source. Study findings were collected in response to an inquiry by the Center for Outcomes and Effectiveness Research at AHRQ in July 1997 to all principal investigators (PIs) funded between 1989 and 1997. The request was for investigators to identify their "most salient findings" and supply material for up to three slides.

Study Design. A taxonomy of 11 non–mutually exclusive categories was used to group the investigators' salient findings by characteristics of methodology or purpose. Two health services researchers assigned findings to up to three categories for each discrete study.

Principal Findings. Responses were received from 61 (64 percent) of the 91 PIs, reporting on 115 studies. Of the 246 category assignments made, descriptive epidemiology was the most common (24 percent), followed by comparative effectiveness (17 percent) and economic assessments (12 percent). Most studies were retrospective analyses of administrative data. Viewed within a conceptual framework for assessing the impact of research, OER has built a solid foundation for future quality improvement efforts by identifying problems, generating hypotheses, and developing new methodologies and has had limited impact on health care policies, practices and outcomes.

Conclusions. OER has had moderate but significant success meeting initial expectations for the field. Challenges for the next generation of OER include advancing from hypothesis generation to definitive studies of effectiveness, and acceleration of the process by which findings effect policy, practice, and outcomes.

Key Words. Outcomes research, effectiveness, impact, research policy, evaluation

Ed Koch, during his tenure as mayor of New York City, was well known for asking voters, "How'm I doing?" Although the constituency is more readily identifiable and easier to poll in elective politics than in health care research, such an assessment is essential in determining future directions for a federal research agency such as the Agency for Healthcare Research and Quality (AHRQ; formerly the Agency for Health Care Policy and Research, AHCPR). As the focus on outcomes and effectiveness research (OER) within the health services community enters its second decade, the staff of the Center for Outcomes and Effectiveness Research (COER) at AHRQ is asking this question to identify opportunities, challenges, and needs in the coming years within the context of both federally and non-federally funded OER.

In addition to internal interest in targeting funding of research to improve health care most efficiently, this assessment is stimulated by increasing demands for accountability for federal investments in research. Concurrent with purchasers' demands for evidence of value from health care delivery, the Government Performance and Results Act of 1993 (GPRA) stresses the importance of government agencies to monitor their results and outcomes to improve programmatic direction and management (U. S. Congress 1993). In short, the research enterprise itself must be accountable to its "investors" and stakeholders. To our knowledge, federal agencies have focused their initial GPRA responses on process indicators (e.g., number of grants funded) and success stories rather than systematic assessments of outcomes. The Advisory Committee to the Director of the National Institutes of Health (NIH) was told that "NIH plans to furnish DHHS with selected anecdotes of past and current scientific advances. This documentation is primarily qualitative . . ." (National Institutes of Health 1997). This article discusses the outcomes of OER, including those identified in a survey of investigators within the contexts of early expectations for the work and a framework for understanding the work's impact.

The opinions in this article are those of the authors and do not represent official policy of the Agency for Health Care Policy and Research or the U. S. Department of Health and Human Services. This study was presented, in part, at the 1998 Annual Meeting of the Society of General Internal Medicine, Chicago, April 1998.

Address correspondence to Daniel Stryer, M.D., Center for Outcomes and Effectiveness Research, Agency for Health Care Policy and Research, Rockville, MD. Sean Tunis, M.D., M.Sc. is Director, Coverage and Analysis Group, Health Care Financing Administration, Baltimore, MD. Heddy Hubbard, R.N., M.P.H., and Carolyn Clancy, M.D. are from the Center for Outcomes and Effectiveness Research, Agency for Health Care Policy and Research, Rockville, MD. This article, submitted to *Health Services Research* on March 29, 1999, was revised and accepted for publication on November 16, 1999.

An assessment of the AHRQ-funded outcomes and effectiveness studies is best accomplished with knowledge of the well-documented ideas, trends, and expectations that led to the outcomes movement of the late 1980s (Epstein 1990). In particular, greater attention to patient-oriented outcomes developed, partly from increasing prevalence of chronic diseases and growing recognition of patient autonomy. Concurrently, a number of changes and findings led to what was termed the third revolution in medical care, that is, the era of assessment and accountability (Relman 1988). The Physician Payment Review Commission (PPRC) developed and Congress adopted the Resource-Based Relative Value Scale, a novel reimbursement system under which the financial incentives to perform procedures were reduced. By doing so, the commission postulated that physicians would make management decisions based on medical effectiveness rather than financial gain.

Institution of a prospective payment system in Medicare and increasing emphasis on cost containment in the 1980s led to fears that such efforts could have a deleterious effect on patients. As a result, policymakers focused greater attention on methods to monitor the results of care (Gray 1992).

During the same period, John Wennberg, an epidemioligist at Dartmouth Medical School and others brought increased attention to the effects of practice styles on utilization and outcomes after documenting regional variations in practice patterns that were not explained by differences in medical needs or outcomes. For example, in the late 1980s they noted marked differences between Boston, Massachussetts, and New Haven, Connecticut, in rates of hospitalization, length of stay, readmission, and reimbursement but found that mortality rates were nearly identical. Their findings suggested that there would have been a reduction of approximately $300 million in hospital expenditures and a need for 739 fewer beds if the rates for New Haven occurred in Boston in 1982 (Wennberg, Freeman, and Culp 1987; Wennberg et al 1989). While researchers and policymakers were unsure how these changes could be accomplished, experts at an invitational conference on quality assessment in health care, sponsored by the Health Care Financing Administration (HCFA), agreed that more knowledge of the effects of variations in practice styles among providers and institutions was needed (Relman 1988).

Building on these emerging findings, William Roper, then the Administrator for the HCFA, and others announced the Effectiveness Initiative, a major federal effort to extend the use of Medicare claims databases beyond quality control to improve the knowledge base for medicine. A four-step process was outlined: monitoring to identify trends in processes or end results;

analysis of variations in terms of utilization patterns and outcomes; assessment of the effectiveness of interventions; and feedback and education (Roper et al. 1988).

An important premise underlying this vision was that optimal medical practices could be determined from analysis of computerized data routinely gathered in the process of patient care. This represented an appealing hypothesis: that data collected for administrative purposes could be mined to generate new knowledge about which practices work best. Researchers and policy makers expected that analytic efforts such as database analysis, systematic literature reviews, decision analysis, and guideline development could substitute for expensive clinical trials that often require years to complete and may have limited generalizability. Importantly, the initiative assumed that physicians would rapidly change their practices when presented with new information and that ensuing improvements in effectiveness and appropriateness would result in substantial and measurable savings.

The promise of large economic savings resulting from better information on the outcomes of care created unprecedented expectations. While AHRQ-funded investigators have been recognized for their leadership in developing the discipline's methodologies and a knowledge base of existing health care practices, the OER enterprise has also been criticized for failing to measure up to standards demanded by some policymakers. Specifically, critics have noted that this research has not fulfilled its promise as an inexpensive method to improve clinical practice and ultimately save money (Anderson 1994; Sheldon 1994).

To provide some baseline information for comparing AHRQ-funded OER with original expectations for this research and to identify future directions and challenges for OER, a survey of research grant principal investigators (PIs) was conducted. The survey sought to identify common themes derived from the findings that OER investigators identified as most important. It also was intended to provide insight into whether there is congruence between what researchers and what others—policymakers, providers, and payers—consider to be most important.

METHODS

A letter was sent to all 95 PIs for Patient Outcomes Research Teams (PORT) and Medical Treatment Effectiveness Program (MEDTEP) studies funded by AHRQ between 1989 and 1997. The letter asked investigators to "identify the most salient findings" from each study and provide sufficient material in "succinct, bullet-point format" for up to three slides. No specific criteria were provided to define "salience," nor were PIs instructed about the number of findings to be submitted. Several investigators were contacted to clarify their findings. Also, because of their core role in AHRQ's program in OER, PIs from the PORT projects and minority MEDTEP centers who did not respond were contacted.

After reviewing a sample of responses, a taxonomy for classifying the findings was developed to group results by purpose or methodology. Eleven non–mutually exclusive, descriptive categories were identified and defined after pilot testing and review of an original list of 14 categories by the authors and other COER staff (Table 1). Two physician/health services researchers reviewed the complete set of "salient" findings submitted by investigators and assigned each to one or more categories. Disagreements were resolved by discussion. Some PIs listed findings from a number of separate studies

Table 1: Categories for Findings Reported by Principal Investigators

Descriptive epidemiology: natural history (follows patient without intervention), description of incidence or prevalence of condition, treatment, or complication, or identification of risk factors.

Comparative effectiveness: comparison of outcomes associated with different treatments, diagnostic approaches, or other management strategies, that is, "What works best."

Economic assessment: compares costs and outcomes for different interventions, or looks at the costs of diseases or treatment without examining outcomes.

Patient-reported outcomes: studies that report general or disease-specific quality of life, functional status, patient preferences, or patient satisfaction.

Practice variations: comparison of rates of utilization of treatment, or service by some categorical group, such as region, demographic feature, provider, institution, or time period.

Sociology of health care: describes patient or provider knowledge, beliefs and attitudes about conditions, treatments, process of care, etc.

Methodological development: design, refinement, or validation of a technique for measuring, analyzing, or defining. Includes survey methods and risk adjustment or severity adjustment methods.

Modeling: use of modeling techniques for decision analysis, predicting outcomes of interest, or determining cost-effectiveness of specific interventions.

Quality of care: findings related to measurement of health care quality or provider performance.

Systematic review/meta-analysis: summarizes existing literature with or without statistical analysis.

Legal/legislative/regulatory: describes or measures impact of an action of courts or government rulings that are related to health care.

while others listed a number of findings from an individual study. Findings for each discrete study were assigned to up to three descriptive categories.

RESULTS

Responses were received from 61 (64 percent) researchers and a total of 246 category assignments were made (Table 2). Nonrespondents tended to be investigators for smaller grants in terms of time and funding or for grants that had been recently funded at the time of the survey. Also, a number of investigators could not be contacted. "Descriptive epidemiology" was the largest category, representing 24 percent of the important findings reported. Studies categorized as descriptive epidemiology were those that described patterns of health care conditions or utilization. For example, one study identified the major risk factors for chlamydia infection, which included patients under age 30 who presented with cervical friability, had multiple sexual partners, and were African American (Sonnenberg 1995). Another study found that 15-year, disease-specific mortality rate for men with prostate cancer and a Gleason score of 2 to 4 (a grading system based on tumor differentiation) was 4 to 7 percent (Albertsen et al. 1998). Another primarily descriptive category, "practice variations," was the fifth largest category. These studies documented variations in practice patterns that occurred in different geographic areas,

Table 2: Classification of Salient Findings

Classification	*Number*	*Percent*[*]
Descriptive epidemiology	58	24
Comparative effectiveness	42	17
Economic assessment	29	12
Patient-reported outcomes	25	10
Practice variations	23	9
Sociology of health care	17	7
Methodological development	17	7
Modeling	16	7
Quality of care	10	4
Systematic review or meta-analysis	5	2
Legal, legislative or regulatory	4	2
Total	246	100

[*] Percentage does not add up to 100 because of rounding.

types of providers, demographic groups, or time intervals. The two-fold variation in use rates of electrophysiologic studies and implantable cardioverter defibrillator for management of malignant arrhythmias is representative of this category (Clark, Alexander, Baker, et al. 1998).

The second, third, and fourth largest categories focused on an array of outcomes associated with clinical interventions, including clinical, economic, and quality of life. Comparative effectiveness studies documented associations between different practice patterns and outcomes, such as one that found that laparoscopic cholecystectomy is at least as safe as open cholecystectomy, despite the higher rate of common bile duct injuries associated with the former (Shea, Healey, Berlin, et al. 1996). While some of the studies in this category were prospective observational studies, most were retrospective analyses of claims or other administrative databases.

Less frequently reported categories included "modeling," "quality of care," and "systematic review/meta-analysis." These represented efforts to build upon previous research, making it more usable by researchers, providers, and policymakers. The remainder of the topics consisted of a broad range, including "sociology of health care," "methodological development," and "legal/legislative/regulatory."

DISCUSSION

When PIs were asked to identify their most salient findings, a large portion of responses identified studies that were descriptive, providing important information about who is affected by a disease, how patients are managed, and what the consequences of management are. Although some studies compared the outcomes associated with alternative clinical interventions, few of these studies were definitive because they were retrospective analyses with limited ability to control for selection bias (Maklan, Greene, and Cummings 1994). Instead, the studies were often hypothesis generating, leading to refinement of specific clinical questions and subsequent studies. In addition, methodological studies were performed, adding to the foundation necessary for subsequent work. These include techniques for disease severity adjustment, meta analysis, measurement of patient-oriented outcomes, and development of models that related the process of care to those outcomes. Two such patient-oriented outcomes are the American Urological Association (AUA) Symptom Scale and the Visual Function (VF)-14, which have replaced previous gold standards in determining appropriateness and outcomes in prostate and cataract operations, respectively (Barry, Cockett, Holtgrewe, et al. 1993; Schein, Steinberg, Cassard, et al. 1995). The essential contributions of these methodological advances are cited in an analysis of private sector activity in OER (Mendelson et al. 1998).

The survey results were also revealing for what was not reported. Few studies specifically addressed the primary expectations for OER, which were changes in practice resulting in improvements in quality and outcomes, and cost reductions. Investigators infrequently identified findings which had been successfully disseminated and implemented at either the individual or institutional level.

Framework for Assessing Impact

The findings of this survey stimulated thinking about the process by which research findings are linked over time to changes in policy, practice, and outcomes. A model depicting potential levels of impact was developed by the authors and others at AHRQ following review of the PI findings and is shown in Figure 1. Level 1 impacts represent the foundation of studies that identify problems, generate hypotheses, establish the effectiveness of interventions, and develop new tools to explore these problems. These studies include findings which conflict with current clinical consensus and practice. For example, the literature reviews done by the Low Birthweight PORT under the direction

Figure 1: Diagram of Levels of Outcomes Research

of Robert Goldenberg found that many common practices to prevent preterm delivery were ineffective (Goldenberg and Rouse 1998). Level 1 studies also include those in which new measurement instruments, such as the VF-14 in the management of cataract disease (Steinberg, Tielsch, Schein, et al. 1994), and new severity measures, such as the Total Illness Burden Index (Greenfield, Sullivan, Dukes, et al. 1995), are developed and validated. These studies, however, do not lead quickly to change, and when change does occur, attribution back to the original studies may be difficult to establish. Despite this, the value of level 1 impact should not be minimized. Results at this level provide a foundation for subsequent studies and are a prerequisite to changes in practice and improvements in quality and outcomes.

Level 2 impacts are those in which a policy or program is created as a result of the research. Potential conduits of level 2 impacts include what we have termed "change agents" in today's health care delivery environment; these are health plans, professional organizations, legislative bodies, regulators, accrediting bodies, the media, industry, and patients. For example, the Thrombolytic Predictive Instrument, developed by Harry Selker, a general internist at New England Medical Center, and colleagues, is offered by a number of manufacturers as a standard feature on new electrocardiogram machines (Selker 1999). In some circumstances, where there is a strong link between process and outcomes, evidence for impact at lower levels strongly suggests that improvement in outcomes (level 4 impact, discussed below) will occur. Beta-blocker usage in elderly patients after myocardial infarctions (MI), based on the work of Stephen Soumerai and others (Soumerai, McLaughlin, Spiegelman, et al. 1997), was included in HEDIS 3.0, a set of quality indicators developed by the National Committee for Quality Assurance to evaluate managed care organizations (National Committee for Quality Assurance 1998). The use of beta-blockers in this way will probably reduce mortality in these patients, assuming rates increase in response to the measure's inclusion. The work of the Stroke Prevention PORT led by David Matchar (Matchar et al. 1994; Matchar, Samsa, and Cohen 1996) supported efforts to improve anticoagulation rates for atrial fibrillation. Peer review organizations in 42 states have implemented 73 quality improvement projects. Based on remeasurement data, anticoagulation rates have increased from 58.4 percent to 71.1 percent, resulting in an estimated 1,285 strokes prevented (Health Care Financing Administration 1998).

Level 3 impacts are those in which there is a documented change in what clinicians or patients do. For example, when patients in two managed care settings viewed the Shared Decisionmaking Program for BPH (benign prostatic hyperplasia), an interactive videodisc, 27 percent of men who initially favored surgery changed their minds and opted for conservative therapy. Overall rates for transurethral resection of the prostate declined to less than half that of the years before implementation of the interactive program and to a greater extent than those in the control group (Leviton, Goldenberg, Baker, et al. 1999).

Finally, research that results in changes in health outcomes, including clinical, economic, quality of life, and satisfaction, represents level 4 impacts. Examples of this level of impact include comparisons of outcomes between patients managed in chest pain diagnostic and treatment units and those managed under standard care. These studies found that total costs and length of stay for intervention patients were significantly reduced and satisfaction with care was increased as compared with control patients (Roberts, Zalenski, Mensah, et al. 1997; Rydman, Zalenski, Roberts, et al. 1997). A related study, not funded by AHRQ, found that rates of missed myocardial infarctions decreased from 4.5 to 0.4 percent ($p < .001$) (Graff, Dallara, Ross, et al. 1997). Level 4 impacts are most credible and of greatest value if evidence is from multiple institutions and geographic regions.

This framework represents a conceptual approach for understanding how research affects population health and emphasizes the role and importance of research that leads indirectly to changes in practice or outcomes. Also, the framework illustrates the prolonged time course and the incremental and often circuitous nature of research. Knowledge may develop through multiple studies addressing focused, seemingly unrelated questions, making it difficult to cite a particular study or a group of studies as the cause of change. Finally, it shows a number of levels at which health care's many interrelated factors, such as organization, delivery, and financing, can affect efforts to improve outcomes. For example, it is impossible to attribute the national decline in rates of surgery for BPH solely to an improved knowledge of outcomes of the different therapies and increased involvement of patients in the decision-making process. Some of the change is also attributable to a shift of patients from fee-for-service to managed care and the development of additional nonsurgical options.

The framework should not be taken to mean that research findings that do not lead to demonstrable impacts on practice or outcomes are without merit. The burden of such an expectation would be unreasonable given the present state of research funding and infrastructure and the time required for changes in practice to affect outcomes. Difficulty also lies in measuring changes in outcomes. As mentioned above, where there is a strong link between process and outcome, improvements in practice may be reasonably assumed to lead to improvements in patient outcomes.

Because of the last decade's dramatic changes in health care delivery systems and experience suggesting that some of the original hopes were not realistic, OER may be more constructively assessed using this framework rather than initial expectations. Viewed within this context, most of the "salient" findings cited by investigators would be classified as level 1 impacts: identifying problems, generating hypotheses, developing new methodologies, and providing a basis for future investigation.

Some OER studies have already led to higher levels of impact, having been embraced by change agents, providers, and patients. They have also resulted in documented changes in policies as well as the processes and outcomes of care. For example, work by the Ischemic Heart Disease PORT, headed by Elizabeth DeLong, a health services research at Duke University, found that percutaneous transluminal coronary angioplasty (PTCA) mortality rates are associated with both provider and hospital PTCA volume (Hannan, Racz, Ryan, et al. 1997). These findings in part led to recommendations by the American College of Cardiology and the American Heart Association to raise annual volume standards for hospitals and providers. The Low Birthweight PORT helped to define indications for the use of antenatal steroids and reasons for their underuse (Gardner and Goldenberg 1995). This information and an active dissemination effort resulted in an increase in their use from 33 percent to 68 percent of appropriate cases in the study's 13 hospitals across the country (Leviton, Goldenberg, Baker, et al. 1999).

The framework, however, does not capture a number of important, more general contributions of AHRQ's program in OER. It has played a major role in the expansion of interest and capacity in outcomes research in both the public and private sectors (Mendelson et al. 1998). Nearly all major pharmaceutical firms, 49 specialty societies, and numerous hospitals and managed care organizations now report activity in OER (American Medical Association 1996; Mendelson et al. 1998). Hundreds of researchers have gained experience in OER through AHCPR-supported (and now AHRQ-supported) training, dissertation grants, and other grant programs. The program has helped drive the paradigmatic shift away from clinical intuition and toward evidence as the basis for clinical decisions. Educators and specialty societies now rely on systematic reviews of information rather than expert opinion and individual experiences as the basis for their teachings and guidelines. Finally, the relatively modest budget of AHRQ and its approach to identifying research priorities have led to consideration of population health needs in deciding topics of study. A recent Institute of Medicine report emphasized the importance of disease burden, in addition to scientific opportunity, in guiding prioritization of biomedical research (National Institutes of Health 1998).

Limited Success or Changing Expectations

Viewed from the perspective of the principal investigators, OER has not, for the most part, led to rapid and efficient improvements in quality and decreases in costs as had been expected. Valuable knowledge has been produced and some policies and practices have been affected, but the findings cited by PIs do not reflect a dramatic impact on the content or outcomes of health care. In retrospect, expectations for OER were probably inflated, in part by proponents who were overly optimistic about the potential for mining administrative databases, in part by researchers anxious for funding, and in part by policymakers' need to identify fast, effective solutions to the problem of soaring health care costs.

The prolonged time course from initial investigation to measurement of improvement in outcomes is a second factor that has contributed to the disparity between expectations and outcomes for OER. In some areas that are less developed in terms of clinical research, resources have been concentrated in developing infrastructure, capabilities, and methodologies rather than conducting effectiveness studies themselves. Our present approaches to study design, data collection, analysis, and publication are not geared to rapid turnaround. For example, the average time between manuscript submission and publication is eight and one-half months (Goodman et al. 1994). A study of technology transfer demonstrated that innovations take four to six years before they achieve 25 citations in the medical literature (Altman and Goodman 1994). The assumption that providers would rapidly change their practices in response to evidence has proven overly optimistic (Davis et al. 1995), increasing the challenge and time needed to implement new findings.

In addition, this research is more expensive than anticipated. Although less costly than most randomized controlled trials, outlays for obtaining and analyzing administrative databases are substantial. For example, in the first year and a half of the Acute Myocardial Infarction PORT, costs for data acquired from HCFA were $183,761 (Weis 1991). The 14 initial PORTs committed 10 to 50 percent of their resources to database analysis (U. S. Congress 1995). Finally, improvements in quality of care are no longer thought to be synonymous with cost reductions in care. Although there are examples of better quality services costing less, such as the use of chest pain diagnostic and treatment units, in general this expectation has proven to be unrealistic. Previous estimates that 25 percent of health care is inappropriate and could be eliminated without affecting quality have not been supported (Naylor 1998; U. S. Congress 1994). Also, underuse of health services appears to be a significant problem (Brook, Kamberg, Mayer-Oakes, et al. 1990; Carlisle, Leape, Bickel, et al. 1999). This, coupled with the possibility that improvements in appropriateness may lead to a reallocation of services, suggests that efforts may actually increase costs in some situations.

Implications for the Future

As federally funded OER approaches its second decade, it faces a number of challenges. The fundamental demands for rapid improvement in quality and efficiency of health care will continue despite a growing appreciation for the value of more basic studies. OER must advance beyond its foundation of increasingly detailed descriptions of practice, hypothesis generation, and methodological advancements and provide more definitive answers to "what works." It should embrace a more systematic approach to identifying promising findings and respond to these with follow-up studies that may stimulate needed changes in care. A strategic approach to dissemination of research results to change agents should be developed. Rather than depending on serendipity, dissemination should become an active effort in which key findings are provided in useable form to health plans, professional organizations, provider groups, policymakers, and the media.

OER will benefit from a stronger link to related fields necessary for accelerating the translation of research findings into practice. A better understanding of the type, quality, and quantity of evidence necessary to stimulate improvement is needed. For example, practitioners and policymakers were unwilling to embrace wholeheartedly thrombolytics for acute myocardial infarctions until there were multiple large, randomized, controlled trials; yet they have been more willing to embrace the results of an observational study showing that the benefits of beta-blockers post MI extend to the elderly. Also, an adequate understanding of the process of knowledge diffusion is lacking. The "readiness for change" transtheoretical model suggests that behavior evolves through stages from precontemplation, to contemplation, to preparation, to initiation, and to maintenance of change. While this model suggests a method for targeting interventions to improve care rapidly and efficiently, evidence of its effectiveness is limited (Cohen, Halvorsen, and Gosselink 1994). Additional effective methods of implementation must be identified. Finally, capacity for real-time data collection and analysis must be developed to accelerate the pace of change in practice and outcomes. For example, the National Eyecare Outcomes Network of the American Academy of Ophthalmology may be a model for efficient assessment of outcomes and improvement in care (National Eyecare Outcomes Network 1999).

In addition to highlighting needs and opportunities for OER, this evaluation also has implications for funding organizations. These groups should institute procedures whereby new research needs can more rapidly lead to targeted funding. Also, governmental and nongovernmental research entities must consider new procedures to shorten the review cycle or identify other mechanisms that more closely match the real-time needs of decision makers. Proposals to AHRQ are funded six to nine months after submission, and applications that undergo a single revision may be funded 18 months after an investigator starts to prepare an application. Finally, if researchers are expected to actively disseminate their findings, funding organizations must be willing to provide support for these activities.

So, how has OER done? By attempting to assess the outcome of outcomes research to inform future directions, AHRQ and other funders share common challenges. Demands for return on investment are unlikely to abate and are not limited to the federal government or to health services research. Not only has the Robert Wood Johnson Foundation grappled with how best to evaluate the impact of its nearly $350 million in annual grants and contracts (Schroeder 1998), but private companies in all sectors are often forced to sacrifice long-term research and development for shorter-term, applied projects (Pearlstein and Russakoff 1996). Determining the proper balance between the need and desire for immediate improvements in health care (level 3 and level 4 impacts) with methodological and hypothesis-generating studies (level 1) will undoubtedly continue to challenge stakeholders and funders.

As this broad initiative enters its second decade, OER is now recognized as an essential part of the foundation of efforts to assess and improve the quality of care. The results of the first decade of OER indicate, however, an imperfect alignment between researchers' interests and decision makers' needs. Close collaboration between these stakeholders will be needed to ensure that future efforts are as effective and efficient as possible.

ACKNOWLEDGMENTS

We are indebted to Dr. John Eisenberg and the members of the Center for Outcomes and Effectiveness Research for their thoughtful comments.

REFERENCES

Albertsen, P., J. Hanley, D. Gleason, and M. Barry. 1998. "Competing Risk Analysis of Men Aged 55 to 74 Years at Diagnosis Managed Conservatively for Clinically Localized Prostate Cancer." *Journal of the American Medical Association* 280 (11): 975–80.

Altman, D., and S. Goodman. 1994. "Transfer of Technology from Statistical Journals to the Biomedical Literature: Past Trends and Future Predictions." *Journal of the American Medical Association* 272 (2): 129–32.

American Medical Association, Department of Practice Parameters. 1996. *Outcomes Research Resource Guide,* 1996/1997 Edition. Publication No. OP047496. Chicago: American Medical Association.

Anderson, C. 1994. "Measuring What Works in Health Care." *Science* 263 (5150): 1080–82.

Barry, M., A. Cockett, H. Holtgrewe, J. McConnell, S. Sihelnik, and H. Winfield. 1993. "Relationship of Symptoms of Prostatism to Commonly Used Physiological and Anatomical Measures of the Severity of Benign Prostatic Hyperplasia." *Journal of Urology* 150 (2, Part 1): 351–58.

Brook, R. H., C. J. Kamberg, A. Mayer-Oakes, M. H. Beers, K. Raube, and A. Steiner. 1990. "Appropriateness of Acute Medical Care for the Elderly: An Analysis of the Literature." *Health Policy* 14 (3): 225–42.

Carlisle, D. M., L. L. Leape, S. Bickel, R. Bell, C. Kamberg, B. Genovese, W. J. French, V. S. Kaushik, P. R. Mahrer, M. H. Ellestad, R. H. Brook, and M. F. Shapiro. 1999. "Underuse and Overuse of Diagnostic Testing for Coronary Artery Disease in Patients Presenting with New-Onset Chest Pain." *American Journal of Medicine* 106 (4): 391–98.

Clark C., M. Alexander, L. Baker, K. McDonald, and M. Hlatky. 1998. "Racial and Ethnic Differences in the Use of Diagnostic and Treatment Procedures for Sudden Cardiac Death" Abstracts, Association of Health Services Research National Meeting, p. 263. Washington, DC.

Cohen, S., H. Halvorsen, and C. Gosselink. 1994. "Changing Behavior to Improve Disease Prevention." *Preventive Medicine* 23 (3): 284–91.

Davis, D., M. Thomson, A. Oxman, and R. Haynes. 1995. "Changing Physician Performance: A Systematic Review of the Effect of Continuing Medical Education Strategies." *Journal of the American Medical Association* 274 (9): 700–705.

Epstein, A. 1990. "The Outcomes Movement—Will It Get Us Where We Want To Go?" *New England Journal of Medicine* 323 (4): 266–70.

Gardner, M., and R. Goldenberg. 1995. "The Clinical Use of Antenatal Corticosteroids." *Clinical Obstetrics and Gynecology* 38 (4): 746–54.

Goldenberg, R., and D. Rouse. 1998. "Prevention of Premature Birth." *New England Journal of Medicine* 339 (5): 313–20.

Goodman, S., J. Berlin, S. Fletcher, and R. Fletcher. 1994. "Manuscript Quality Before and After Peer Review and Editing at Annals of Internal Medicine." *Annals of Internal Medicine* 121 (1): 11–21.

Graff, L. G., J. Dallara, M. A. Ross, A. J. Joseph, J. Itzcovitz, R. P. Andelman, C. Emerman, S. Turbiner, J. A. Espinosa, and H. Severance. 1997. "Impact On the Care of the Emergency Department Chest Pain Patient from the Chest Pain Evaluation Registry (CHEPER) Study." *American Journal of Cardiology* 80 (5): 563–68.

Gray, B. 1992. "The Legislative Battle Over Health Services Research." *Health* 11 (4): 38–66.

Greenfield, S., L. Sullivan, K. Dukes, R. Silliman, R. D'Agostino, and S. Kaplan. 1995. "Development and Testing of a New Measure of Case Mix for Use in Office Practice." *Medical Care* 33: AS47–AS55.

Hannan, E., M. Racz, T. Ryan, B. D. McCallister, L. W. Johnson, D T. Arani, A. D. Guerci, J. Sosa, and E. J. Topol. 1997. "Coronary Angioplasty Volume: Outcome Relationships for Hospitals and Cardiologists." *Journal of the American Medical Association* 277 (11): 892–98.

Health Care Financing Administration. 1998. *PRO Results: Bridging the Past with the Future.* September 1998. Baltimore, MD: Health Care Financing Administration.

Leviton, L., R. Goldenberg, C. Baker, R. M. Schwartz, M. C. Freda, L. J. Fish, S. P. Cliver, D. J. Rouse, C. Chazotte, C. Merkatz, and J. M. Raczynski. 1999. "Methods to Encourage the Use of Antenatal Corticosteroid Therapy for Fetal Maturation: A Randomized Controlled Trial." *Journal of the American Medical Association* 281 (1): 46–52.

Maklan, C., R. Greene, and M. Cummings. 1994. "Methodological Challenges and Innovations in Patient Outcomes Research." *Medical Care* 32 (7): JS13–JS21.

Matchar, D., D. McCrory, H. Barnett, and J. Feussner. 1994. "Medical Treatment for Stroke Prevention." *Annals of Internal Medicine* 121 (1): 41–53.

Matchar, D., G. Samsa, and S. Cohen. 1996. "Should We Just Let the Anticoagulation Service Do It? The Conundrum of Anticoagulation for Atrial Fibrillation." *Journal of General Internal Medicine* 11 (12): 768–70.

Mendelson, D., C. Goodman, R. Ahn, and R. Rubin. 1998. "Outcomes and Effectiveness Research in the Private Sector." *Health Affairs* 17 (5): 75–90.

National Committee for Quality Assurance. 1998. HEDIS 3.0/1998. Washington, DC: National Committee for Quality Assurance.

National Eyecare Outcomes Network. 1999. http://www.eyenet.org/member/ops/neon/. San Francisco: American Academy of Ophthalmology.

National Institutes of Health. 1997. *Proceedings of the 75th Meeting of the Advisory Committee to the Director, National Institutes of Health.* December 4, 1997. http://www.nih.gov/welcome/director/120497.htm.

———. 1998. *Scientific Opportunities and Public Needs: Improving Priority Setting and Public Input at NIH.* July 1998. Washington, DC: National Institutes of Health, Institute of Medicine, Committee on the NIH Research Priority-Setting Process.

Naylor, C. D. 1998. "What is Appropriate Care?" *New England Journal of Medicine* 338 (26): 1918–20.

Pearlstein, S., and D. Russakoff. 1996. "Rewired: Market-Driven Research Shakes the Ivory Tower; At Bell Labs, the Product is King; Telecommunications Competition Alters Role of Scientific Curiosity. *Washington Post,* November 25, 1996, p. A1.

Relman, A. 1988. "Assessment and Accountability: The Third Revolution in Medical Care." *New England Journal of Medicine* 319 (18): 1220–22.

Roberts, R., R. Zalenski, E. Mensah, R. J. Rydman, G. Ciavarella, L. Gussow, K. Das, L. M. Kampe, B. Dickover, M. F. McDermott, A. Hart, H. E. Straus, D. G. Murphy, and R. Rao. 1997. "Costs of an Emergency Department—

Based Accelerated Diagnostic Protocol vs Hospitalization in Patients with Chest Pain." *Journal of the American Medical Association* 278 (20): 1670–76.

Roper, W., W. Winkenwerder, G. Hackbarth, and H. Krakauer. 1988. "Effectiveness in Health Care: An Initiative to Evaluate and Improve Medical Practice." *New England Journal of Medicine* 319 (18): 1197–1202.

Rydman, R., R. Zalenski, R. Roberts, G. A. Albrecht, V. M. Misiewicz, L. M. Kampe, and M. McCarren. 1997. "Patient Satisfaction with an Emergency Department Chest Pain Observation Unit." *Annals of Emergency Medicine* 29 (1): 109–15.

Schein, O., E. Steinberg, S. Cassard, J. Tielsch, J. Javitt, and A. Sommer. 1995. "Predictors of Outcome in Patients Who Underwent Cataract Surgery." *Ophthalmology* 102 (5): 817–23.

Schroeder, S. 1998. "Reflections on the Challenges of Philanthropy. *Health Affairs* 17 (4): 209–16.

Selker H. 1999. Personal communication, January 13, 1999.

Shea, J., M. Healey, J. Berlin, J. R. Clarke, P. F. Malet, R. N. Staroscik, J. S. Schwartz, and S. V. Williams. 1996. "Mortality and Complications Associated with Laparoscopic Cholecystectomy: A Meta-Analysis. *Annals of Surgery* 224 (5): 609–20.

Sheldon, T. 1994. "Please Bypass the PORT." *British Medical Journal* 309 (6948): 142–43.

Sonnenberg, F. 1995. "A Clinical Decision Aid for Genital Chlamydia in Women." NTIS Accession No. PB97-251989, Final Report. Rockville, MD: Agency for Health Care Policy and Research.

Soumerai, S., T. McLaughlin, D. Spiegelman, E. Hertzmark, G. Thibault, and L. Goldman. 1997. "Adverse Outcomes of Underuse of Beta-Blockers in Elderly Survivors of Acute Myocardial Infarction." *Journal of the American Medical Association* 277 (2): 115–21.

Steinberg, E., J. Tielsch, O. Schein, J. C. Javitt, P. Sharkey, S. D. Cassard, M. W. Legro, M. Diener-West, E. B. Bass, and A. M. Damiano. 1994. "The VF-14: An Index of Functional Impairment in Patients with Cataract." *Archives of Ophthalmology* 112 (5): 630–38.

U. S. Congress. 1993. "Government Performance and Results Act of 1993." Public Law 103-62, 103rd Congress, Section 2. Washington, DC: U. S. Congress.

———. 1994. *Identifying Health Technologies That Work: Searching for Evidence.* Office of Technology Assessment, U. S. Congress, Report No. OTA-H-608, September, 1994. Washington, DC: U. S. Government Printing Office.

———. 1995. *Tools for Evaluating Health Technologies: Five Background Papers.* Office of Technology Assessment, U. S. Congress, Report No. BP-H-142, February, 1995. Washington, DC: U. S. Government Printing Office.

Weis, K. 1991. "PORT Data Requests Protocol." Agency for Health Care Policy and Research, April 25, 1991. Internal communication.

Wennberg, J., J. Freeman, and W. Culp. 1987. "Are Hospital Services Rationed in New Haven or Over-Utilised in Boston?" *Lancet* 1 (8543): 1185–89.

Wennberg, J., J. Freeman, R. Shelton, and T. Bubolz. 1989. "Hospital Use and Mortality among Medicare Beneficiaries in Boston and New Haven. *New England Journal of Medicine* 321 (17): 1168–73.